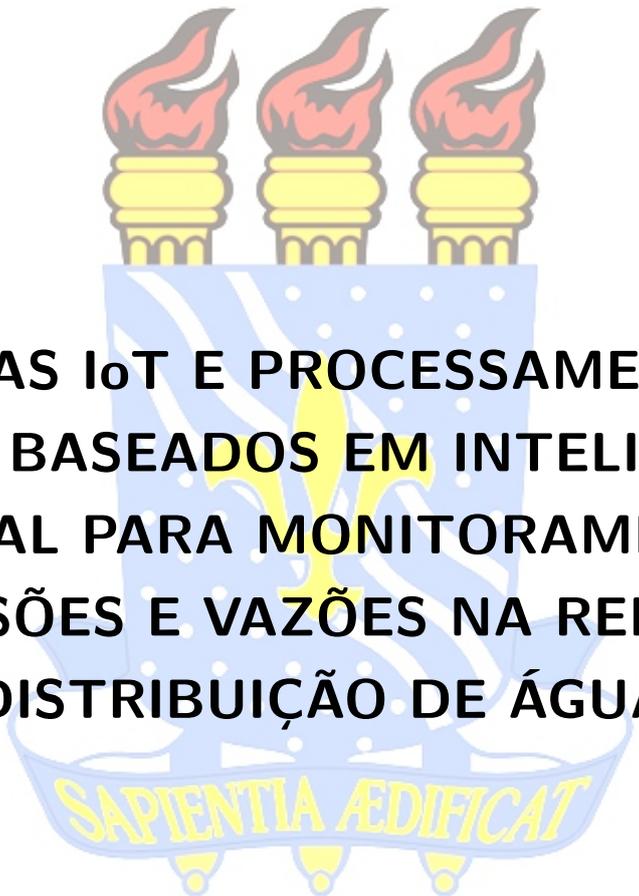


UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE ENERGIAS ALTERNATIVAS E RENOVÁVEIS
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

Jonatha Bizerra Silva

The logo of the Universidade Federal da Paraíba is centered in the background. It features three lit torches at the top, a shield with a blue and yellow design, and a banner at the bottom with the Latin motto "SAPIENTIA AEDIFICAT".

**SISTEMAS IoT E PROCESSAMENTO DE
DADOS BASEADOS EM INTELIGÊNCIA
ARTIFICIAL PARA MONITORAMENTO DE
PRESSÕES E VAZÕES NA REDE DE
DISTRIBUIÇÃO DE ÁGUA**

João Pessoa
2024

Jonatha Bizerra Silva

**SISTEMAS IoT E PROCESSAMENTO DE DADOS
BASEADOS EM INTELIGÊNCIA ARTIFICIAL PARA
MONITORAMENTO DE PRESSÕES E VAZÕES NA
REDE DE DISTRIBUIÇÃO DE ÁGUA**

Dissertação apresentada ao Programa de Pós Graduação em Engenharia Elétrica da Universidade Federal da Paraíba como exigência para a obtenção do título de Mestre em Engenharia Elétrica.

Universidade Federal da Paraíba

Centro de Energias Alternativas e Renováveis

Programa de Pós Graduação em Engenharia Elétrica

Orientador: Prof. Dr. Juan Moises Maurício Villanueva

João Pessoa

2024

Catálogo na publicação
Seção de Catalogação e Classificação

S586s Silva, Jonatha Bizerra.

Sistemas IoT e processamento de dados baseados em inteligência artificial para monitoramento de pressões e vazões na rede de distribuição de água / Jonatha Bizerra Silva. - João Pessoa, 2024.

79 f. : il.

Orientação: Juan Moises Maurício Villanueva.

Coorientação: Cícero da Rocha Souto.

Dissertação (Mestrado) - UFPB/CEAR.

1. Abastecimento de água - monitoramento. 2. Rede Neural Artificial (RNA). 3. Outliers. I. Villanueva, Juan Moises Maurício. II. Souto, Cícero da Rocha. III. Título.

UFPB/BC

CDU 628.1(043)

UNIVERSIDADE FEDERAL DA PARAÍBA – UFPB
CENTRO DE ENERGIAS ALTERNATIVAS E RENOVÁVEIS – CEAR
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA – PPGE

A Comissão Examinadora, abaixo assinada, aprova a Dissertação

**SISTEMAS IoT E PROCESSAMENTO DE DADOS BASEADOS EM INTELIGÊNCIA ARTIFICIAL PARA
MONITORAMENTO DE PRESSÕES E VAZÕES NA REDE DE DISTRIBUIÇÃO DE ÁGUA**

Elaborada por

JONATHA BIZERRA SILVA

como requisito parcial para obtenção do grau de
Mestre em Engenharia Elétrica.

COMISSÃO EXAMINADORA



PROF. DR. JUAN MOISES MAURICIO VILLANUEVA
Orientador – UFPB



PROF. DR. CÍCERO DA ROCHA SOUTO
Coorientador – UFPB

Documento assinado digitalmente

gov.br

CLEONILSON PROTASIO DE SOUZA

Data: 21/11/2024 12:09:31-0300

Verifique em <https://validar.iti.gov.br>

PROF. DR. CLEONILSON PROTASIO DE SOUZA
Examinador Interno – UFPB

Documento assinado digitalmente

gov.br

IVANOVITCH MEDEIROS DANTAS DA SILVA

Data: 15/11/2024 13:37:29-0300

Verifique em <https://validar.iti.gov.br>

PROF. DR. IVANOVITCH MEDEIROS DANTAS DA SILVA
Examinador Externo – UFRN

Documento assinado digitalmente

gov.br

RAFAEL MOURA DUARTE

Data: 25/11/2024 09:44:12-0300

Verifique em <https://validar.iti.gov.br>

PROF. DR. RAFAEL MOURA DUARTE
Examinador Externo – UNIVASF

AGRADECIMENTOS

Ao CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), à UFPB, a CAGEPA e a Amazon AWS, pelo apoio financeiro, por meio do edital Projeto 400046/2023-7, Chamada 64/2023 AWS-CNPq - Faixa A. Ao orientador deste trabalho, Prof. Juan Mauricio Villanueva e à equipe do LENHS (Laboratório de Eficiência Energética e Hidráulica em Saneamento) pelo suporte prestado para o desenvolvimento dessa pesquisa.

"Eu sou o silêncio em si."

(Mestre Pena Branca)

RESUMO

A água é um recurso natural essencial para a manutenção da vida na terra. Embora a maior parte do nosso planeta seja composta por água, apenas 25 % corresponde à água doce, e apenas 1 % está disponível para consumo. O Brasil detém aproximadamente 12 % desse percentual. Os sistemas de abastecimento de água acompanham o desenvolvimento das cidades, o crescimento populacional e sua distribuição geográfica, resultando em sistemas de tubulações com características heterogêneas, como diferentes diâmetros e idades de instalação, o que gera diversos desafios na operação e manutenção, além de comprometer a garantia dos requisitos técnicos mínimos de consumo. As perdas nos sistemas de abastecimento, inerentes a qualquer rede, têm recebido atenção crescente nos últimos anos devido à escassez hídrica e ao risco de contaminação da água tratada. A gestão eficiente dos recursos hídricos é crucial para a sobrevivência, ressaltando a necessidade de pesquisas que abordem as perdas associadas à distribuição de água para a população. Neste contexto, este trabalho tem como objetivo desenvolver uma metodologia baseada na *Typicality and Eccentricity Data Analytics* (TEDA) para a detecção de *outliers* em dados coletados em uma bancada experimental de monitoramento de abastecimento de água no Laboratório de Eficiência Energética e Hidráulica em Saneamento da Universidade Federal da Paraíba (LENHS/UFPB), além da aplicação de uma Rede Neural Artificial (RNA) para previsão da vazão no sistema monitorado. Os resultados indicam que, ao utilizar a RNA sem *outliers* nos dados, o modelo alcançou um erro percentual médio absoluto (MAPE) de 3,1 %, enquanto a aplicação da RNA com *outliers* nos dados de teste resultou em um MAPE de 5,2 %. Quando *outliers* foram introduzidos apenas nos dados de teste, o modelo apresentou um MAPE de 5,2 %, evidenciando a vulnerabilidade do sistema a dados atípicos. Após a aplicação do TEDA para a detecção de anomalias e a substituição dos dados detectados por leituras anteriores, a RNA obteve um MAPE de 3,0 %, refletindo uma melhora significativa na precisão das previsões. Esses avanços têm o potencial de impactar positivamente a gestão de recursos hídricos, minimizando perdas e promovendo uma utilização mais sustentável da água, especialmente em um cenário onde as perdas nos sistemas de distribuição são significativas. Esta pesquisa contribui para uma melhor compreensão do comportamento da vazão e da pressão nas redes de distribuição, possibilitando uma gestão mais eficaz dos recursos naturais.

Palavras-chave: TEDA, *outliers*, Rede Neural Artificial.

ABSTRACT

Water is a natural resource essential for sustaining life on Earth. Although most of our planet is composed of water, only 25 % is freshwater, and just 1 % is available for consumption. Brazil holds approximately 12 % of this percentage. Water supply systems accompany the development of cities, population growth, and geographical distribution, resulting in pipeline systems with heterogeneous characteristics, such as varying diameters and installation ages, which present various challenges in operation and maintenance, as well as compromising the guarantee of minimum technical consumption requirements. Losses in water supply systems, inherent to any network, have received increasing attention in recent years due to water scarcity and the risk of contamination of treated water. Efficient management of water resources is crucial for survival, highlighting the need for research addressing losses associated with water distribution to the population. In this context, this work aims to develop a methodology based on Typicality and Eccentricity Data Analytics (TEDA) for detecting outliers in data collected from an experimental water supply monitoring bench at the Laboratory of Energy Efficiency and Hydraulic Engineering in Sanitation at the Federal University of Paraíba (LENHS/UFPB), as well as the application of an Artificial Neural Network (ANN) for predicting flow in the monitored system. The results indicate that, when applying the ANN without the presence of outliers, the model achieved a Mean Absolute Percentage Error (MAPE) of 3.1 %. In contrast, when outliers were introduced only in the test data, the model resulted in a MAPE of 5.2 %, highlighting the system's vulnerability to atypical data. Following the application of TEDA for detecting anomalies and replacing detected data with previous readings, the ANN yielded a MAPE of 3.0 %, reflecting a significant improvement in prediction accuracy. These advancements have the potential to positively impact the management of water resources by minimizing losses and promoting a more sustainable use of water, especially in a scenario where losses in distribution systems are significant. The research contributes to a better understanding of the behavior of flow and pressure in distribution networks, enabling more effective management of natural resources.

Keywords: TEDA, *outliers*, Artificial Neural Network.

LISTA DE ILUSTRAÇÕES

Figura 1 – Sistema de Abastecimento de Água.	21
Figura 2 – Capacidade distribuída e consumo medido entre os anos de 2012 e 2016 nas cidades de Caruaru e Salgueiro, em Pernambuco, Brasil.	23
Figura 3 – Aplicações da Internet das Coisas.	25
Figura 4 – Representação de neurônios biológicos.	30
Figura 5 – Modelo artificial de neurônio biológico.	31
Figura 6 – Conjunto Motor Bomba.	33
Figura 7 – Bancada Experimental.	37
Figura 8 – Idealização do Sistema.	38
Figura 9 – Diagrama de fluxo dos dados.	41
Figura 10 – Controle do Sistema Supervisório.	44
Figura 11 – Correlação entre as variáveis de entrada da RNA e as vazões estimadas.	45
Figura 12 – Rede Neural Artificial com 5 entradas e duas saídas.	45
Figura 13 – Perfil de consumo de água por hora para os períodos pré-pandêmico e pandêmico.	48
Figura 14 – Frequência do inversor e resposta da vazão medida.	49
Figura 15 – Funcionamento do Sistema.	50
Figura 16 – Arquitetura do Sistema.	51
Figura 17 – Armazenamento em um dataframe dos dados adquiridos do servidor AWS da Amazon.	52
Figura 18 – Disposição dos dados armazenados no dataframe.	53
Figura 19 – Ajuste do horário de emulação.	53
Figura 20 – Dados medidos no LENHS.	54
Figura 21 – Dados adquiridos no LENHS, com inserção de cinco <i>outliers</i>	55
Figura 22 – Dados adquiridos no LENHS, com <i>outliers</i> corrigidos.	55
Figura 23 – Acurácia do TEDA na variação do percentual de <i>outliers</i> inseridos na leitura da vazão e pressão.	57
Figura 24 – Variação do F1-SCORE do TEDA no ajuste da sensibilidade de detecção de outliers.	58
Figura 25 – Dados de entrada da RNA.	59
Figura 26 – Rede Neural Artificial utilizada.	60
Figura 27 – Resultado obtido com a aplicação da RNA para previsão da vazão normalizada.	61
Figura 28 – Resultado obtido com a aplicação da RNA para previsão da vazão desnormalizada.	62

Figura 29 – Gráfico com a previsão do consumo, considerando a presença de outliers gerados em 3 % dos dados.	63
Figura 30 – Gráfico com a previsão do consumo com a presença de outliers gerados a uma taxa de 3 %.	64
Figura 31 – Grafico do conjunto de dados com a presença de outliers gerados a uma taxa de 3 % apenas nos dados de teste.	66
Figura 32 – Grafico com a previsão do consumo com a presença de outliers gerados a uma taxa de 3 % apenas nos dados de teste.	67
Figura 33 – Grafico do TEDA aplicado na detecção de outliers a uma taxa de 3 % apenas nos dados de teste.	68
Figura 34 – Outliers substituídos após a detecção do TEDA.	69
Figura 35 – Aplicação da RNA após o uso do TEDA.	70
Figura 36 – Fluxograma Experimental do TEDA e RNA executando em tempo real.	72
Figura 37 – TEDA e RNA <i>onlines</i>	72

LISTA DE TABELAS

Tabela 1 – Especificações técnicas Mini Gerador Hidrelétrico DB-268.	34
Tabela 2 – Especificações técnicas do Transmissor de Pressão Industrial TPI-PRESS.	35
Tabela 3 – Especificações técnicas do Sensor de Fluxo de Água YF-S201.	35
Tabela 4 – Especificações técnicas da ESP32.	36
Tabela 5 – Tabela dos <i>outliers</i> interpolados.	56

LISTA DE ABREVIATURAS E SIGLAS

AWS	Amazon Web Services
ANA	Agência Nacional de Águas e Saneamento Básico
EDA	Análise Exploratória de Dados
CAGEPA	Companhia de Água e Esgotos da Paraíba
COMPESA	Companhia Pernambucana de Saneamento
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
MSE	Erro Médio Quadrático
MAPE	Erro Percentual Absoluto Médio
ETA	Estação de Tratamento de Água
UTS	Estação de Tratamento Simplificado
IA	Inteligência Artificial
IoT	Internet das Coisas
LENHS	Laboratório de Eficiência Energética e Hidráulica em Saneamento
MQTT	<i>Message Queuing Telemetry Transport</i>
RNA	Rede Neural Artificial
RN	Rio Grande do Norte
SNIS	Sistema Nacional de Informação Sobre Saneamento
TEDA	<i>Typicality and Eccentricity DataAnalytics</i>
UNICAMP	Universidade Estadual de Campinas
UFPB	Universidade Federal da Paraíba

SUMÁRIO

1	INTRODUÇÃO	14
1.1	Pertinência e motivação do trabalho	14
1.2	Objetivos	15
1.2.1	Objetivos Específicos	15
1.2.2	Contribuições	16
1.3	Organização do trabalho	16
2	TRABALHOS RELACIONADOS	18
3	FUNDAMENTAÇÃO TEÓRICA	20
3.1	Sistemas de Abastecimento de Água	20
3.1.1	Perdas e problemas enfrentados na distribuição de água	22
3.1.2	Utilização de Internet das Coisas no Monitoramento de Sistemas de Abastecimento	24
3.2	Técnicas de Detecção de Outliers	26
3.2.1	<i>Typicality and Eccentricity Data Analytics - TEDA</i>	27
3.3	Modelos de Previsão	30
3.3.1	Redes Neurais Artificiais	30
4	MATERIAIS E MÉTODOS	33
4.1	Materiais	33
4.1.1	Bancada Experimental no LENHS	33
4.1.2	Servidor AWS da Amazon	38
4.1.3	SCADA BR	39
4.2	Métodos	39
4.2.1	Métricas de Avaliação	42
4.2.1.1	Erro Médio Quadrático	42
4.2.1.2	Erro Percentual Absoluto Médio	42
4.2.2	Pré-Processamento dos Dados	42
4.2.3	Detecção e correção de <i>outliers</i>	43
4.2.4	Modelo para Previsão da Vazão	43
5	RESULTADOS	47
5.1	Descrição do Conjunto de Dados	47
5.2	Aquisição de dados e envio para a nuvem	49
5.3	Detecção de Outliers utilizando o TEDA	52

5.3.1	Aplicação do TEDA <i>offline</i>	52
5.3.2	Aplicação do TEDA <i>online</i>	56
5.4	Aplicação da RNA para previsão da vazão	58
5.4.1	Utilização da RNA Treinada com Outliers: Avaliação sem Aplicação do TEDA	62
5.4.2	Utilização da RNA Treinada sem Outliers: Avaliação do Desempenho em Dados de Teste com Outliers, sem a Aplicação do TEDA	64
5.4.3	Utilização da RNA Treinada sem Outliers: Avaliação do Desempenho em Dados de Teste com Outliers, com a Aplicação do TEDA	68
5.5	Detecção de outliers e previsão da vazão em tempo real	70
6	CONCLUSÕES	73
	REFERÊNCIAS	75

1 INTRODUÇÃO

1.1 PERTINÊNCIA E MOTIVAÇÃO DO TRABALHO

A água desempenha um papel crítico nos processos físicos, químicos e biológicos, tendo influência desde o funcionamento do corpo humano até questões climáticas a nível global (BRUNI, 1993).

Apesar de o planeta ter 71 % da sua superfície coberta por água, 97,5 % de toda a água da Terra, de acordo com a Agência Nacional de Águas e Saneamento Básico (ANA), é salgada, sendo ela imprópria tanto para o consumo como para o plantio. Apenas 2,5 % da água existente no planeta é doce, e dessa, apenas 1 % está disponível para o consumo direto, pois a maior parte encontra-se nas geleiras e armazenadas em aquíferos. No Brasil, apenas 12 % da água doce está disponível para uso, e apenas 25 % desse percentual se encontra na região Nordeste. (CASTRO, 2022).

Considerando a água como um recurso finito, pode-se interpretar esses dados como uma escassez relativa de água potável. Nessa perspectiva, o gerenciamento dos recursos hídricos torna-se uma questão crucial para garantir o bem-estar das populações e a sustentabilidade ambiental.

Historicamente, o uso dos recursos hídricos é um aspecto fundamental no desenvolvimento do homem em sociedade. Civilizações antigas, como o Egito surgiram às margens do Rio Nilo dependendo de sua água para irrigação, transporte e outras necessidades básicas (CAMINOS, 1994). Nesse cenário, o desenvolvimento de uma civilização está diretamente ligado à sua capacidade de administrar seus recursos naturais, sendo os sistemas de abastecimento de água um dos primeiros serviços implantados nos conglomerados humanos.

Com o crescimento populacional, a expansão da agricultura irrigada e o desenvolvimento industrial no Brasil, a demanda por água potável aumentou significativamente Castro (2022), elevando o consumo dos recursos hídricos disponíveis. Isso levou à necessidade de implementar políticas públicas para regular o uso dos recursos hídricos (SCHEIBE e TRINDADE, 2014), definida pela Lei 9.433/97 (Lei das Águas) (Presidência da República, 1997), que traz medidas de conservação, incluindo o racionamento de água e a redução de perdas associadas à distribuição e ao consumo excessivo e desordenado.

O Art. 2º da Lei das Águas traz como objetivo desta lei em seu parágrafo III: *"a prevenção e a defesa contra eventos hidrológicos críticos de origem natural ou decorrentes do uso inadequado dos recursos naturais."* Logo, percebe-se a necessidade de um gerenciamento

eficiente dos recursos hídricos para manutenção da vida, sendo necessário o desenvolvimento de pesquisas que visem dirimir as perdas associadas à distribuição de água para população.

Frente a esses desafios, o desenvolvimento de tecnologias, como a Internet das Coisas (*Internet of Things* - IoT) e técnicas de Inteligência Artificial (IA), oferece novas oportunidades para otimizar o gerenciamento dos recursos hídricos, devido à capacidade de processar volumes significativos de dados (ZANCAN; PASSADOR; PASSADOR, 2023). A IoT permite a coleta de dados em tempo de aquisição de sistemas de distribuição de água e infraestrutura relacionada, enquanto a IA possibilita a análise desses dados para identificar padrões, prever demandas e detectar vazamentos e outras anomalias de forma mais eficiente (PATAKA, 2021).

A aplicação dessas tecnologias pode resultar em uma gestão mais eficaz dos recursos hídricos, contribuindo para a conservação da água e a redução de desperdícios. Além disso, essas soluções podem ajudar a mitigar os impactos das mudanças climáticas, que afetam a disponibilidade e a distribuição da água em todo o mundo. Lucena (2020) afirma que a melhoria na prestação do serviço e a redução de custos passam, necessariamente, pelo aperfeiçoamento da operação dos sistemas utilizados no abastecimento.

É evidente a importância de investir em pesquisa e desenvolvimento de soluções inovadoras para o gerenciamento inteligente dos recursos hídricos. O propósito da gestão dos recursos hídricos consiste em ampliar tanto a disponibilidade quanto a qualidade da água, visando suprir as necessidades fundamentais e reduzir a demanda por usos não essenciais por meio da otimização dos processos de utilização (NUNES; TEIXEIRA, 2010).

Soluções vêm sendo desenvolvidas com a aplicação de Internet das Coisas e técnicas de Inteligência Artificial, as quais se apresentam como importantes ferramentas para o aumento da eficiência de utilização dos recursos hídricos por meio da diminuição das perdas e da melhoria dos sistemas de distribuição (VELAYUDHAN et al., 2022).

1.2 OBJETIVOS

O objetivo geral deste trabalho consiste no desenvolvimento de um sistema de medição IoT e uso de IA para o monitoramento de uma rede de distribuição de água como ferramenta de combate às perdas.

1.2.1 OBJETIVOS ESPECÍFICOS

Este trabalho tem os seguintes objetivos específicos:

- Desenvolver e validar a utilização do TEDA (*Typicality and Eccentricity Data Analytics*) para detecção e posterior correção de *outliers* em dados de pressão e vazão

coletados de uma bancada experimental de distribuição de água;

- Treinar e avaliar o desempenho de uma Rede Neural Artificial na previsão da vazão, utilizando os dados tratados após a detecção de *outliers*;
- Aplicar a RNA para realizar previsões de vazão no sistema de abastecimento, analisando o impacto da detecção de *outliers* na acurácia das previsões;
- Aplicar algoritmos de detecção de erro na coleta de dados de pressão e vazão em um sistema de distribuição de água, e realizar a previsão da demanda no ponto monitorado.

1.2.2 CONTRIBUIÇÕES

1. Otimizar o processo de aquisição de dados de um sistema de distribuição de água por meio do uso de metodologias de detecção e correção de *outliers* em grandes volumes de dados;
2. Desenvolver uma plataforma de previsão de demanda em sistemas de abastecimento de água, com a finalidade de fornecer ferramentas de tomada de decisões para os operadores dos centros de controle operacional das redes de distribuição de água.

1.3 ORGANIZAÇÃO DO TRABALHO

Este trabalho está organizado em cinco capítulos, cada um com um papel fundamental na construção e demonstração da metodologia proposta para a otimização de sistemas de distribuição de água, utilizando IoT e inteligência artificial (IA).

No primeiro tópico, foram apresentadas a introdução e os objetivos do trabalho, destacando a relevância do tema. A motivação para o desenvolvimento de um sistema de medição IoT e o uso de IA no monitoramento de redes de distribuição de água foram abordados, com a exposição dos objetivos gerais e específicos.

No segundo capítulo são apresentados trabalhos relacionados, apresentando o uso de internet das coisas e inteligência artificial no gerenciamento dos sistemas de distribuição de água.

No terceiro capítulo, realiza-se a fundamentação teórica, discutindo sistemas de abastecimento de água, perdas e desafios enfrentados na distribuição. Neste tópico, também são abordadas as possibilidades do uso de IoT no monitoramento e correção de falhas na coleta de dados. Além disso, conceitos como a detecção de *outliers* por meio da técnica TEDA são apresentados, visto que são aplicados diretamente nas análises subsequentes.

O quarto capítulo descreve a metodologia aplicada. A partir de uma visão geral da bancada experimental utilizada no LENHS, detalha-se a aplicação das técnicas de monitoramento e predição implementadas. São apresentadas as técnicas empregadas para detecção de *outliers* e a posterior utilização de redes neurais para predição de vazão.

No quinto capítulo, também são apresentados e examinados os resultados, comparando-se as variações nos parâmetros das técnicas utilizadas, especialmente TEDA e RNA, e suas influências nas previsões de vazão.

No fechamento deste trabalho, o sexto capítulo apresenta as conclusões, confrontando os resultados obtidos com os objetivos propostos. Discute-se, ainda, as potenciais aplicações futuras deste projeto, além de propostas de continuidade e expansão do estudo para novos cenários e desafios.

2 TRABALHOS RELACIONADOS

Nesta seção será apresentada uma revisão bibliográfica relacionada com o tema da pesquisa, evidenciando a relevância do tema, as principais metodologias utilizadas, e os desafios e oportunidades.

Em Byer (2005), os autores, de maneira pioneira neste campo de pesquisa, criaram e testaram um sistema de monitoramento *online* para distribuição de água. Nesse estudo, eles conseguiram detectar substâncias contaminantes presentes nos sistemas de distribuição. Realizaram experimentos práticos para monitorar parâmetros como pH, turbidez, condutividade e carbono orgânico total. Com esse monitoramento identificaram substâncias contaminantes no sistema de distribuição. Esse estudo representa uma abordagem pioneira e significativa no monitoramento em tempo real de sistemas de distribuição de água.

Mohammadpour et al. (2015) analisou a aplicação de diferentes algoritmos na detecção de problemas relacionados à qualidade da água nos sistemas de distribuição. Os autores obtiveram um erro absoluto médio (MAE - *Mean Absolute Error*) de 0,0052 na aplicação de *Support Vector Machine* (SVM) e de 0,0044 ao utilizarem redes neurais na previsão do índice proposto para mensurar a qualidade da água. Nesse estudo, os autores compararam o uso do SVM com redes neurais, encontrando resultados semelhantes para ambos os métodos. Eles concluíram que a aplicação de redes neurais e SVM otimizou o monitoramento do sistema de distribuição.

Em Zhang et al. (2017), foi comparada a aplicação de janelas temporais duplas móveis (*Dual Time-moving windows*) com os algoritmos AD e ADAM para detecção de anomalias em trezentas leituras de dados do pH da água de um sistema fluvial. Os autores observaram melhores resultados na aplicação do método *Dual Time-moving windows*, com uma redução de falsos positivos de 9,6 % e 5,3 % nos algoritmos AD e ADAM, respectivamente, para 1,3 % com o *Dual Time-moving windows*. Contudo, neste trabalho, os autores se restringiram apenas à detecção de anomalias.

Velayudhan et al. (2022) realizaram um estudo sobre a utilização de IoT no monitoramento de sistemas de distribuição. Os autores apresentaram o estado da arte com um levantamento da literatura sobre o tema e concluíram que as aplicações de monitoramento utilizando IoT não cobriam estratégias de controle em malha fechada e previsão para sistemas de distribuição de água. Além disso, os autores destacam que um sistema de distribuição de água eficiente, com alta produtividade e desempenho, pode ser obtido por meio do uso de sistemas IoT automatizados integrados à análise preditiva, o que está alinhado à proposta deste trabalho.

Ebisi et al. (2023) afirma que o uso de IoT e IA integram as estratégias mais promissoras na detecção de vazamentos de água na infraestrutura em escala industrial e em cidades inteligentes. Os autores aplicaram técnicas de inteligência artificial para detecção de vazamentos em sistemas de distribuição.

Saritha et al. (2023) propõem um mecanismo baseado em IoT para monitoramento da qualidade da água. Os autores demonstram que a utilização de IoT é capaz de monitorar em tempo real o PH, a temperatura da água. Garg et al. (2023) realizam o monitoramento da qualidade da água por meio do emprego de IoT e *Machine Learning*. Os autores aplicam um mecanismo de monitoramento das propriedades da água e previsão se a água está segura para o consumo. Os autores chegaram a uma acurácia de 96,12 % na detecção da qualidade da água.

Em Puviyarasi et al. (2024) é apresentada uma estratégia de monitoramento da qualidade da água. Para isso, os autores utilizam um microcontrolador ESP32 para aquisição dos dados e envio para um servidor remoto, o *Firebase*. No trabalho desenvolvido, os autores utilizaram sensores para medir a turbidez, temperatura e pH da água. O sistema desenvolvido demonstrou eficiência ao monitorar a qualidade da água em diferentes cenários, alertando os operadores sempre que os valores de qualidade ultrapassavam os limiares. O *hardware* utilizado foi projetado para ser de baixo custo, consumindo pouca energia. Contudo, os autores não realizam previsões com os dados adquiridos. O alerta gerado no monitoramento consistiu apenas na indicação de que os valores adquiridos estão fora dos parâmetros configurados no sistema proposto.

Após a revisão dos trabalhos relacionados, observou-se que as pesquisas têm mostrado resultados satisfatórios na aplicação de IoT e técnicas de inteligência artificial na detecção de anomalias no monitoramento de sistemas de distribuição. Contudo, para que uma rede neural possa realizar previsões satisfatórias, é necessário que ela reconheça previamente que um dado é uma anomalia, ou que o dado seja identificado como tal antes de ser aplicado na rede neural.

De acordo com a revisão bibliográfica, pode-se destacar a relevância do uso de sistemas de monitoramento e aplicação de técnicas de IA em redes de abastecimento de água, com a finalidade de melhorar o serviço prestado à população e garantir a qualidade de água. Desta forma, este trabalho de dissertação se fundamenta na aplicação de técnicas de coleta de dados utilizando ferramentas de IoT, pré-processamento dos dados para a geração de bancos de dados consolidados, e uso de metodologias de previsão de demanda de água baseadas em inteligência artificial.

3 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo são abordados conceitos teóricos necessários para o entendimento deste trabalho. Inicialmente, são apresentadas as principais perdas envolvidas nos sistemas de abastecimento de água e as estratégias utilizadas como solução do problema. A utilização de Internet das Coisas, a correção dos erros encontrados na coleta de dados, as técnicas utilizadas para detecção de *outliers*, e os modelos de previsão baseados em Redes Neurais Artificiais também serão abordados nesse capítulo.

3.1 SISTEMAS DE ABASTECIMENTO DE ÁGUA

Orellana et al. (2018) definem sistemas de abastecimento de água como um conjunto de infraestruturas, equipamentos e instalações com a finalidade de prover água, atendendo a requisitos técnicos de quantidade e qualidade adequadas ao consumo humano. Para que a água potável chegue até as residências é preciso que haja uma infraestrutura capaz de bombear a água captada nos mananciais até os consumidores. Contudo, o custo da energia com o bombeamento é um dos fatores que mais afeta o custo operacional dos sistemas de abastecimento (LUCENA, 2020), resultando diretamente no aumento da fatura paga pelo consumidor final.

Na Figura 1, observa-se a estrutura de um sistema de abastecimento. A água bruta é captada nos mananciais hídricos e conduzida para o tratamento por meio do sistema de adução, o qual consiste em uma tubulação que liga a captação da água bruta à Estação de Tratamento de Água (ETA) ou à Estação de Tratamento Simplificado (ETS).

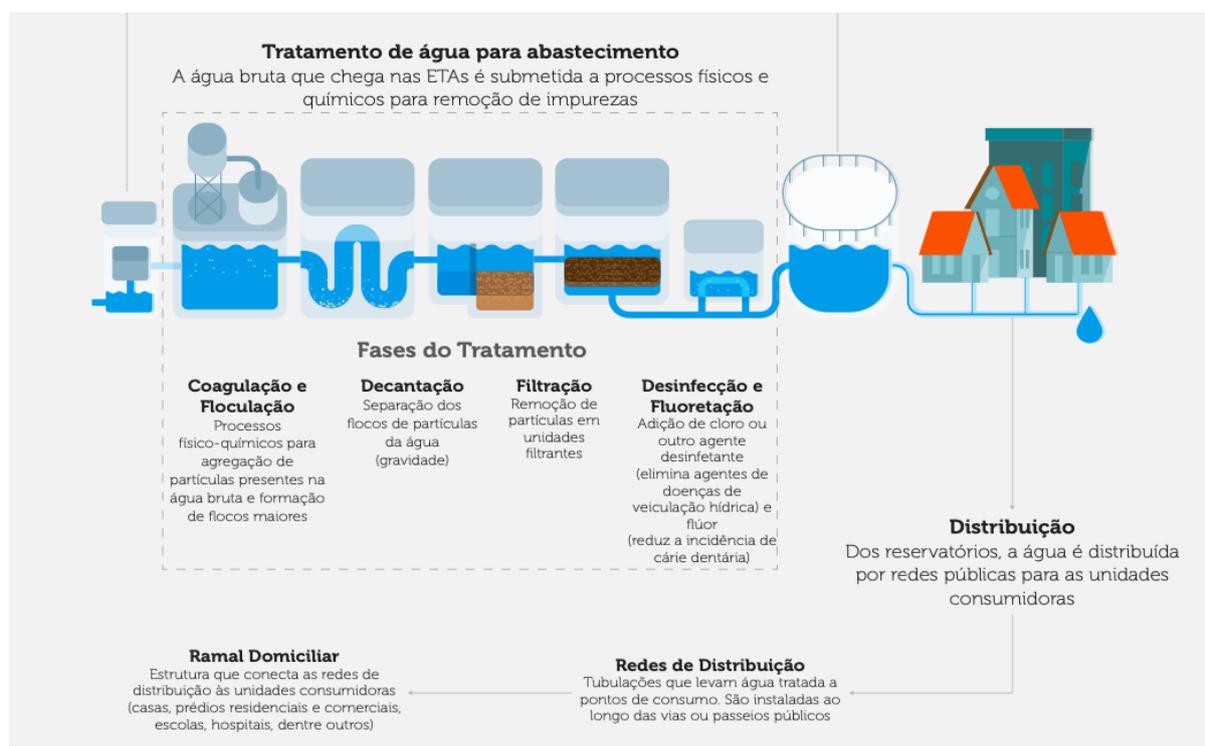
Após a captação, a água bruta passa pela coagulação/floculação, decantação, filtração e também pela desinfecção/fluoretação. Essas etapas consistem de processos físicos e químicos aplicados à água bruta para remoção de impurezas.

Os sistemas de abastecimento de água acompanham o desenvolvimento das cidades, o crescimento da população e a sua distribuição geográfica, resultando assim em sistemas de tubulações com características heterogêneas: diferentes diâmetros, idades de instalação, entre outras. Isso também traz diversos desafios na operação e manutenção, na garantia dos requisitos técnicos mínimos de consumo (TUCCI, 2008).

Orellana et al. (2018) ainda aponta que o processo de envelhecimento dos sistemas de fornecimento de água é inevitável. Conforme os diversos componentes desses sistemas atingem o fim de sua vida útil, a probabilidade de ocorrência de vazamentos aumenta. Além disso, rupturas e interrupções no fornecimento tornam-se mais frequentes. Muitas

tubulações acabam acumulando incrustações, resultando em um nível elevado de aspereza interna. Isso, por sua vez, leva à redução do diâmetro hidráulico, diminuindo as pressões, por vezes abaixo dos valores recomendados por normas técnicas. Conseqüentemente, ocorre um aumento significativo no consumo de energia elétrica em sistemas pressurizados por bombas.

Figura 1 – Sistema de Abastecimento de Água.



Fonte: (BRASIL, 2023)

Neste cenário, é fundamental o desenvolvimento de pesquisas que visem à redução dos custos relacionados à operação dos sistemas de abastecimento, assim como a redução das perdas.

Em síntese, sistemas comuns de abastecimento de água consistem na utilização de bombas hidráulicas para captação da água em poços ou estações de tratamento e encaminhamento para reservatórios distribuídos nos centros urbanos. Esses sistemas de abastecimento representam um dos principais componentes da infraestrutura de um município, tendo a função de fornecer água em quantidade e qualidade adequadas para a população (MIQUELIN; SOLER, 2016).

Estima-se que, no território brasileiro, seja produzido um volume de água de 17,9 bilhões de m³/ano, o qual refere-se à quantidade que os prestadores de serviço disponibilizam para consumo, sendo medida ou estimada na saída da Estação de Tratamento de Água

(ETA). Contudo, apenas 11,6 bilhões de m³/ano de água chega ao consumidor. Na região Nordeste, o percentual de perda chega a 46,7 % (BRASIL, 2023).

O estado da Paraíba tem grande parte do seu território localizado na região semi-árida do Nordeste brasileiro. A economia está baseada na agricultura, pecuária, indústria e turismo (LUNGUINHO, 2007). A Paraíba, assim como os demais estados que fazem parte da Região Nordeste, sofre com escassez de recursos hídricos, repercutindo diretamente no processo de desenvolvimento local.

O trabalho desenvolvido por Lucena e Nascimento (2019) apresentou um diagnóstico da operação de parte do sistema de abastecimento de água de Campina Grande, assim como apresentou as principais limitações e potencialidades da infraestrutura do sistema. Por meio da metodologia proposta, os autores afirmam que o custo da energia com o bombeamento é um dos fatores que mais afeta o custo operacional dos sistemas de abastecimento no Estado da Paraíba.

Pereira e Cunico (2020) identificou e avaliou as condições de esgotamento sanitário e de abastecimento de água nos diferentes graus de vulnerabilidade socioambiental no município de João Pessoa. Por meio do monitoramento proposto, os autores relacionaram o problema da distribuição de água e a coleta de esgoto em áreas de vulnerabilidade socioambiental.

3.1.1 PERDAS E PROBLEMAS ENFRENTADOS NA DISTRIBUIÇÃO DE ÁGUA

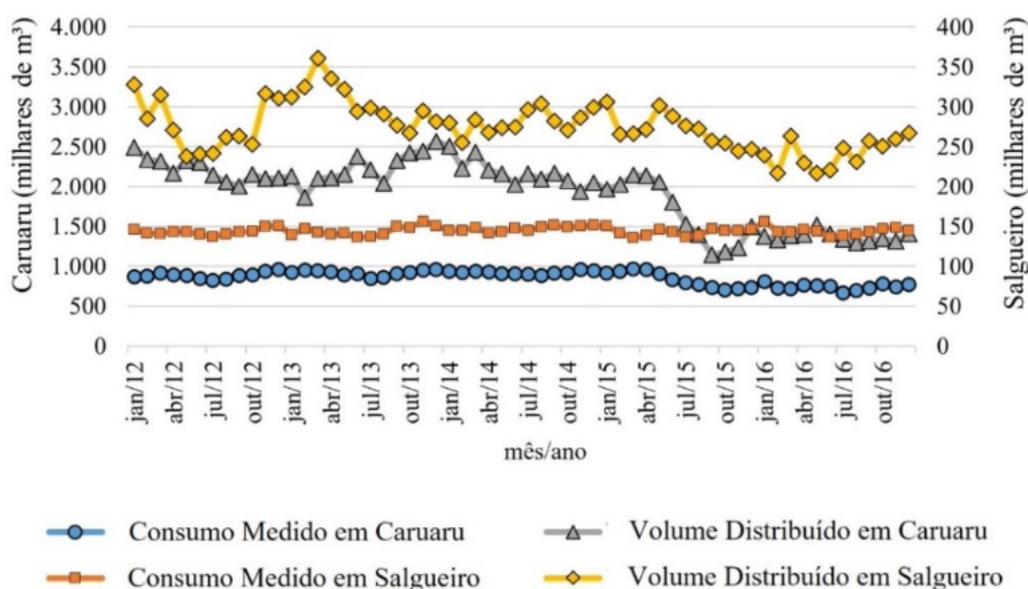
Apesar de as perdas serem inerentes a qualquer sistema de abastecimento de água, esse assunto tem recebido atenção nos últimos anos devido à escassez hídrica e ao risco de contaminação da água tratada.

O Diagnóstico dos Serviços de Água e Esgoto publicado pelo SNIS (BRASIL, 2019) apresenta relatórios das perdas dos sistemas de distribuição por ano nas diferentes regiões do Brasil. O SNIS-NE 2022 aponta que no território brasileiro, 37,8 % da água tratada não chega às residências. Em termos quantitativos, a cada cem litros de água que sai das estações elevatórias, apenas 62,2 litros são contabilizados pelos consumidores (BRASIL, 2023).

No estudo conduzido por Figueiredo et al. (2023), que analisou o sistema de distribuição de água das cidades de Caruaru e Salgueiro, no Estado de Pernambuco, foi constatada uma perda aproximada de 50 % da água tratada pela consencionária local entre os anos de 2012 e 2016. Os autores estimaram um volume de água perdido na ordem de 72,25 milhões de metros cúbicos. As perdas podem ser classificadas em reais ou aparentes, também chamadas de perdas comerciais.

Na Figura 2 apresenta-se uma distribuição do volume de água distribuído nas cidades de Caruaru e Salgueiro, assim como o consumo medido nas residências das duas cidades. Observa-se, pela diferença das séries temporais apresentadas o volume de água perdido entre janeiro de 2012 e outubro de 2016.

Figura 2 – Capacidade distribuída e consumo medido entre os anos de 2012 e 2016 nas cidades de Caruaru e Salgueiro, em Pernambuco, Brasil.



Fonte: (FIGUEIREDO et al., 2023)

Frente a essas perdas, uma das medidas de enfrentamento à escassez de água é o fornecimento intermitente (GALAITSI et al., 2016), no qual o sistema de distribuição de água funciona por períodos limitados. Galaitsi et al. (2016) ainda afirmam que essa medida pode ser tomada por diversos fatores, como a insuficiência dos recursos hídricos disponíveis ou altas perdas de água na distribuição.

Nos últimos anos, têm crescido os investimentos para redução das perdas, sejam elas reais ou comerciais (DUTRA; OLIVEIRA, 2017). Nesse contexto, Figueiredo et al. (2023) propôs alternativas de combate a este cenário, com o objetivo de mitigar os efeitos da escassez hídrica e do abastecimento intermitente. Os autores relatam o problema enfrentado pela população de Caruaru e Salgueiro no ano de 2012, quando a Companhia Pernambucana de Saneamento (COMPESA) realizou rodízio no abastecimento de até três dias, chegando a 21 dias sem o fornecimento de água potável para população.

Outra medida de enfrentamento às perdas é o uso da automação, apontada por Figueiredo et al. (2023) como um caminho para solução do problema. Essa solução é defendida pelo autor juntamente com a setorização das cidades em distritos de medição

e controle. Em seu trabalho, Figueiredo et al. (2023) ainda afirma que o monitoramento e controle da pressão na rede é uma importante estratégia para redução do consumo de energia utilizada no bombeamento.

Frente a este cenário, o uso de Internet das Coisas vem sendo utilizada como uma ferramenta eficaz de monitoramento dos sistemas de abastecimento.

3.1.2 UTILIZAÇÃO DE INTERNET DAS COISAS NO MONITORAMENTO DE SISTEMAS DE ABASTECIMENTO

A Internet das Coisas (IoT) recebe esse nome devido a concepção de interconectar todas as coisas, possibilitando acesso aos dispositivos conectados, assim como a possibilidade de geração de dados (SOUSA; PROFA; BORIN, 2019). Pataca (2021) define a Internet das Coisas como a ampla capacidade de diferentes dispositivos se comunicarem quando conectados à internet. Sua aplicação tem um papel importante na gestão integrada de serviços essenciais tais como fornecimento de água, gás e energia.

Além da necessidade de monitoramento na gestão dos recursos hídricos, o uso de IoT se torna uma importante ferramenta devido à sua capacidade de geração de dados, os quais são necessários para o desenvolvimento de estudos ambientais de planejamento e gestão de recursos hídricos (ZUFFO, 2011).

O campo de aplicação da IoT é amplo, como ilustrado na Figura 3, abrangendo desde aplicações domésticas, como cozinhas e salas inteligentes, até contextos ambientais e situações de emergência ou crise.

Figura 3 – Aplicações da Internet das Coisas.



Fonte: (PATAKA, 2021) adaptado.

Sousa, Profa e Borin (2019) apresentam o uso de IoT como uma solução para aquisição eficaz de dados aplicado ao monitoramento inteligente de sistemas de distribuição de água. Os autores propuseram um projeto de integração na Universidade Estadual de Campinas - UNICAMP - campus Barão Geraldo, com a proposta de desenvolver um sistema capaz de aumentar a eficiência energética e de saneamento. O projeto partiu da necessidade de desenvolver um sistema de gerenciamento da distribuição de água capaz de informar os locais em que ocorre variação no consumo normal, caracterizando assim um possível vazamento. Atualmente, essa identificação ocorre ou por usuários que informam

vazamentos visíveis ou pela medição mensal do consumo, que apenas direciona a região do vazamento, havendo ainda a necessidade de detectar o local específico. Este problema é real e enfrentado pelas concessionárias, devido à falta de automação no monitoramento dos sistemas de distribuição.

A ausência de dados de monitoramento de pressão e vazão na rede de distribuição, somada aos desafios mencionados no Item 3.1.1, como as perdas aparentes, que incluem furtos de água, erros de medição e falhas administrativas no sistema de faturamento, justifica a necessidade de desenvolver soluções de monitoramento e automação que possam mitigar esses problemas.

Neste contexto, Velayudhan et al. (2022) destaca o uso de Internet das Coisas juntamente com a aplicação de Inteligência Artificial como as tecnologias mais proeminentes para solucionar esse cenário, devido à capacidade de medir, processar e analisar os atributos e eventos na rede de abastecimento.

Essas tecnologias foram aplicadas na cidade de Fontain Valley, California, nos Estados Unidos. A SENSUS utilizou inteligência artificial e IoT como ferramentas de gestão dos recursos hídricos na cidade. Como resultados, a empresa conseguiu reduzir em 23 % as perdas no fornecimento de água (SOUSA; PROFA; BORIN, 2019).

Apesar de IoT ser uma ferramenta para solução do problema enfrentado no monitoramento, é possível que os dados adquiridos pelos sensores apresentem erros de medição, ocasionados por fatores diversos. Frente a isso, faz-se necessário o emprego de algoritmos que identifiquem esses erros de aquisição, os quais são abordados na Seção 3.2.

3.2 TÉCNICAS DE DETECÇÃO DE OUTLIERS

Falhas na leitura dos sensores de monitoramento podem estar presentes em um conjunto de dados, os quais são consideradas erros, que se não tratados, ou ignorados, podem comprometer a análise de tendências (PETERSON; WESTERN; CHENG, 2018). No estudo realizado por Figueiredo et al. (2023), os autores excluíram os dados que apresentavam inconsistências, como valores negativos, ou desvio dos dados na comparação com parâmetros de referência ou com as informações históricas da COMPESA, deixando explícita a presença de *outliers* nos dados medidos.

Nesse contexto, é possível a aplicação de algoritmos que detectem *outliers* na leitura dos registros medidos. Ventura et al. (2017) em seu estudo apresenta métodos de tratamento de dados meteorológicos que trabalham com detecção de *outliers*. O autor cita as seguintes técnicas que podem ser utilizadas na detecção de erros em um conjunto de dados:

- Z-Score: O método Z-score é uma métrica que representa o comportamento anormal de itens em termos de sua distância em relação à média e ao desvio padrão de uma coleção de dados. Estimar o Z-score consiste em normalizar os dados, de modo que a média seja 0 e o desvio padrão seja 1 (VENKATAANUSHA et al., 2019).
- Boxplot: Também conhecida como diagrama de caixa, é uma ferramenta gráfica utilizada na análise estatística para representar a distribuição de um conjunto de dados e identificar possíveis *outliers*. Boxplots estão entre as ferramentas de análise exploratória de dados (EDA) mais amplamente utilizadas na prática estatística (DOVOEDO; CHAKRABORTI, 2014).

Silva e Nóbrega (2021) utilizaram o Boxplot e o Z-Score para avaliar a presença de dados anômalos no comportamento da precipitação pluvial acumulada no Semiárido do estado do Rio Grande do Norte.

Além das técnicas citadas, técnicas de inteligência artificial também podem ser usadas para detecção de *outliers*, como, por exemplo, a técnica de *clustering* que procura e agrupa dados de maneira a categorizar dados que possuem a mesma similaridade. Devido à sua funcionalidade, este algoritmo pode ser utilizado para detectar anomalias em um conjunto de dados (SALVADOR; CUNHA; CORRÊA, 2009).

Outra técnica para detecção de *outliers* é a *Typicality and Eccentricity Data Analytics*, a qual tem como vantagem a capacidade de identificar *outliers* em grandes volumes de dados do tipo *streaming*. Esta técnica será abordada na Subseção 3.2.1.

3.2.1 TYPICALITY AND ECCENTRICITY DATA ANALYTICS - TEDA

Uma estrutura de análise de dados baseada na tipicidade e excentricidade foi introduzida por Angelov (2013). O TEDA é uma técnica que exige um baixo esforço computacional, inserida no repertório de análise de dados, destinada, principalmente à detecção de *outliers*.

Inicialmente proposta por Angelov (2013), a técnica visa identificar anomalias em fluxos de dados. O algoritmo baseia-se em dois conceitos fundamentais: tipicidade e excentricidade. A tipicidade representa a semelhança de uma amostra com o restante do conjunto de dados, enquanto a excentricidade expressa a diferença da amostra em relação ao mesmo conjunto.

Essa técnica, por ser simples de implementar, pode ser implementada em diversas plataformas, incluindo microcontroladores, tendo como foco principal a aplicação em um fluxos de dados.

O TEDA utiliza diferentes métricas de distância, como a Euclidiana, Mahalanobis e a Distância Cosseno, para analisar tanto a tipicidade quanto a excentricidade dos dados. Ao

receber uma nova entrada, o TEDA calcula a tipicidade em relação aos grupos existentes, facilitando sua classificação. Dessa forma, o TEDA não se limita apenas à detecção de *outliers*, mas também pode ser aplicado em processos de clusterização e classificação.

Uma das vantagens do algoritmo é sua capacidade de ser utilizado sem a necessidade de treinamento prévio, podendo ser utilizado tanto *online* quanto *offline*. A partir do momento em que os dados são recebidos, o algoritmo evolui com a classificação, exigindo um baixo esforço computacional e tornando-se uma opção viável para sistemas embarcados.

O TEDA destaca-se por seu desempenho quando considerada a relação custo-benefício. Suas aplicações abrangem diversas áreas, como veículos, compressão de dados, processamento de áudio e aplicações industriais. Comparado aos métodos estatísticos convencionais, o TEDA se destaca por não exigir conhecimento prévio dos dados. Sua utilidade é evidente em séries temporais, pois dispensa o conhecimento matemático do modelo e a compreensão da distribuição dos dados, necessários nos métodos apresentados na Seção 3.2, tornando-se uma ferramenta valiosa para casos do mundo real, nos quais as distribuições nem sempre seguem padrões gaussianos.

O algoritmo TEDA mensura a tipicidade e excentricidade para cada nova amostra, utilizando diversas distâncias, como euclidiana, cosseno ou Mahalanobis. Com um baixo esforço computacional, o algoritmo destaca-se por sua rapidez e eficácia (ANGELOV, 2013).

Ao ser considerado o *datastream* como um vetor, a soma das distâncias euclidianas para cada um dos elementos é calculada. A Equação 3.1 apresenta a soma das distâncias de uma amostra específica (x) para cada um dos outros k elementos.

$$\pi_k(x) = \sum_{i=1}^k d(x, x_i) \quad (3.1)$$

Dessa maneira, a excentricidade da amostra de dados x , no instante de tempo k , pode ser dada pela equação, para $k > 2$ e o denominador > 0 :

$$\xi_k = \frac{2\pi_k(x)}{\sum_{i=1}^k \pi_k(x_i)} \quad (3.2)$$

A equação pode ser simplificada para que a excentricidade pudesse ser calculada de forma recursiva, não precisando calcular todas as distâncias para cada nova amostra. Caso contrário, precisaria armazenar todas as amostras na memória para atualização do cálculo da distância de cada amostra em relação às demais. Para o cálculo da excentricidade, faz-se necessário calcular a média e a variância do conjunto de amostras recebidas.

A média de x no instante k é dada pela Equação 3.3, a qual usa a média anterior calculada. Para a primeira amostra considera-se a média como sendo o primeiro valor

recebido.

$$\mu_k = \frac{k-1}{k}\mu_{k-1} + \frac{1}{k}x_k, \text{ sendo } \mu_1 = x_1 \quad (3.3)$$

A variância é calculada mediante a aplicação da Equação 3.4. A variância assume o valor igual a zero para o primeiro dado lido.

$$\sigma_k^2 = \sum_{i=1}^k \frac{(x_i - \mu_k)^T (x_i - \mu_k)}{k} \quad (3.4)$$

A Equação 3.5 calcula a excentricidade no instante de tempo k :

$$\xi_k = \frac{1}{k} + \frac{(\mu_k - x_k)^T (\mu_k - x_k)}{k\sigma_k^2} \quad (3.5)$$

Observa-se que para o cálculo da excentricidade basta possuir os valores da média e variância do conjunto dos dados, os quais na aplicação computacional são atualizados recursivamente a cada novo dado recebido. A tipicidade pode ser calculada como o complemento da excentricidade.

Após o cálculo, a excentricidade é normalizada, garantindo que a soma de todos os termos de excentricidade seja igual a 1.

Outros métodos comuns utilizam o desvio padrão das amostras para o cálculo de *outliers*, contudo, essa abordagem é restrita a distribuições gaussianas (COSTA et al., 2015). Diferentemente, o TEDA oferece outra abordagem para o cálculo dos *outliers* que não precisa que a distribuição das amostras assemelhe-se à distribuição normal.

Utilizando-se a inequação Chebyshev, Equação 3.6, em que a probabilidade de uma amostra é menor ou igual a $\frac{1}{m^2}$, em que X é o dado de entrada, μ é a média, σ é o desvio padrão e m é um valor positivo maior que 1:

$$Pr(\|X - \mu\| \geq m\sigma) \leq \frac{1}{m^2} \quad (3.6)$$

Utilizando o trabalho de Bernieri, Betta e Liguori (1996), Angelov (2013) adaptou o uso da excentricidade normalizada para a expressão definida na Equação 3.7, a qual pode ser utilizada para caracterizar se um dado é ou não um *outlier*.

$$\zeta > \frac{m^2 + 1}{2k} \quad (3.7)$$

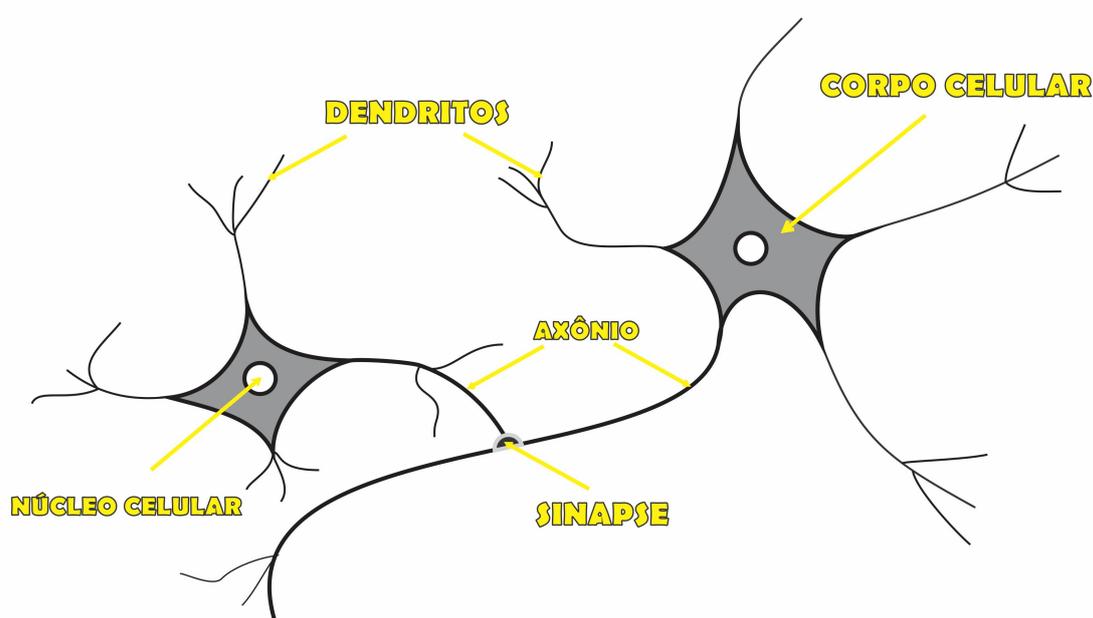
Uma vez detectado um dado anômalo, ele pode ser excluído ou substituído por outro utilizando uma técnica de correção de dados.

3.3 MODELOS DE PREVISÃO

3.3.1 REDES NEURAIS ARTIFICIAIS

O cérebro humano é responsável pelo desempenho de diversas funções vitais, sendo o centro de comando dos mecanismos de locomoção e processamento dos sinais sensoriais. Tal processamento é devido a ativação sináptica que ocorre entre células neuronais (Figura 4). O pulso elétrico conduzido no neurônio é propagado quimicamente na fenda sináptica, o que possibilita a intercomunicação neural, e a interligação em rede entre bilhões de neurônios .

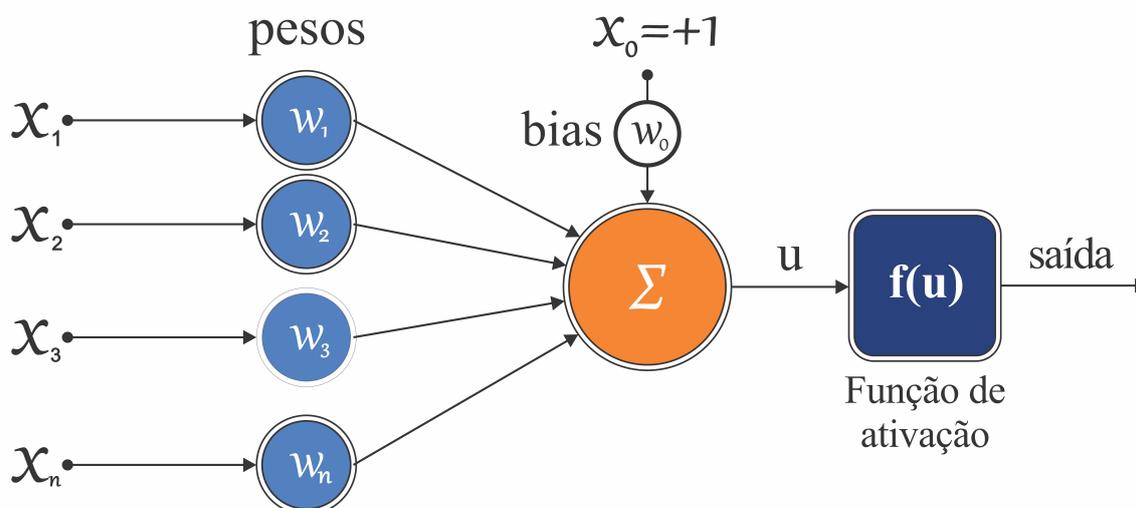
Figura 4 – Representação de neurônios biológicos.



Fonte: (NETO; BONINI, 2010) adaptado.

Matematicamente, uma célula neural pode ser simulada (Figura 5) a partir da soma ou subtração dos sinais de entrada e da aplicação de uma função de ativação, que retira a linearidade da operação. Essa remoção da linearidade é essencial, pois permite que a rede neural modele relacionamentos complexos nos dados. Quando fala-se sobre a linearidade, refere-se à relação entre a soma ponderada dos sinais de entrada e a saída da rede. A função de ativação, ao introduzir não-linearidades, possibilita que a rede aprenda e represente padrões mais sofisticados, transformando a combinação linear dos sinais em uma saída que pode capturar a complexidade do comportamento do sistema em análise.

Figura 5 – Modelo artificial de neurônio biológico.



Fonte: (NETO; BONINI, 2010) adaptado.

A exploração da atividade neural e a aplicação de sistemas computacionais que simulem seu processamento é um campo vasto de pesquisa que possibilita ao homem um passo importante em sua evolução, pois otimiza os recursos computacionais na resolução de questões complexas, não resolvidas analiticamente, como trabalho desenvolvido por Liimatainen et al. (2021) que utilizaram redes neurais convolucionais para classificação de padrões de localização de proteínas.

Kriegeskorte e Golan (2019) apresentam em seu estudo que os trabalhos com redes neurais artificiais são inspirados no funcionamento de redes neurais biológicas, mediante a implementação de unidades neurais modeladas para combinar múltiplas entradas e produzir uma saída única. Nesse processo, comumente a célula realiza a soma das entradas transformando sua resposta em uma saída não linear.

A rede neural artificial é composta pela camada de entrada, camada oculta e pela camada de saída. Seu treinamento se dá mediante o ajuste dos pesos sinápticos entre neurônios artificiais (KRIEGESKORTE; GOLAN, 2019).

A otimização da rede neural se dá por meio do ajuste da função de custo, a qual pode ser otimizada pelo gradiente descendente. A partir do balanceamento dos pesos sinápticos é possível configurar a rede para aquisição da resposta desejada. Chen, Wang e Lai (2020) utilizaram esse princípio para classificação e reconhecimento de imagens. Para isso, os pesquisadores implementaram uma rede neural profunda, com camadas internas para a extração das características.

Os otimizadores, assim como o gradiente descendente, são usados para facilitar o encontro do mínimo global na função de custo, contudo com implementações que visam à convergência de forma mais rápida (LI et al., 2021). São características dos otimizadores:

- Acelerar a convergência do algoritmo e para que se consiga chegar ao mínimo global da função de custo o mais rápido possível, de forma a treinar o mais rápido;
- Prevenir ou pular mínimos locais. Algumas funções de custo geram respostas com mínimos locais, fazendo com que o gradiente descendente fique preso em um desses mínimos. Os otimizadores fazem ajustes para que essa situação não ocorra;
- Simplificação da definição manual de parâmetros, principalmente a taxa de aprendizagem. Os otimizadores controlam a taxa de aprendizagem, otimizando-a, sem que o usuário precise defini-la de forma manual.

Os principais otimizadores usados atualmente em redes neurais são: Momentum Optimizer, Nesterov, AdaDelta, RMSProp, Adam, AdaMax, Nadam. Entre esses, o otimizador Adam se destaca por suas características herdadas do AdaGrad e RMSProp. O Adam tem um taxa de aprendizagem adaptativa (KINGMA; BA, 2014). Isso significa que diferentes partes da rede neural podem possuir diferentes taxas de aprendizagem, sendo tal controle difícil de ser efetuado manualmente. Essa característica acelera a convergência da função de custo, e apresenta uma alta performance em relação aos outros otimizadores.

4 MATERIAIS E MÉTODOS

Neste capítulo é apresentada a metodologia e os materiais que foram usados no trabalho para atingir os objetivos propostos.

4.1 MATERIAIS

Nesta seção são apresentados os materiais e métodos utilizados na pesquisa. Este trabalho, que conta com a parceria institucional do CNPq, Amazon AWS, CAGEPA e UFPB, está integrado aos resultados de um grupo de pesquisa que visa aplicar ferramentas de inteligência artificial e computação em nuvem utilizado sobre uma rede de abastecimento de água no Laboratório de Eficiência Energética e Hidráulica em Saneamento da Universidade Federal da Paraíba (LENHS/UFPB).

4.1.1 BANCADA EXPERIMENTAL NO LENHS

O LENHS conta com uma rede de tubulação pressurizada por uma bomba centrífuga, acionada por um motor de indução trifásico, conforme ilustrado na Figura 6. Esse conjunto é responsável por pressurizar a rede, que simula um sistema de distribuição de água.

Figura 6 – Conjunto Motor Bomba.



Fonte: (FLORES, 2021).

O acionamento da bomba é automatizado por meio da utilização de um inversor de frequência gerenciado por um software computacional, LabVIEWTM - *Laboratory Virtual*

Instruments Engineering Workbench, o qual atua como *software* supervisor (FLORES, 2021).

A bancada experimental disponível no LENHS, a qual foi utilizada neste trabalho, é composta por um conjunto de sensores integrados a um mini gerador hidrelétrico modelo DB-268, cujas especificações estão detalhadas na Tabela 1. Além disso, inclui um transmissor de pressão industrial, modelo TPI-PRESS, cujas características também podem ser encontradas na Tabela 2. Para medir o fluxo de água, é utilizado um sensor de fluxo YF-S201, com suas especificações apresentadas na Tabela 3.

Todos esses componentes estão conectados a um microcontrolador ESP32, cujas características são descritas na Tabela 4. A escolha do microcontrolador ESP32 se fundamenta nas suas características de conectividade integrada, baixo consumo energético, capacidade de processamento, custo e acessibilidade. Para esta aplicação, a ESP32 se mostra mais vantajosa que a *Raspberry Pi* devido esta necessitar maior consumo energético. A ESP32 também apresenta vantagens frente ao Arduino na aplicação proposta, devido a sua capacidade de processamento e conectividade (SOUSA, 2024). O microcontrolador desempenha funções essenciais, como a coleta de dados dos sensores, o armazenamento temporário dessas informações e o envio dos dados coletados para um servidor web, que está hospedado na plataforma Amazon.

O mini gerador hidroelétrico e os sensores estão ligados a uma tubulação de meia polegada que tem o fluxo de água controlado por um sistema supervisor.

Tabela 1 – Especificações técnicas Mini Gerador Hidrelétrico DB-268.

Característica	Especificação
Modelo	DB-268
Tensão de Saída	12VDC
Corrente de Saída	0 - 150mA
Potência	1,8W
Resistência do Fio	$10,5 \pm 0,5\Omega$
Resistência de Isolamento	$10M\Omega$
Velocidade de Rotação	2000 - 3000 RPM
Faixa de Fluxo	2,5 - 25L/min
Pressão Máxima Fechada	0,6Mpa
Pressão Máxima Aberta	1,2Mpa
Folga Axial	0,2 - 1,0mm
Diâmetro da Rosca	1/2" (G $\frac{1}{2}$)
Temperatura de Operação	0 - 80°C
Dimensões	84 x 48,50 x 29,50mm
Peso	79g

Tabela 2 – Especificações técnicas do Transmissor de Pressão Industrial TPI-PRESS.

Característica	Especificação
Material do Invólucro	Aço Inoxidável AISI 316L
Sensor Piezoresistivo	Aço Inoxidável AISI 316L
Material do Corpo	Aço Inoxidável 304L (316L opcional)
Faixa de medição	0 a 40 mca
Grau de Proteção	IP-65 / IP-66
Conexão	1/2 , 1/4 NPT ou BSP
Unidades de Engenharia	bar, kgf/cm ² , psi, kPA, mPA, mH ₂ O, mmH ₂ O, mmHg
Conector Elétrico	DIN 43650
Alimentação	12...30
Sinal de Saída	4...20 mA
Precisão	0,25 % ou 0,5 % F.E
Tempo de Resposta	≤ 1 milissegundo
Proteção	Contra inversão de polaridade e contra rádio frequência
Temperatura de Contato com Fluido	-20 a 100°C

Tabela 3 – Especificações técnicas do Sensor de Fluxo de Água YF-S201.

Característica	Especificação
Modelo	YF-S201
Tensão de Operação	4,5 a 18V
Corrente Máxima	15mA (5V)
Faixa de Fluxo (Vazão)	1 - 30L/min
Pressão Máxima	1.75MPa
Pulsos por Litro	450
Temperatura de Trabalho	-25 a 80°C
Exatidão	10 %
Comprimento do Cabo	15cm
Conexão	1/2"

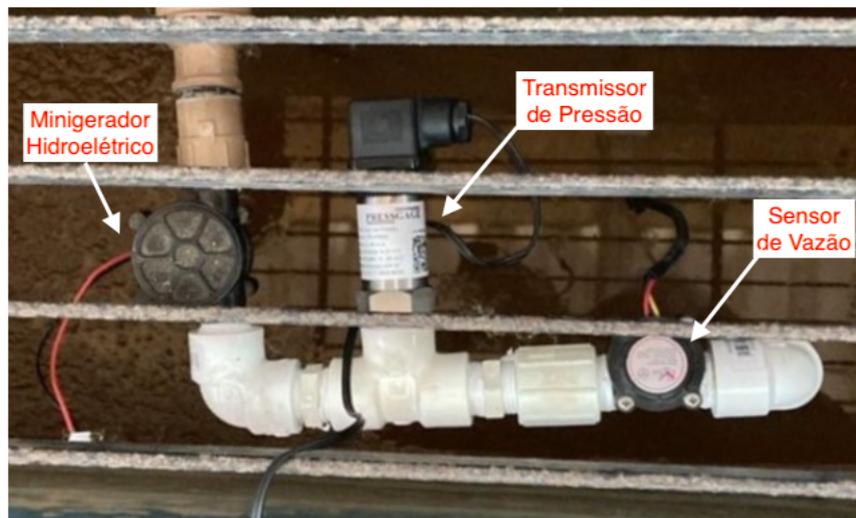
Tabela 4 – Especificações técnicas da ESP32.

Característica	Especificação
Alimentação	2,2V 3,3V DC
Corrente de Consumo	Média de 80mA
Temperatura de Operação	-40°C +85°C
Processador	Xtensa® Dual-Core 32-bit LX6
Processador Secundário	ULP 8MH com consumo de 150uA
Frequência de Operação	80MHz 240MHz
Memória FLASH	4MB
Memória RAM	520KB
Memória ROM	448KB
GPIOs (Entradas e Saídas)	34 GPIOs de 3.3V e 12mA
Conversores ADC	18 ADC com 12-bit de resolução (4096 bits)
Conversores DAC	2 ADC com 8-bit de resolução (256 bits)
WiFi	2,4 GHz, 802.11 b/g/n/e/i (802.11n até 150 Mbps)
Bluetooth	Bluetooth Low Energy v4.2 (BLE)
Criptografia	AES, RSA, SHA e ECC
Segurança	WPA/WPA2/WPA2-Enterprise/WPS
Protocolos de Rede	IPv4, IPv6, SSL, TCP/UDP/HTTP/FTP/MQTT
Temporizadores	4 Timers de 64-bit
Interfaces de Módulos	Cartão SD, UART, SPI, SDIO, I2C, LED PWM

A placa de circuito impresso com a ESP32, assim como o circuito de alimentação da placa utilizada na aquisição dos dados foi desenvolvida por Oliveira (2023). O autor construiu um sistema capaz de medir a tensão da bateria de alimentação e os dados de pressão d'água e vazão d'água na tubulação na qual os sensores estão conectados, possibilitando assim criar uma base de dados extensa o suficiente para a aplicação de algoritmos de inteligência artificial, que está entre os objetos desta pesquisa.

Na Figura 7 é apresentado o *setup* de leitura dos dados. A água que escoar pela tubulação passa pelo minigerador, que alimenta o circuito encapsulado e a bateria.

Figura 7 – Bancada Experimental.

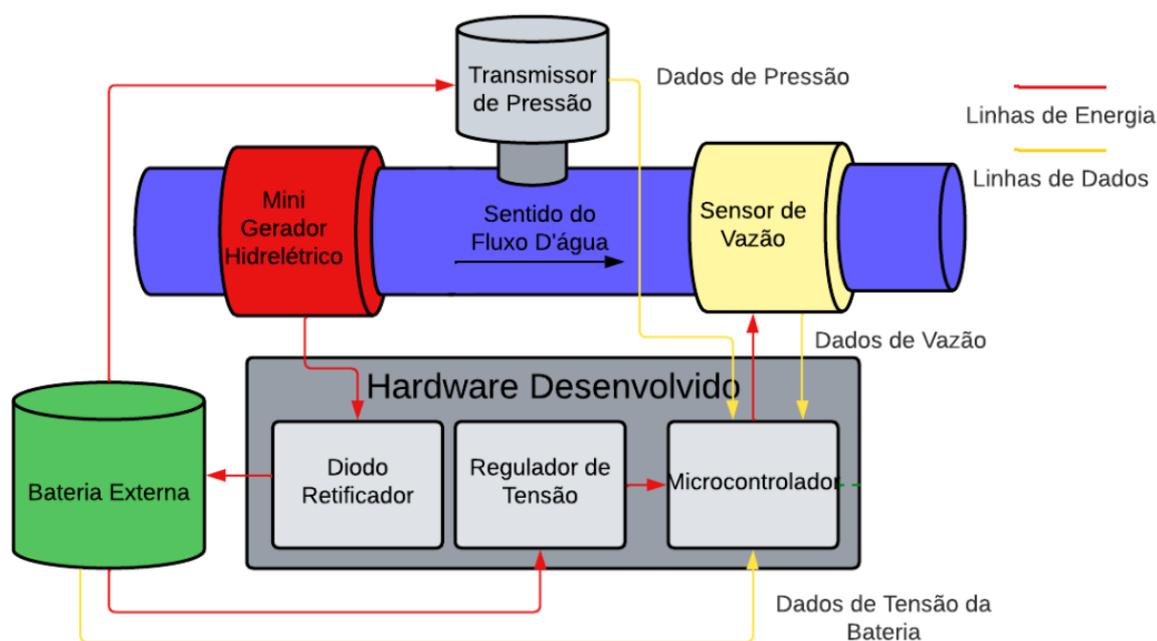


Fonte: De Oliveira et al. (2023) Adaptado.

Os dados do transmissor de pressão e do sensor de vazão são medidos pela ESP32 e enviados via Wi-Fi. Durante esse processo, o microcontrolador ativa o modo *Active Mode*, que permite a comunicação contínua e o funcionamento pleno do Wi-Fi para um servidor remoto. Para otimizar o consumo de energia, a ESP32 é colocada em modo *Deep Sleep* quando não estiver transmitindo dados. Isso significa que, enquanto estiver inativo, o microcontrolador funcionará de maneira eficiente para economizar bateria (GATIAL; BALOGH; HLUCHY, 2020). Assim que for necessário enviar os dados, o ESP32 retorna ao *Active Mode*, garantindo uma transmissão eficaz e rápida das informações para o servidor.

Na aquisição dos dados utilizados neste trabalho foi realizada uma leitura a cada um minuto. No período entre leituras o microcontrolador permaneceu no modo *Deep Sleep*. A arquitetura do sistema proposto mostrando o posicionamento dos sensores e o fluxo de dados está exemplificado na Figura 8.

Figura 8 – Idealização do Sistema.



Fonte: De Oliveira et al. (2023) adaptado.

4.1.2 SERVIDOR AWS DA AMAZON

A atualização do Marco Legal de Saneamento Básico no ano de 2020, estabeleceu entre seus objetivos o de aprimorar as condições estruturais do saneamento básico no País. Para isso, foram estabelecidos desafios para as empresas de saneamento até o ano de 2033, entre eles, a redução e controle das perdas de água, aumento da eficiência na distribuição de água tratada, estímulo à racionalização de seu consumo pelos usuários e fomento à eficiência energética.

Nesse contexto, com o apoio financeiro e a parceria institucional do CNPq, da Amazon AWS (*Amazon Web Services*), da Companhia de Água e Esgotos da Paraíba (CAGEPA) e da UFPB, estão sendo realizadas pesquisas voltadas ao desenvolvimento de algoritmos baseados em Inteligência Artificial, implementados em *Python*, para o monitoramento de grandezas hidráulicas e elétricas utilizando ferramentas da AWS IoT.

Estudos utilizando a plataforma da AWS Amazon e IoT tem sido utilizados, como o trabalho desenvolvido por Patel, Maiti e Muthuswamy (2023) que propôs um projeto de sistema para a transformação digital de uma fresadora CNC para a Indústria 4.0, usando uma rede de sensores baseada em Robot Operating System 2 (ROS 2) conectada à plataforma IoT Amazon Web Services (AWS).

Entre outros trabalhos que utilizam a plataforma da AWS, cita-se o desenvolvido

por Arthi., Nishuthan e Vignesh (2023) que propôs um sistema de baixo custo que usa Internet das Coisas, Amazon Web Services (AWS) IoT e *Machine Learning* (ML) para maximizar o rendimento e a produtividade de práticas agrícolas.

Esta pesquisa utilizará os serviços da Amazon para armazenamento dos dados em nuvem, pré-processamento, e aplicação dos algoritmos de inteligência artificial para predição da vazão na bancada experimental proposta.

4.1.3 SCADA BR

A automação é essencial para os sistemas industriais. Para otimizar e agilizar os processos, as indústrias frequentemente operam de forma remota por meio de *Supervisory Control and Data Acquisition* (SCADA) (UPADHYAY; SAMPALLI, 2020) .

O SCADA é uma tecnologia que permite o monitoramento de diferentes dispositivos, possibilitando o envio de informações de controle. Essa tecnologia oferece flexibilidade e otimiza o trabalho do operador, uma vez que não é necessário que ele fique monitorando pontos remotos que estão operando normalmente.

O SCADA BR é um sistema supervisório disponibilizado como *software* livre, o qual permite uma ampla gama de aplicações que vão desde automações prediais ou residenciais, a automação de processos industriais ou monitoramento de redes de distribuição de água (FABIO; MELO; PINA, 2015).

O SCADA BR possui funcionalidades como: configuração de tipos de dados, monitoramento de diferentes fontes de dados, configuração de menu principal, controle de variáveis, eventos gráficos e alarmes.

A utilização do SCADA aumenta os índices de produtividade na implementação no monitoramento de sistemas operacionais. O uso do SCADA em sistemas elétricos tem sido um divisor de águas na eficiência do gerenciamento de recursos energéticos. A automação de sistemas de água através do SCADA melhora a eficiência e a resposta a emergência (ALMAS et al., 2014).

Neste trabalho, o SCADA BR foi utilizado para monitoramento gráfico dos sensores e atuação no inversor de frequência, controlando assim a pressão no sistema de distribuição. Por meio de sua tela é possível acompanhar graficamente a vazão e pressão no ponto monitorado, assim como observar a detecção de *outliers* nos dados adquiridos.

4.2 MÉTODOS

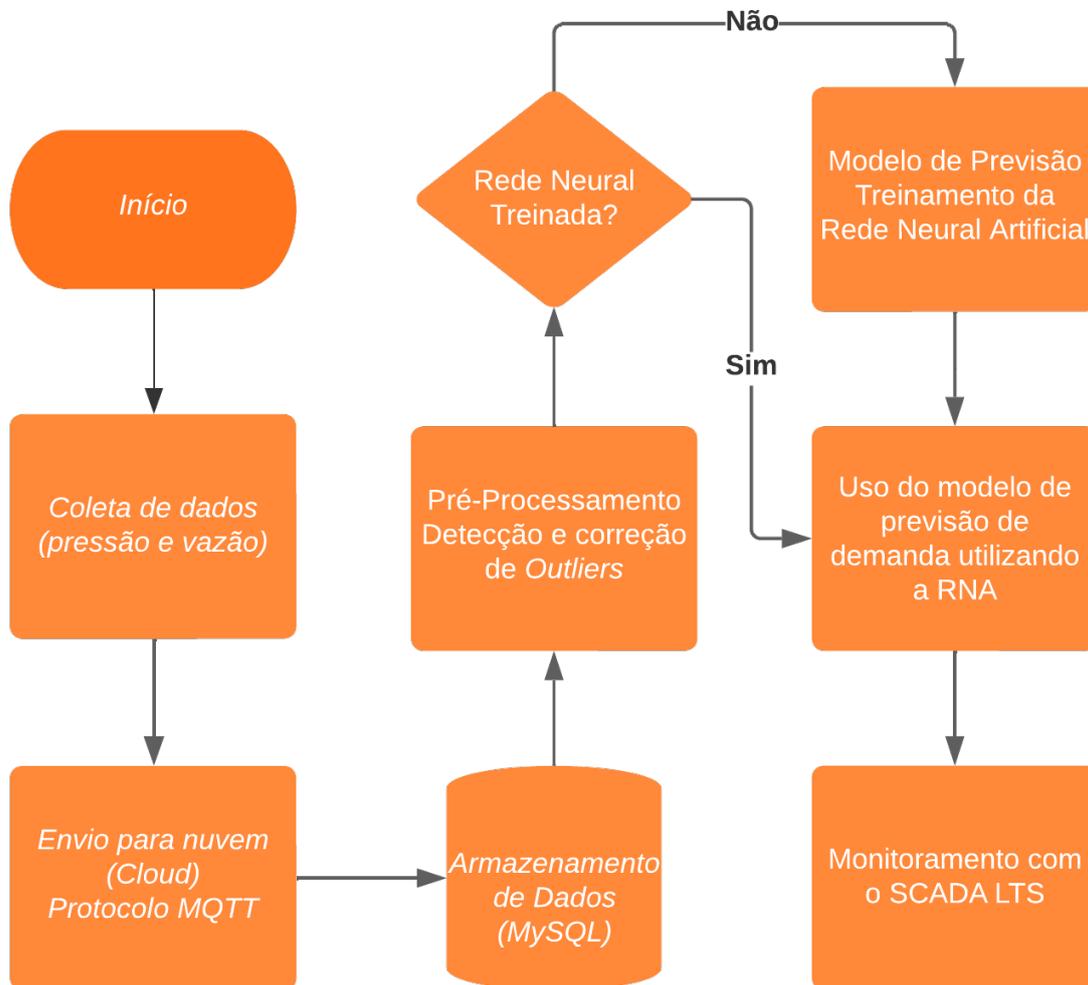
O TEDA é uma ferramenta para detecção de *outliers* em um conjunto de dados. Neste trabalho, este algoritmo será utilizado para detecção de anomalias no conjunto de

dados coletados.

Para tratamento dos dados são utilizados algoritmos desenvolvidos em *Python*, como aplicação de redes neurais para prever estados da rede de abastecimento. Uma vantagem da linguagem *Python* é o vasto conjunto de bibliotecas para processamento, análise de dados, inteligência artificial. O uso de *Python* permite realizar uma integração de banco de dados, e serviços na nuvem com eficiência e robustez, assim como possibilita a implantação de sistemas escaláveis e distribuídos, necessários para aplicações em ambientes de produção e monitoramento contínuo.

Na Figura 9 apresenta-se o fluxograma que representa o procedimento de tratamento de dados coletados. Os módulos do fluxograma estão apresentados na Seção 4.2.2 e Seção 4.2.3.

Figura 9 – Diagrama de fluxo dos dados.



Fonte: Autoria Própria.

O início representa o sistema hidráulico sendo ligado começando o processo de coleta de dados de pressão, vazão através da ESP e dos sensores instalados no sistema experimental. Os dados são coletados em intervalos de tempos controlados e enviados para a nuvem usando o protocolo MQTT. Os dados são armazenados em um servidor MySQL e aplicado o TEDA para detecção e posterior correção dos *outliers* encontrados. Os dados corrigidos são usados pela RNA para estimar as vazões futuras no sistema, sendo apresentado graficamente no SCADA.

4.2.1 MÉTRICAS DE AVALIAÇÃO

O desempenho da rede neural pode ser avaliado mediante o uso de métricas como o Erro médio quadrático e o Erro Percentual Absoluto Médio, os quais foram utilizados neste trabalho.

4.2.1.1 ERRO MÉDIO QUADRÁTICO

O *Mean Squared Error* (MSE) de um estimador é a média dos quadrados dos erros, ou seja, a diferença quadrática média entre os valores estimados e o valor atual a qual, pode ser medida mediante aplicação da Equação (3.1):

$$MSE = \frac{1}{n} \sum_{t=1}^n (A_t - P_t)^2 \quad (4.1)$$

em que A_t é o valor real e P_t é o valor estimado.

4.2.1.2 ERRO PERCENTUAL ABSOLUTO MÉDIO

Já o MAPE é uma medida da precisão de um método de previsão em estatísticas, a qual expressa a precisão como uma proporção definida pela Equação (3.2):

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - P_t}{A_t} \right| * 100\% \quad (4.2)$$

em que: A_t é o valor atual e P_t é o valor previsto. O valor absoluto neste cálculo é somado para cada ponto previsto e dividido pelo número de pontos ajustados n . Multiplicar por 100 % faz com que seja um erro em porcentagem.

Esses indicadores serão utilizados para avaliar o desempenho da rede neural treinada para previsão da vazão na bancada experimental.

4.2.2 PRÉ-PROCESSAMENTO DOS DADOS

Para pré-processamento e processamento dos dados, serão utilizadas as bibliotecas *Pandas* e *Numpy* para aplicação das técnicas de padronização e normalização dos dados. O *Pandas* é uma biblioteca *open source* disponível na linguagem *Python* que oferece ferramentas de leitura e escrita de arquivos com dados em diversos formatos: CSV, Microsoft Excel, SQL *databases*, e arquivos de texto. Já o *Numpy* é uma biblioteca de cálculo numérico em *Python*, amplamente utilizada por sua eficiência e versatilidade na análise de dados. Como ferramenta de codificação será utilizado o *Google Colaboratory*, que consiste de uma ferramenta *online* de desenvolvimento em *Python* disponibilizada pela *Google*.

4.2.3 DETECÇÃO E CORREÇÃO DE *OUTLIERS*

Durante a aquisição de dados, é comum a presença de erros de leitura, seja por problemas no sensor, ou fatores externos que gerem ruídos na leitura dos sinais captados.

Um *outlier* é um dado que está incoerente em relação aos demais, o qual, se não tratado afetará o desempenho da RNA. Para que esse problema seja dirimido, propõe-se, para detecção dos *outliers*, a aplicação do TEDA da Equação 3.4, 3.5 e 3.7.

Os *outliers* detectados serão substituídos por meio de técnicas de preenchimento de dados ausentes. Ventura et al. (2017) cita como exemplo desses métodos: Regressão Linear, *Support Vector Machine* ou substituição da lacuna pela média dos dados adjacentes. Neste trabalho, para análise dos dados *offline*, propõe-se a substituição do dado anômalo pela interpolação dos dados anteriores. Para isso é utilizada a interpolação polinomial com o método de Lagrange:

$$P(x) = \sum_{i=0}^{n-1} y_i \cdot \ell_i(x) \quad (4.3)$$

Onde o termo $\ell_i(x)$ é o polinômio base de Lagrange, definido como:

$$\ell_i(x) = \prod_{\substack{j=0 \\ j \neq i}}^{n-1} \frac{x - x_j}{x_i - x_j} \quad (4.4)$$

Aplicando as equações 4.3 e 4.4 é possível interpolar o valor de $P(x)$ em um ponto x a partir de n pontos conhecidos.

Para a aplicação *online*, caso o dado seja identificado como outlier, utiliza-se a extrapolação linear, conforme:

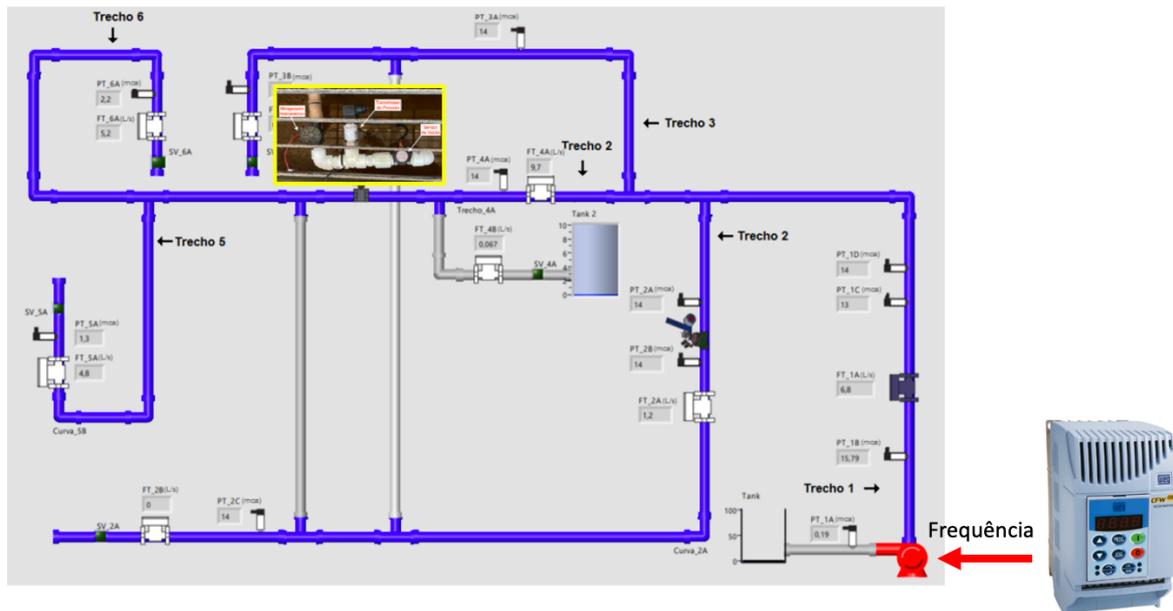
$$y_{n+1} = y_n + (y_n - y_{n-1}) \quad (4.5)$$

Dessa maneira, é possível substituir o *outlier* identificado pela extrapolação do valor y_{n+1} com base nos dois últimos valores conhecidos y_n e y_{n-1} .

4.2.4 MODELO PARA PREVISÃO DA VAZÃO

Como apresentado na Seção 4.1.1, o LENHS possui uma rede de tubulação pressurizada por uma bomba centrífuga acoplada a um motor de indução trifásico, o qual pressuriza uma rede de tubulação que simula um sistema de distribuição de água. Para controle do sistema é utilizado um inversor de frequência, com o qual é possível controlar a rotação do conjunto motor-bomba. Na Figura 10 é possível observar uma representação do sistema.

Figura 10 – Controle do Sistema Supervisório.

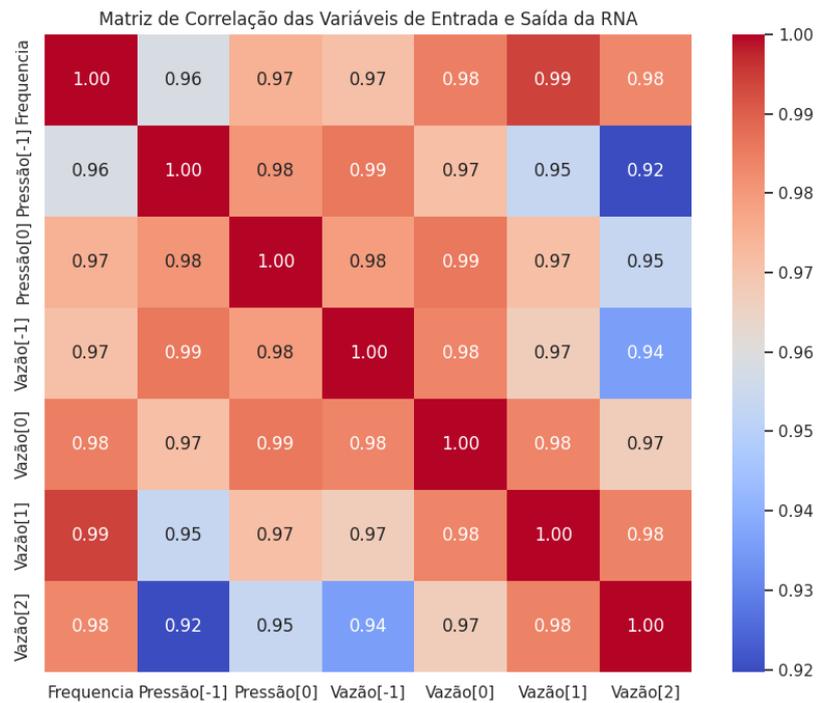


Fonte: Autoria Própria.

Modificando a frequência do inversor pelo SCADA BR é possível modificar a pressão e a vazão no ponto de medição no LENHS. Em vista disso, foram selecionadas cinco variáveis para entrada do modelo de previsão, sendo elas: a frequência do inversor, a medição de pressão e vazão atuais e anteriores.

Com o devido agrupamento dos dados da vazão e pressão, foram criadas entradas atrasadas, correspondente aos dados de vazão anteriores, as quais consistiram na variável de entrada da RNA. Esses dados correlacionados com as vazões estimadas estão representadas na Figura 11. Observa-se que as vazões estimadas apresentam forte correlação com a frequência do inversor e os valores de pressão e vazão medidos.

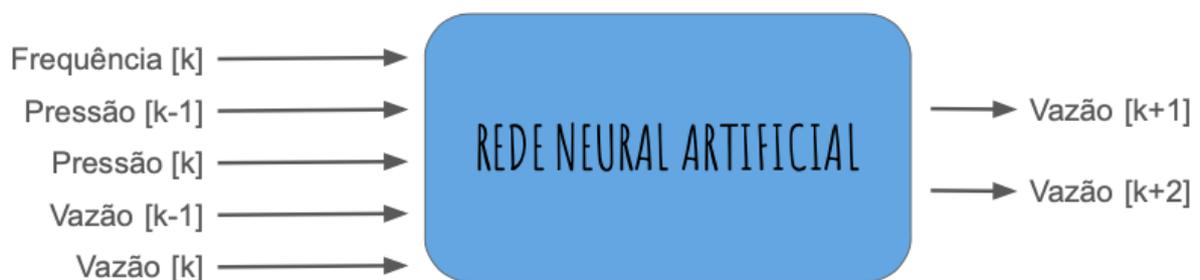
Figura 11 – Correlação entre as variáveis de entrada da RNA e as vazões estimadas.



Fonte: Autoria Própria.

Na Figura 12, é apresentada a representação dos dados de entrada e saída da Rede Neural Artificial (RNA). Nessa figura, k corresponde ao instante atual, enquanto $k + 1$ e $k + 2$ representam as previsões futuras. De forma semelhante, $k - 1$ indica as amostras defasadas em relação a um período de amostragem.

Figura 12 – Rede Neural Artificial com 5 entradas e duas saídas.



Fonte: Autoria Própria.

É esperada a ausência de dados gerados pela automatização do atraso na coluna. Como maneira de limpar os dados, foram excluídas todas as linhas com dados ausentes. Após esse procedimento, os dados foram normalizados mediante a divisão de todos os valores pelo número máximo registrado, deixando-os prontos para alimentação da rede neural.

O procedimento proposto de padronização é um requisito obrigatório para readequação dos dados, de maneira a possibilitar o treinamento da rede neural. Visando o melhoramento da resposta de rede, foi implementada a normalização dos dados de entrada, pela divisão do valor máximo esperado.

Os dados foram divididos em treino (72 %) e teste (28 %). Com o objetivo de flexibilizar o tamanho da quantidade de amostras para trabalhos futuros, o algoritmo foi desenvolvido para realizar essa proporção independentemente da quantidade de dados, ajustando assim os dados para a proporção estabelecida.

Os dados de treino foram utilizados para treinamento da rede neural, os quais possuem entradas atrasadas da vazão e pressão medidos. Os dados de entrada são constituídas pelo valor da vazão e pressão medidas na bancada experimental, assim como o valor de entrada no inversor responsável pela pressurização do sistema. Como saída, são previstas as duas vazões futuras a serem medidas pelo sistema.

Durante o treinamento da rede, de forma a verificar a melhor resposta, foram analisadas algumas funções de ativação: *Relu*, *Sigmoid*, *Tanh* e *Softmax*.

Para cada função de ativação haverá um treinamento da rede neural. O erro da RNA foi armazenado ao final de cada iteração em um dicionário. Durante o processo de seleção da função de ativação foi escolhido um valor para o número de épocas que não comprometa o tempo de processamento.

Ao final dos testes foi retornada a função que apresentar menor erro, a qual será utilizada para treinamento final da rede neural. Nesse estudo foi utilizado o otimizador Adam para uma convergência otimizada da rede.

Após treinamento da rede, foram aplicados os dados de teste para predição das vazões futuras, e observado o erro apresentado pela RNA, assim como são examinados graficamente as vazões e pressões lidas na bancada experimental.

5 RESULTADOS

Neste capítulo são demonstrados e discutidos os resultados obtidos neste trabalho, para isso, será apresentada a Descrição do Conjunto de Dados na Seção 5.1, no qual é abordado o padrão de consumo de uma residência. Em seguida é apresentada a aquisição dos dados e envio para a nuvem, apresentada na Seção 5.2, que contém o conjunto de medições da pressão e vazão lidos na bancada experimental. Na Seção 5.3 é apresentada a aplicação do TEDA para detecção de *outliers*, com o objetivo de que os dados após tratados possam ser utilizados no treinamento da rede neural. Na Seção 5.4 é apresentada a aplicação da RNA treinada para previsão de vazões futuras. Por fim, é apresentada a aplicação do TEDA e RNA sendo executados em tempo real, emulando a aplicação da metodologia proposta em um centro de distribuição.

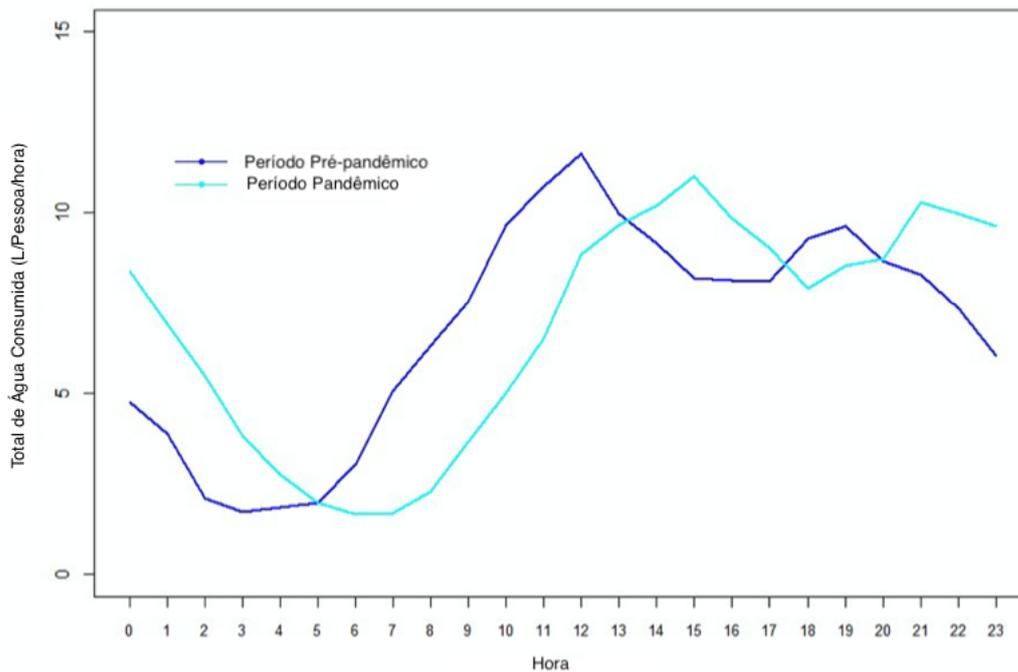
5.1 DESCRIÇÃO DO CONJUNTO DE DADOS

Para pressurização do sistema, foi necessário emular o consumo de uma residência, para que o conjunto de dados se aproxime do medido em uma unidade consumidora.

Cominato et al. (2022) realizou um estudo comparativo entre o consumo de água no período pré-pandemia e durante o período de quarentena em 14 (quatoze) residências brasileiras, localizadas em Joinville, sul do Brasil. Nesse estudo, o autor apresentou o padrão de consumo das residências em estudo nos diferentes meses do ano, estabelecendo para isso um gráfico médio o qual define um padrão de consumo de água médio para o conjunto de residências em estudo (Figura 13). O perfil de consumo de água para o período pré-pandemia ilustrado na Figura 13 está próximo do relatado na literatura (COMINATO et al., 2022).

Os autores observaram que os perfis de consumo total de água para os períodos pré-pandemia e pandemia foram semelhantes, com um padrão bem definido, mostrando apenas uma mudança do pico durante o dia, que passou das 12 horas antes da pandemia para as 15 horas durante a pandemia.

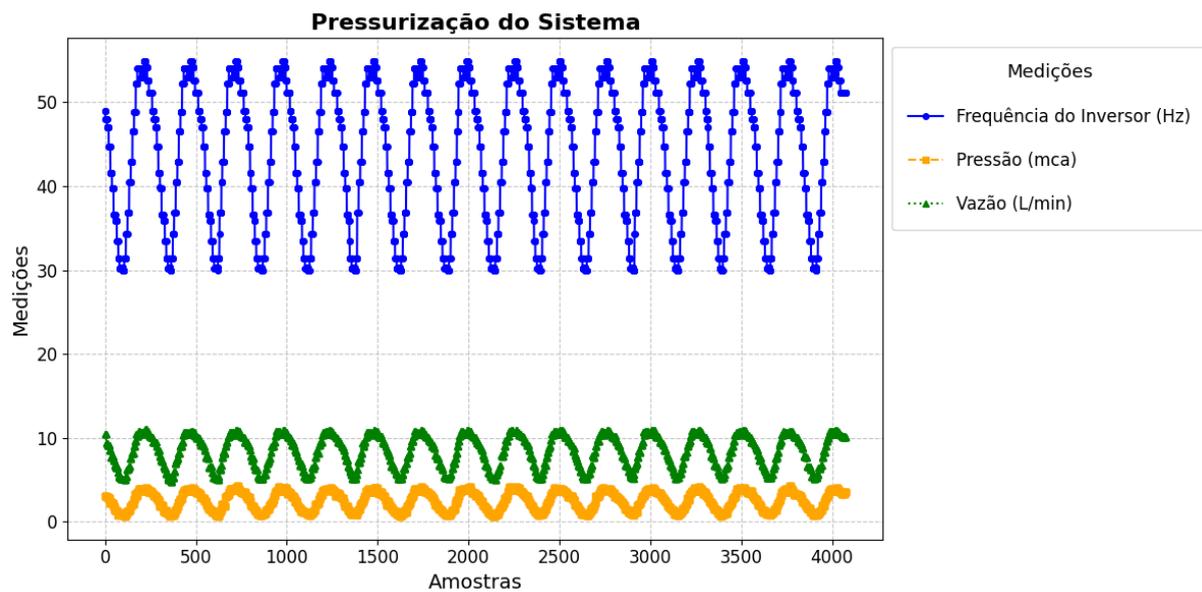
Figura 13 – Perfil de consumo de água por hora para os períodos pré-pandêmico e pandêmico.



Fonte: (COMINATO et al., 2022) adaptado.

Com o objetivo de estabelecer um conjunto de dados semelhante ao apresentado pelo autor, Oliveira (2023) modificou o valor de entrada do inversor para que a curva da vazão se assemelhasse à apresentada por Cominato et al. (2022), conforme apresentado na Seção 4.2.4. A resposta gráfica da relação entre o dado de entrada do inversor no sistema supervisorio e a resposta da vazão lida na tubulação monitorada está apresentada na Figura 14.

Figura 14 – Frequência do inversor e resposta da vazão medida.



Fonte: Autoria própria.

Para construção do gráfico apresentado na Figura 14 utilizou-se uma variação da frequência do inversor variando de 30,0 Hz a 54,91 Hz, em um conjunto de 60 valores definidos. Observa-se que modificando-se a frequência do inversor, é possível modificar a vazão e a pressão do sistema no ponto monitorado.

5.2 AQUISIÇÃO DE DADOS E ENVIO PARA A NUVEM

A partir do sistema de aquisição de dados desenvolvido, foi realizada a aquisição da vazão e pressão no ponto monitorado.

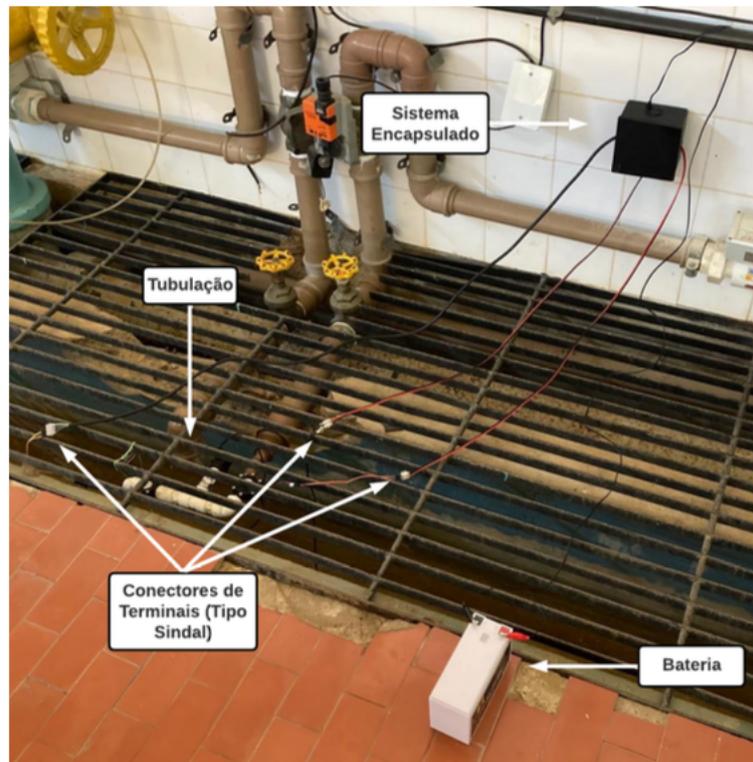
Neste trabalho, o processo de aquisição e armazenamento foi aprimorado. Utilizou-se um *script* para controle do inversor de frequência, com o qual é possível atuar no conjunto motor-bomba por meio do SCADA BR.

A leitura dos dados inicia-se com o sistema supervisório operando com a frequência do inversor em 30 Hz, a placa de aquisição conectada aos sensores e a uma fonte de alimentação externa de 12 V. Modifica-se a frequência do inversor em sessenta valores distintos, com valor máximo igual 54,91 Hz.

O sistema é capaz de medir tanto a pressão quanto a vazão em uma tubulação de meia polegada (Figura 15). Essa informação é importante, pois permite observar o instante em que o fluxo de água na tubulação é suficiente para proporcionar autonomia ao sistema de medição.

O fluxo de água na tubulação gera energia que é direcionada a um conjunto de diodos retificadores, que desempenham um papel importante na regulação da fonte de energia do sistema.

Figura 15 – Funcionamento do Sistema.



Fonte: De Oliveira et al. (2023).

Os dados são armazenados localmente na memória RTC (*Real-Time Clock*) do microcontrolador, que mantém a contagem de tempo de forma precisa e contínua, mesmo quando o dispositivo está desligado ou reiniciado. Essa memória é utilizada para armazenar informações temporais de forma estável. Durante uma hora, as medições são realizadas a cada ciclo de um minuto, resultando em sessenta medições por hora.

A ESP32 permanece no modo *DeepSleep* entre as medições, reduzindo assim seu consumo de energia. Os dados armazenados são então enviados a um *Broker* Público via protocolo MQTT (*Message Queuing Telemetry Transport*) após a ESP32 sair do modo *deepsleep* e se conectar a WiFi.

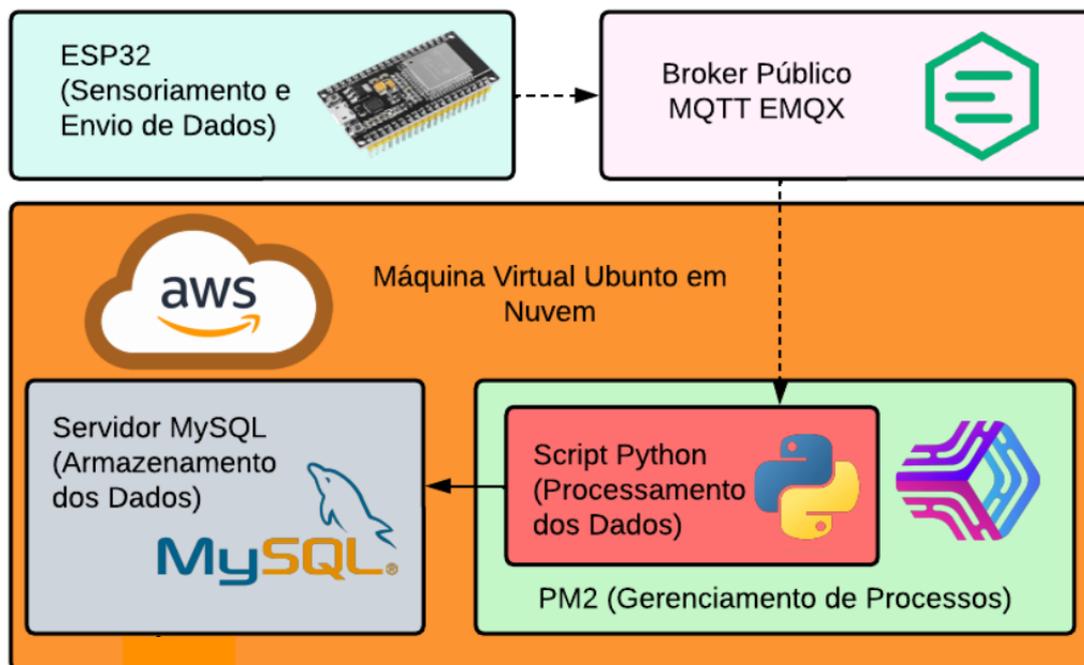
Utilizando-se um *script* em *Python* executado em um servidor hospedado na AWS e gerenciado pelo PM2, consulta-se o *Broker* público e lê-se os novos dados, armazenando-os em um servidor *MySQL*. O *MySQL* possui uma estrutura baseada em arquivos e é capaz de lidar com várias operações de leitura simultaneamente. Essa característica é

particularmente vantajosa em aplicações de IoT com um alto volume de tráfego de dados. (OWENS, 2006).

A partir de uma máquina conectada à internet os dados podem ser acessados e exibidos fazendo-se uma requisição ao banco de dados. Na Figura 16, observa-se a arquitetura do sistema descrito. A ESP32 transmite os dados medidos via protocolo MQTT, utilizando um *broker* público, que atua como um intermediário na transferência de informações.

Os dados são enviados para uma máquina virtual Ubuntu, hospedada na plataforma de computação em nuvem AWS, onde todo o processamento ocorre. Nesta máquina virtual, um *script* Python é executado para tratar os dados recebidos e inseri-los em um banco de dados MySQL, o que garante o armazenamento e a persistência das informações para futuras análises.

Figura 16 – Arquitetura do Sistema.



Fonte: Oliveira (2023) adaptado.

O gerenciamento desse processo é realizado pelo PM2, uma ferramenta que assegura a execução contínua e eficiente dos *scripts* Python. Com essa arquitetura, o sistema possibilita a aquisição em tempo real, a transmissão segura e o armazenamento estruturado dos dados monitorados no sistema de distribuição de água.

5.3 DETEÇÃO DE OUTLIERS UTILIZANDO O TEDA

5.3.1 APLICAÇÃO DO TEDA *OFFLINE*

Com os dados medidos no LENHS, utilizou-se o *Google Colaboratory* para acessar o servidor *MySQL* hospedado na AWS Amazon.

No código em *Python* que foi desenvolvido, utilizaram-se as bibliotecas *tensorflow*, *pandas*, *matplotlib.pyplot*, *datetime*, *sklearn.metrics* e a *math* para as etapas de tratamento dos dados e processamento.

Inicialmente, o servidor AWS da Amazon é acessado por meio da biblioteca *mysql.connector*. Os dados lidos representam a emulação do consumo de água de uma residência no período de um dia.

Com os dados armazenados em um *dataframe*, ilustrado na Figura 17, pode-se observar três colunas: hora, pressão, vazão e tensão, sendo sessenta valores armazenados em cada uma. Como é abordado, esse conjunto de valores representa a emulação de consumo de uma residência.

Figura 17 – Armazenamento em um dataframe dos dados adquiridos do servidor AWS da Amazon.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 60 entries, 0 to 59
Data columns (total 4 columns):
#   Column      Non-Null Count  Dtype
---  -
0   Hora        60 non-null    datetime64[ns]
1   Pressão     60 non-null    float64
2   Vazão       60 non-null    float64
3   Tensão      60 non-null    float64
dtypes: datetime64[ns](1), float64(3)
memory usage: 2.0 KB
```

Fonte: Autoria Própria.

Na Figura 18 são apresentados os primeiros cinco registros contidos nas colunas. Observa-se que o horário apresentado é a hora da aquisição dos dados, devendo-se realizar um ajuste para relacionar a medição à hora do dia correspondente à emulação.

Figura 18 – Disposição dos dados armazenados no dataframe.

	Hora	Pressão	Vazão	Tensão
0	2023-11-22 14:15:34	1.01	12.90	12.47
1	2023-11-22 14:16:34	1.10	12.90	12.47
2	2023-11-22 14:17:34	0.85	11.85	12.46
3	2023-11-22 14:18:34	0.76	11.85	12.46
4	2023-11-22 14:19:34	0.84	12.00	12.46

Fonte: Autoria Própria.

Para ajuste, modificou-se os dados da coluna Hora para que os sessenta valores igualmente espaçados fossem distribuídos entre os horários de um dia completo, com início a 00:00 e término às 23:59 (Figura 19). Para isso, os segundos foram desconsiderados nessa aproximação. Com esse ajuste, cada medição representa a vazão e pressão lidos a cada 24 (vinte e quatro) minutos. Esse valor surge a partir da amostragem de frequências assumida pelo inversor. Como sessenta valores frequência representam o padrão de consumo de um dia, o intervalo entre dois valores representa um intervalo de amostragem de 24 minutos.

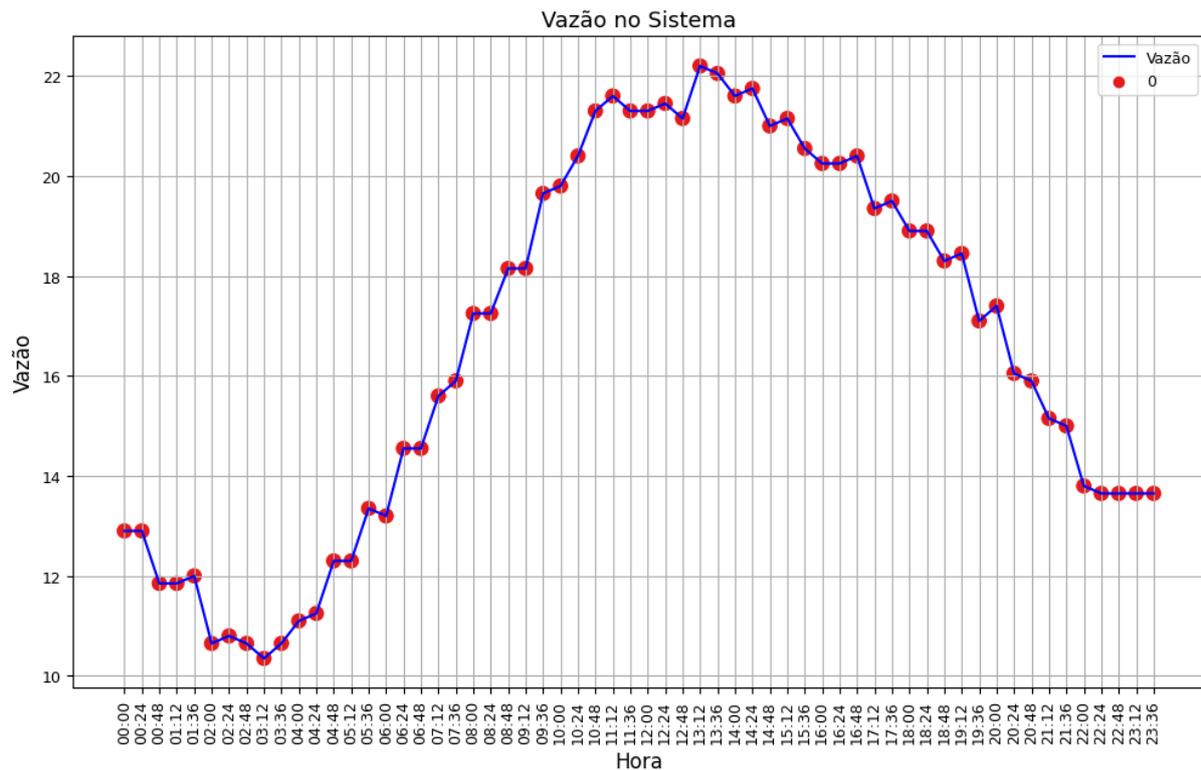
Figura 19 – Ajuste do horário de emulação.

	Hora	Pressão	Vazão	Tensão
0	00:00	1.01	12.90	12.47
1	00:24	1.10	12.90	12.47
2	00:48	0.85	11.85	12.46
3	01:12	0.76	11.85	12.46
4	01:36	0.84	12.00	12.46

Fonte: Autoria Própria.

O gráfico dos valores da vazão coletados pelo sistema IoT, a partir da emulação na planta hidráulica do laboratório, semelhante ao consumo médio de água em uma residência ao longo de um dia está apresentado na Figura 20

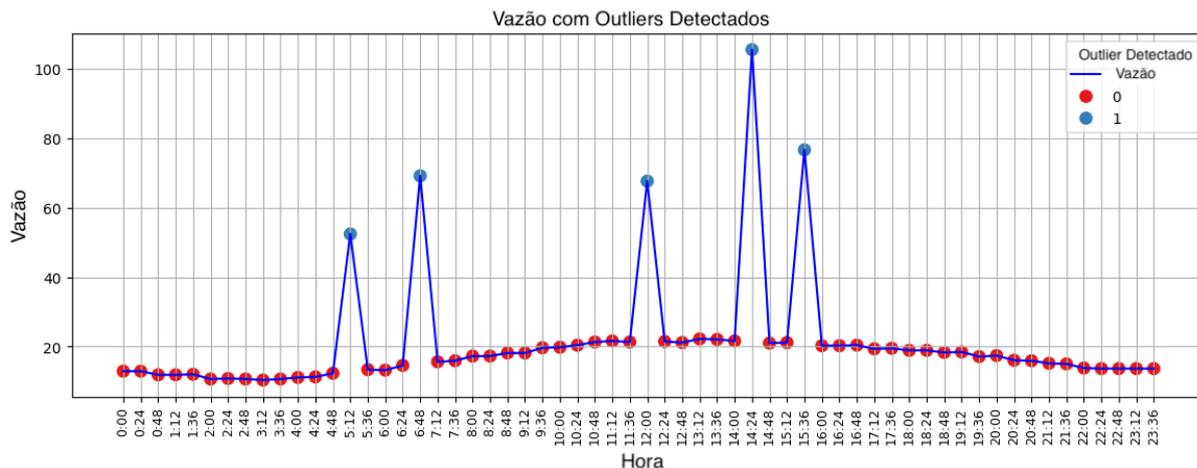
Figura 20 – Dados medidos no LENHS.



Fonte: Autoria Própria.

Contudo, como já abordado na seção 3.2, os dados adquiridos podem conter erros de leitura ou apresentar valores dados faltantes, comprometendo a interpretação do comportamento sistema de distribuição. Para correção dessas anomalias, foi aplicado o algoritmo proposto por Angelov (2012), com o qual é possível identificar a presença de *outliers* nos dados recebidos. Para teste do algoritmo, foram atribuídos valores aleatórios em cinco registros dos dados, para simular falhas na leitura do sensor de vazão, gerando, assim, *outliers* no conjunto de dados lidos. Na Figura 21, é possível observar o conjunto de dados com os *outliers* identificados após aplicação do TEDA.

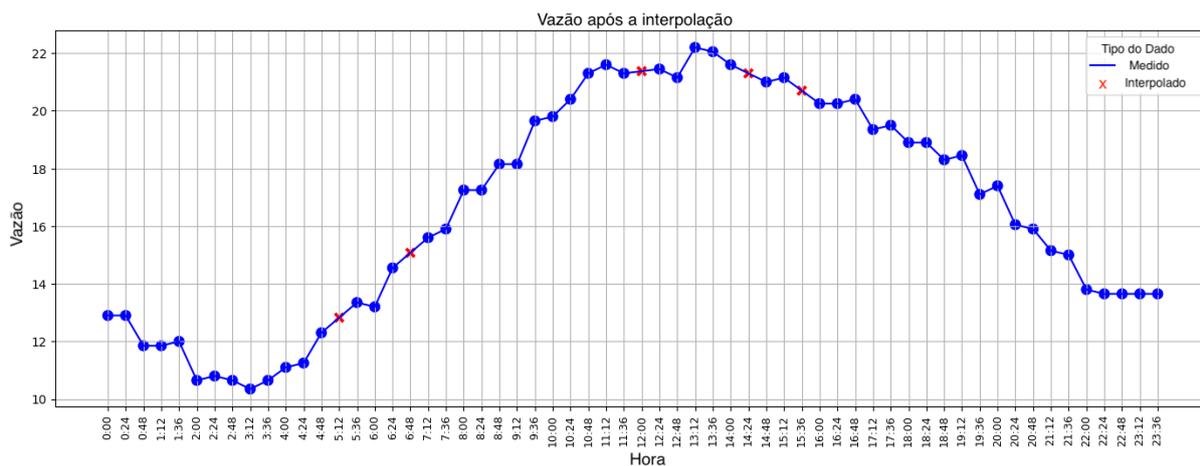
Figura 21 – Dados adquiridos no LENHS, com inserção de cinco outliers.



Fonte: Autoria Própria.

Após a identificação do erro no conjunto de dados, aplicou-se uma interpolação linear, a qual consiste na média entre as amostras anterior e posterior à mostra detectada como outlier, sendo assim utilizada para substituição dos dados identificados pelo algoritmo como outlier. Na Figura 22 é possível observar o novo gráfico após a correção dos dados identificados como heterogêneos pelo TEDA. O erro absoluto com a substituição dos dados reais pelos dados interpolados estão apresentados na Tabela 5.

Figura 22 – Dados adquiridos no LENHS, com outliers corrigidos.



Fonte: Autoria Própria.

Tabela 5 – Tabela dos *outliers* interpolados.

Índice	Valor real (L/min)	Valor interpolado (L/min)	Erro absoluto	Erro relativo (%)
13	12.30	12.82	0.52	4.26
17	14.55	15.07	0.52	3.60
30	21.30	21.37	0.07	0.35
36	21.75	21.30	0.45	2.06
39	20.55	20.70	0.15	0.72

Esse processo de detecção e correção de anomalias nos dados é fundamental para este trabalho pois a presença de um *outlier* irá gerar erros na aplicação da rede neural artificial utilizada para predição das vazões futuras, conforme apresentado na seção 5.4.

5.3.2 APLICAÇÃO DO TEDA *ONLINE*

Neste tópico utilizou-se o TEDA para emulação do sistema executando *online*. Foram adquiridos um conjunto de valores de vazão, pressão e frequência do inversor na rede, de maneira que pudesse simular uma *stream* de dados sequenciais, e, assim, poder examinar o comportamento do TEDA na detecção dos *outliers*.

Os dados utilizados para o estudo foram adquiridos no LENHS. Para realizar a análise de detecção de *outliers*, foi desenvolvido um *script* específico com o objetivo de inserir *outliers* de maneira aleatória no conjunto de dados. A inserção de *outliers* consistiu na adição de um número aleatório entre -100 e 100 ao dado lido, gerando assim uma anomalia no dado medido. O percentual de inserção de *outliers* foi de 3 %, garantindo uma quantidade controlada de anomalias que permite uma análise mais realista do sistema. Esse processo de geração de *outliers* foi aplicado a cada um dos registros presentes no *dataset*, permitindo a inclusão de anomalias controladas.

A seguir, aplicou-se o algoritmo TEDA sobre o conjunto modificado, com o intuito de verificar a capacidade de detecção de *outliers* introduzidos. Esse procedimento foi repetido diversas vezes, variando-se a proporção de *outliers* no conjunto de dados, de modo a avaliar como o TEDA se comporta sob diferentes condições de ruído nos dados.

Para análise do desempenho do TEDA utilizou-se o cálculo da acurácia, a qual é uma métrica de desempenho que indica a proporção de previsões corretas em relação ao total de previsões realizadas. A acurácia varia de 0 (nenhuma previsão correta) a 1 (todas as previsões corretas), representando a fração do total de previsões corretas. A acurácia pode ser obtida por meio da aplicação da Equação 5.1:

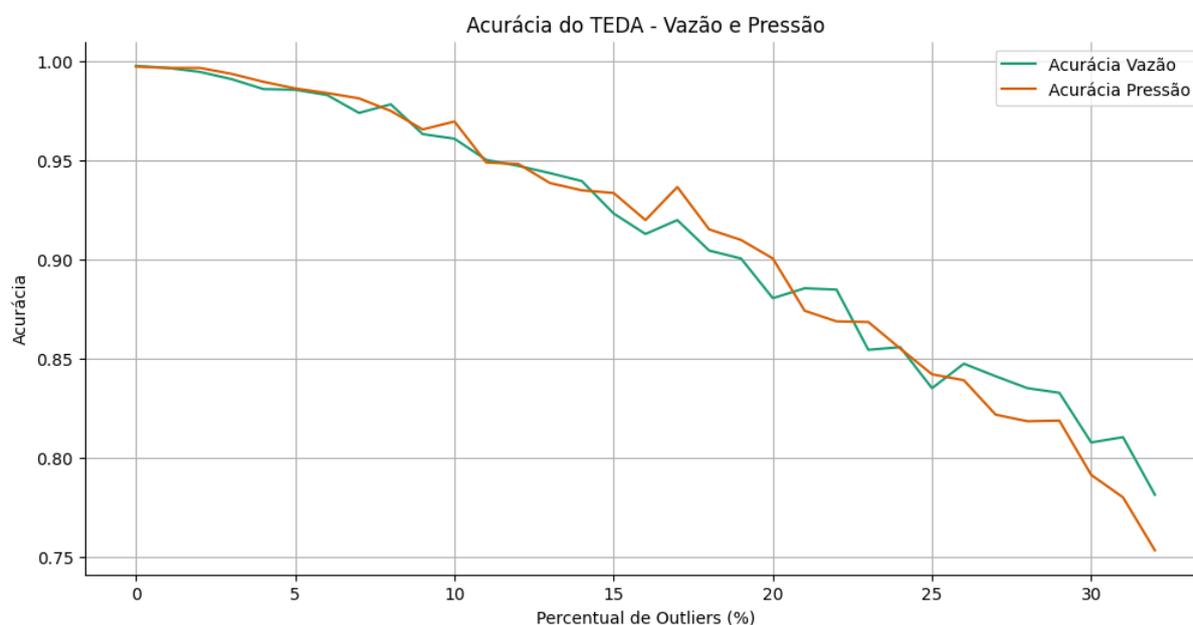
$$\text{Acurácia} = \frac{TP + TN}{TP + TN + FP + FN} \quad (5.1)$$

Em que:

- TP: Verdadeiros Positivos
- TN: Verdadeiros Negativos
- FP: Falsos Positivos
- FN: Falsos Negativos

Na Figura 23 é possível observar os resultados obtidos a partir das iterações realizadas, demonstrando a acurácia da aplicação do TEDA em função do percentual de *outliers* presentes no conjunto de dados:

Figura 23 – Acurácia do TEDA na variação do percentual de *outliers* inseridos na leitura da vazão e pressão.



Fonte: Autoria Própria.

Na Figura 23 observa-se que o TEDA apresenta uma acurácia de acima de 95 % para um percentual de *outliers* até 10 %. Com a presença de mais de 30 % de *outliers* no *dataset*, o TEDA apresentou uma acurácia inferior a 80 %.

O procedimento foi repetido utilizando o mesmo conjunto de dados, porém, desta vez, variando o valor de m (*threshold*) no algoritmo TEDA. O objetivo dessa etapa foi avaliar como diferentes valores de m influenciam a acurácia do TEDA na detecção de *outliers*. Para cada valor de m , o TEDA foi aplicado sobre o mesmo conjunto de dados, permitindo observar as variações no desempenho do algoritmo em termos de precisão e capacidade de identificar corretamente os *outliers*. Esse processo possibilitou a escolha de

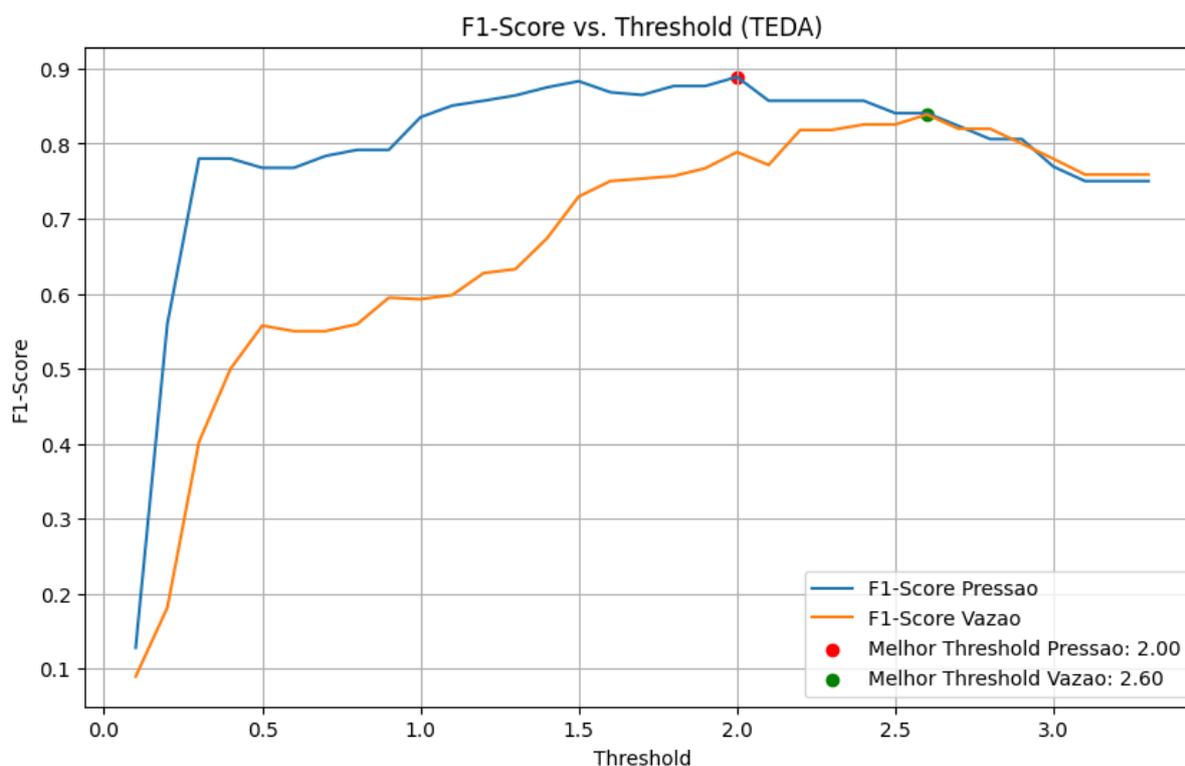
um valor de m que apresentasse a melhor acurácia para os dados analisados, garantindo uma detecção de *outliers* mais eficiente e consistente.

Para mensurar o desempenho do TEDA na detecção dos *outliers* inseridos nas medições da pressão e vazão utilizou-se o $F1 - Score$, conforme Equação 5.2. O $F1 - Score$ é uma métrica de desempenho que representa o equilíbrio entre precisão e sensibilidade em um modelo de classificação.

$$F1\text{-Score} = 2 \times \frac{\frac{TP}{TP+FP} \times \frac{TP}{TP+FN}}{\frac{TP}{TP+FP} + \frac{TP}{TP+FN}} \quad (5.2)$$

Na Figura 24 é possível visualizar os resultados obtidos ao longo das diferentes variações do *threshold*.

Figura 24 – Variação do F1-SCORE do TEDA no ajuste da sensibilidade de detecção de outliers.



Fonte: Autoria Própria.

5.4 APLICAÇÃO DA RNA PARA PREVISÃO DA VAZÃO

Após a realização da etapa de detecção de erros encontrados nos dados lidos, apresentado na Seção 5.3, segue-se para treinamento e aplicação de uma rede neural

artificial para identificação das vazões futuras do sistema de distribuição.

Para treinamento da RNA, a frequência do inversor foi utilizada como um dos parâmetros da camada de entrada da rede neural. Outros parâmetros também utilizados foram a vazão e a pressão medidos pela ESP-32. Em síntese, a rede neural configurada possui cinco neurônios na camada de entrada e dois neurônios de camada de saída. A RNA segue o modelo proposto na Seção 4.2.4 para previsão da vazão.

Objetiva-se que a partir da frequência do inversor, pressão e vazão do sistema, em instantes de tempos distintos, possa se estimar a previsão das vazões em dois intervalos de leitura de dados à frente, nesse caso, correspondendo à resposta do sistema em 48 minutos futuros, os quais consistem em duas amostras de 24 minutos.

O *dataframe* ajustado para a entrada de treinamento da rede está apresentado na Figura 25. As cinco primeiras colunas representam os dados que alimentam a camada de entrada. As duas últimas colunas, correspondentes à leitura da vazão futura, são utilizadas nos neurônios de saída para o treinamento supervisionado da rede neural.

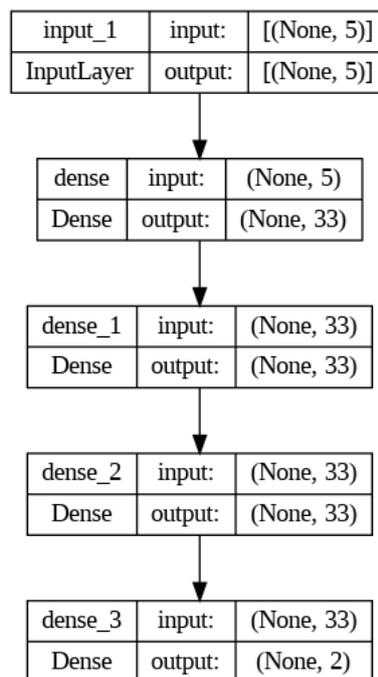
Figura 25 – Dados de entrada da RNA.

	Frequencia	Pressão[-1]	Pressão[0]	Vazão[-1]	Vazão[0]	Vazão[1]	Vazão[2]
0	35.881529	NaN	1.01	NaN	12.90	12.90	11.85
1	35.881529	1.01	1.10	12.90	12.90	11.85	11.85
2	33.459936	1.10	0.85	12.90	11.85	11.85	12.00
3	33.459936	0.85	0.76	11.85	11.85	12.00	10.65
4	31.377329	0.76	0.84	11.85	12.00	10.65	10.80

Fonte: Autoria Própria.

Além das camadas de entrada, utilizaram-se três camadas ocultas, com 33 neurônios cada (Figura 26).

Figura 26 – Rede Neural Artificial utilizada.

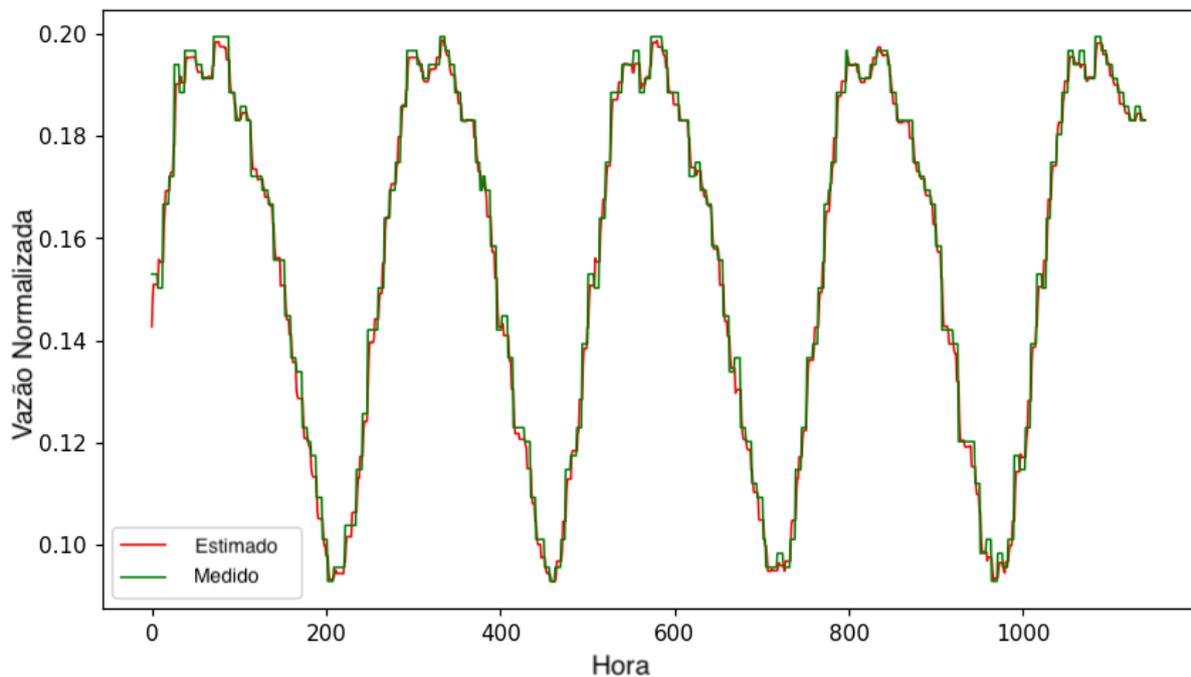


Fonte: Autoria Própria.

Para treinamento da rede, dividiram-se os dados em 72 % no conjunto de treinamento e 28 % para teste, como apresentado na Seção 4.2.2. Para a definição da função de ativação dos neurônios, a rede é treinada com 300 épocas para as seguintes funções de ativação: tangente hiperbólica, sigmoide, relu e softmax. Ao final de cada treinamento, o erro médio quadrático é calculado, e a função que apresenta o menor erro é utilizada no treinamento da RNA com 700 épocas.

Após o treinamento final, a rede é utilizada para previsão da vazão de duas leituras à frente. Na Figura 28, é possível observar a resposta gráfica dos dados da vazão medida e da vazão prevista pela RNA treinada.

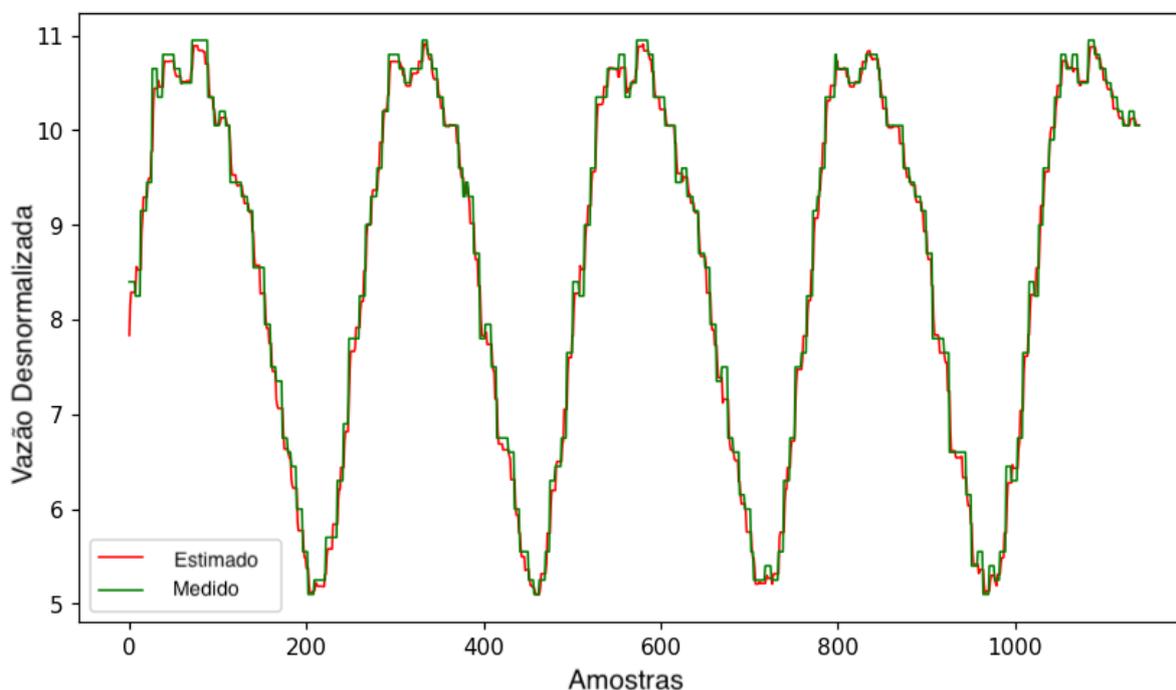
Figura 27 – Resultado obtido com a aplicação da RNA para previsão da vazão normalizada.



Fonte: Autoria Própria.

Observa-se que a saída da rede neural encontra-se normalizada em função do processo de normalização aplicado aos dados antes do treinamento. Esse procedimento é essencial para garantir que as variáveis de entrada e saída estejam na mesma escala, evitando que valores com magnitudes muito diferentes possam influenciar indevidamente o aprendizado da rede. Para restituir os valores à sua escala original e, assim, recuperar a amplitude real das leituras de vazão, foi implementado um processo de desnormalização. Esse processo consistiu em multiplicar o autovetor resultante da rede pelos respectivos autovalores utilizados no estágio de normalização. Dessa forma, foi possível reverter os efeitos da normalização e obter os valores reais de saída. O gráfico contendo os dados desnormalizados é apresentado na Figura 28.

Figura 28 – Resultado obtido com a aplicação da RNA para previsão da vazão desnormalizada.



Fonte: Autoria Própria.

Os resultados obtidos a partir da RNA estão representados na Figura 28. A RNA apresentou um MSE de $4.4284e-05$, o que reflete um bom desempenho, visto que esse valor indica uma diferença mínima entre as previsões da rede e os dados reais. Além disso, o erro MAPE foi de 3,15 %, o que significa que as previsões da rede neural apresentaram um desvio médio de apenas 3,15 % em relação aos valores observados. O MAPE está dentro da faixa de estimação esperada, com um valor abaixo de 5 %.

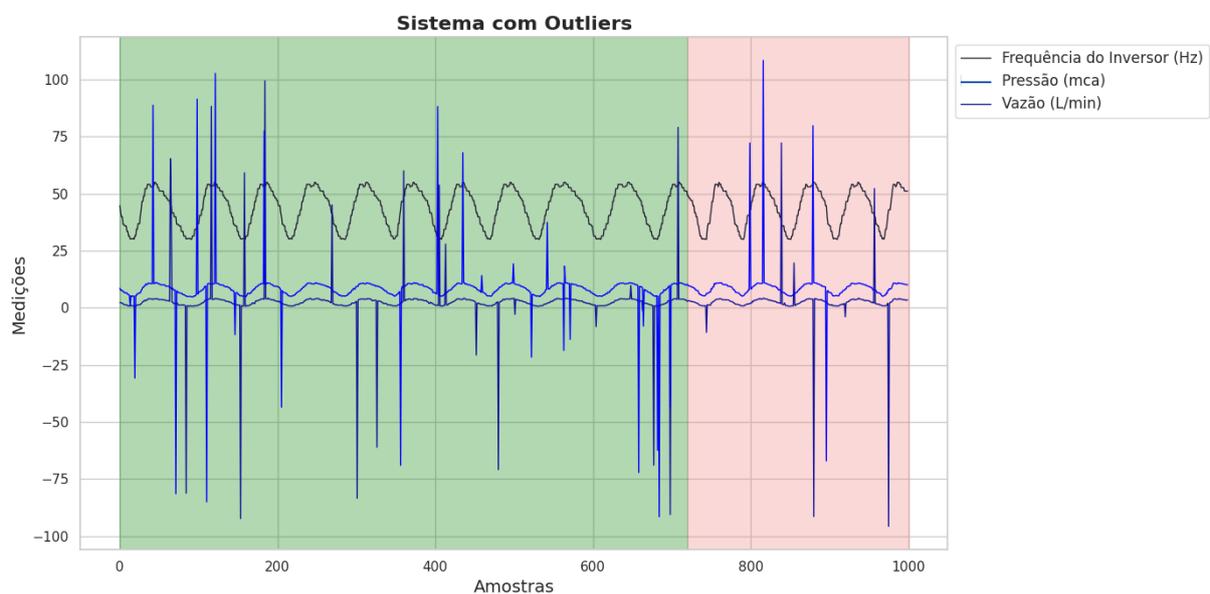
O baixo valor de MAPE decorre da capacidade da rede de lidar com a variabilidade dos dados, mantendo as previsões dentro de um intervalo de erro aceitável para aplicações em tempo real, como o monitoramento de redes de abastecimento. Esses resultados são especialmente importantes para sistemas de saneamento no contexto brasileiro, onde a eficiência na gestão de recursos hídricos é crucial para lidar com desafios como a escassez de água e a otimização de redes de distribuição.

5.4.1 UTILIZAÇÃO DA RNA TREINADA COM OUTLIERS: AVALIAÇÃO SEM APLICAÇÃO DO TEDA

Com o objetivo de identificar o impacto da presença de *outliers* no treinamento e na aplicação da Rede Neural Artificial (RNA), foi aplicado um algoritmo de geração

de *outliers* com uma taxa de 3 %, aleatoriamente, no conjunto de dados de treinamento e teste. A introdução desses *outliers* visa simular condições reais em que dados atípicos podem afetar o desempenho da rede. O conjunto de dados resultante, após a inserção dos *outliers*, pode ser visualizado graficamente na Figura 29, em que a cor de fundo verde indica os dados utilizados para treinamento da RNA e em cor-de-rosa os dados utilizados para teste. Essa visualização permite uma análise mais clara das alterações nos padrões dos dados e seu potencial efeito sobre a capacidade da RNA em aprender e generalizar a partir do conjunto de dados alterado.

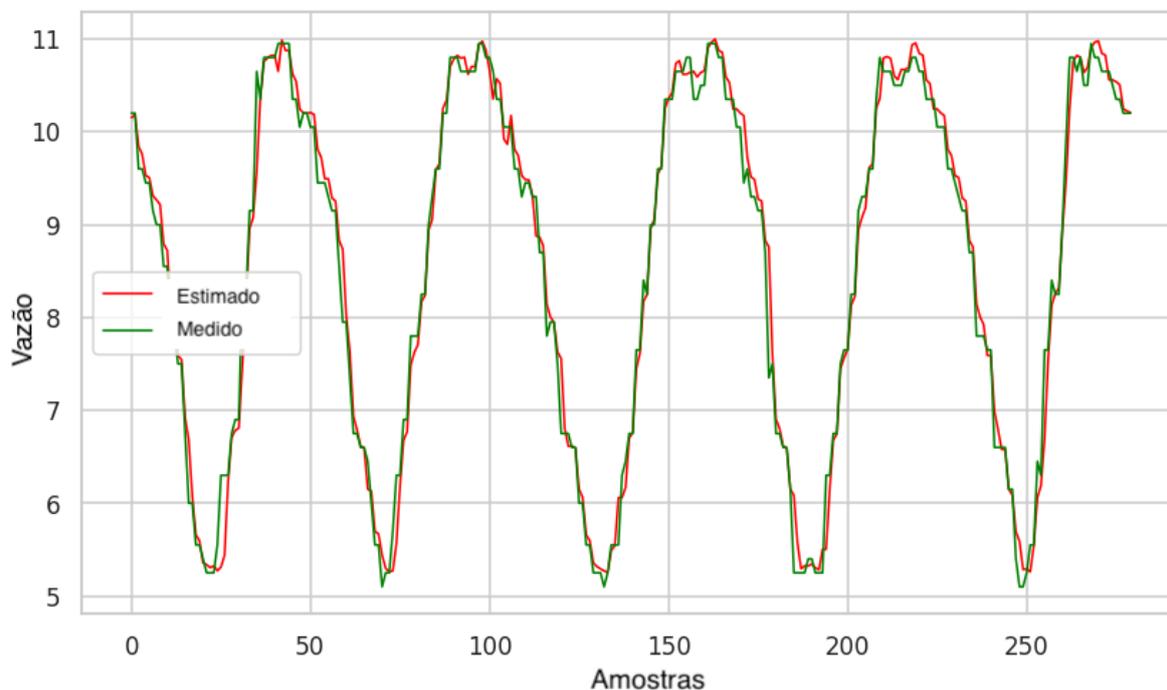
Figura 29 – Gráfico com a previsão do consumo, considerando a presença de outliers gerados em 3 % dos dados.



Fonte: Autoria Própria.

Foi aplicado o mesmo procedimento metodológico descrito na Seção 5.4, onde a RNA foi treinada utilizando 72 % do conjunto de dados e testada com os 28 % restantes. Após o treinamento da rede neural e a aplicação dos dados de teste, a resposta gráfica da previsão da demanda de consumo pode ser observada na Figura 30.

Figura 30 – Gráfico com a previsão do consumo com a presença de outliers gerados a uma taxa de 3 %.



Fonte: Autoria Própria.

Os resultados obtidos a partir da aplicação da RNA com a presença de *outliers* durante o treinamento e nos dados de teste são apresentados no gráfico que exibe os valores de previsão. A RNA demonstrou um erro quadrático médio (*loss*) de $1.6716e-05$, indicando um desempenho robusto na modelagem das previsões de demanda de consumo, mesmo na presença de dados atípicos. Este valor de *loss* sugere que as previsões da rede estão muito próximas dos valores reais, evidenciando a capacidade da RNA em lidar com a variabilidade introduzida pelos *outliers*. Além disso, o MAPE registrado foi de 3,4 %. Esse resultado indica que, em média, as previsões da rede apresentaram um desvio de aproximadamente 3,4 % em relação aos valores reais. Esse nível de precisão é aceitável para aplicações práticas, destacando a eficácia da RNA em condições não ideais.

5.4.2 UTILIZAÇÃO DA RNA TREINADA SEM OUTLIERS: AVALIAÇÃO DO DESEMPENHO EM DADOS DE TESTE COM OUTLIERS, SEM A APLICAÇÃO DO TEDA

A imprevisibilidade dos *outliers* é um desafio significativo no treinamento da RNA. Embora uma rede possa ser treinada em condições em que os sensores estão operando corretamente, a presença de diversos fatores externos, como falhas nos sensores,

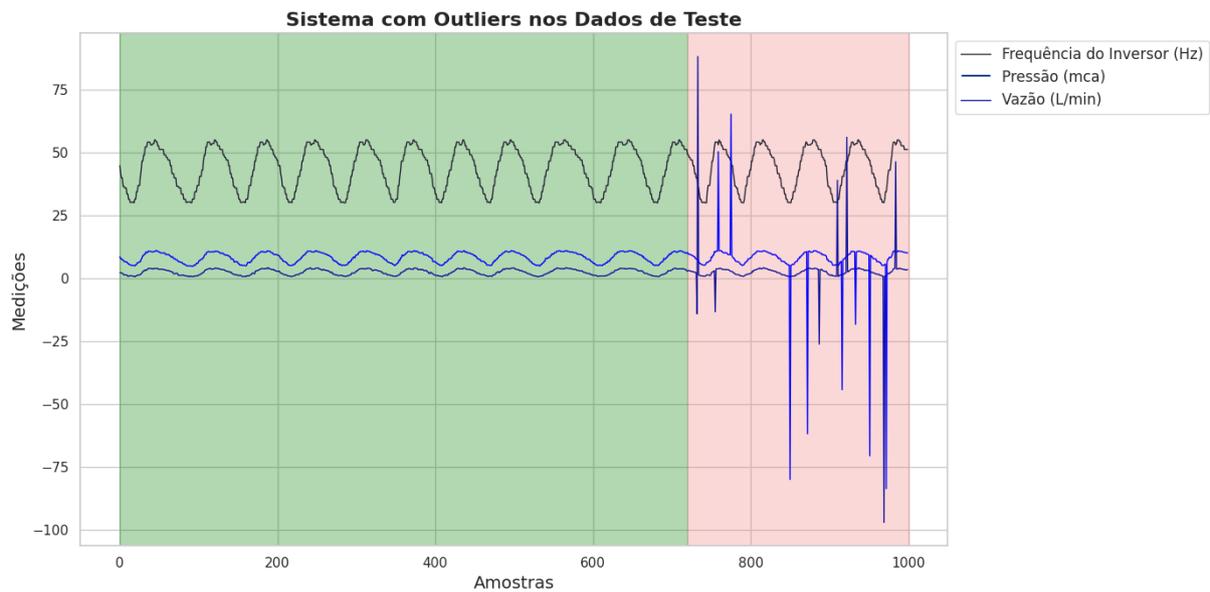
interferências ambientais ou anomalias operacionais, pode resultar em erros nas leituras que não são detectados durante a fase de treinamento. Esses *outliers* podem comprometer a capacidade da rede de generalizar e prever com precisão em situações reais.

Para abordar essa questão, foi criada uma estratégia de treinamento em que a rede foi alimentada com um conjunto de dados livre de *outliers*, garantindo que seu aprendizado fosse baseado em informações consistentes e confiáveis. No entanto, para avaliar a resiliência da RNA frente a dados atípicos, os *outliers* foram deliberadamente inseridos, a uma taxa de 3 %, apenas nos dados de teste. Essa abordagem, apresentada na Figura 31, em que a cor de fundo verde indica os dados utilizados para treinamento da RNA e em cor-de-rosa os dados utilizados para teste, permite observar como a rede se comporta quando confrontada com previsões em um cenário que reflete a imprevisibilidade das condições do mundo real.

Através dessa metodologia, é possível analisar a capacidade da RNA em detectar e lidar com *outliers* que possam ocorrer em situações práticas, ao mesmo tempo em que se assegura que o treinamento inicial não seja prejudicado por dados inconsistentes. Essa análise é crucial para entender a robustez do modelo e sua eficácia na previsão de demandas em sistemas de monitoramento e controle, onde a precisão das leituras é fundamental para a tomada de decisões informadas.

Os resultados obtidos com a aplicação da RNA, quando os *outliers* foram inseridos apenas nos dados de teste, estão apresentados graficamente na Figura 31. Essa abordagem permitiu avaliar o desempenho da rede em um cenário que simula a presença de erros de leitura que podem ocorrer no mundo real.

Figura 31 – Grafico do conjunto de dados com a presença de outliers gerados a uma taxa de 3 % apenas nos dados de teste.

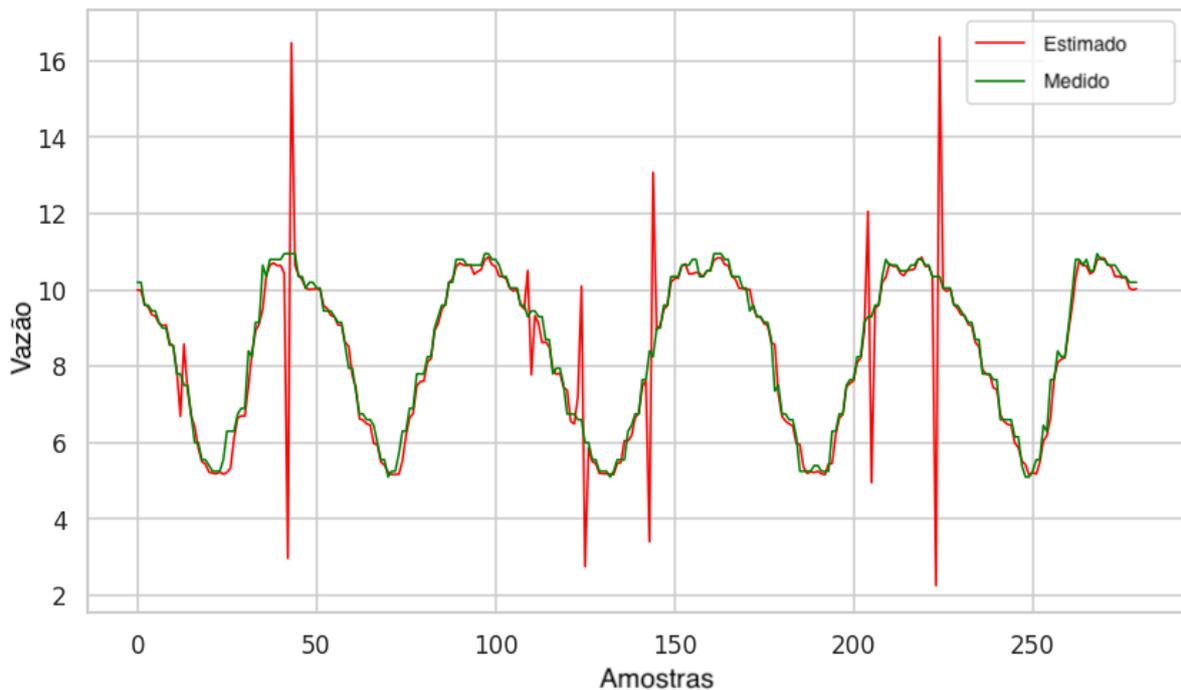


Fonte: Autoria Própria.

Os dados revelaram que, neste contexto, a RNA apresentou um erro quadrático médio (*loss*) de $3.8067e-04$. Esse valor indica um aumento significativo no erro em comparação com as situações anteriores, em que a rede foi treinada sem *outliers*. Esse resultado sugere que a presença de dados atípicos nos dados de teste impactou negativamente a capacidade da RNA de prever com precisão a demanda de consumo, refletindo a dificuldade que a rede enfrenta ao lidar com informações que não se alinham ao padrão de dados em que foi treinada.

Além disso, o erro percentual médio absoluto (MAPE) obtido foi de 5,16 %, indicando que as previsões da RNA apresentaram um desvio médio de aproximadamente 5,16 % em relação aos valores reais. Embora esse nível de erro seja aceitável em muitos contextos, ele representa uma deterioração na precisão em relação ao desempenho anterior, quando a RNA conseguiu manter um MAPE de 3,4 %.

Figura 32 – Grafico com a previsão do consumo com a presença de outliers gerados a uma taxa de 3 % apenas nos dados de teste.



Fonte: Autoria Própria.

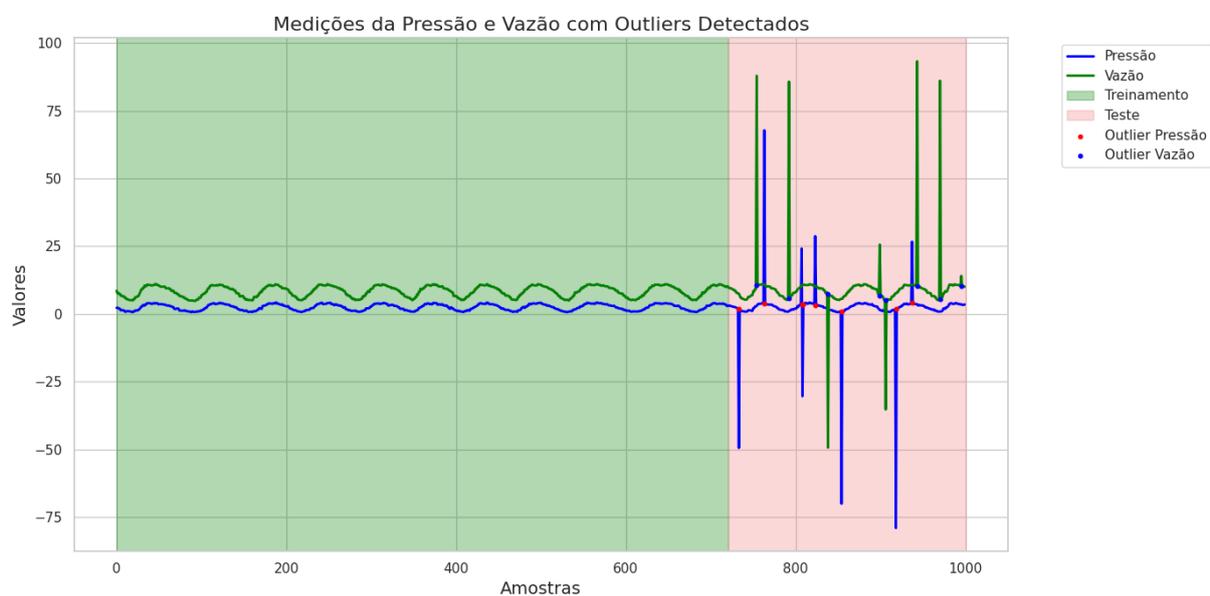
Esses resultados ressaltam a importância de considerar a imprevisibilidade dos *outliers* ao desenvolver modelos preditivos. A capacidade da RNA de manter sua precisão quando treinada com dados consistentes é uma vantagem, mas a introdução de *outliers* nos dados de teste evidencia a vulnerabilidade do modelo a erros não previstos. Portanto, a análise de desempenho em cenários com *outliers* é fundamental para compreender como melhorar a robustez da rede e garantir que ela possa se adaptar a condições variáveis, refletindo mais adequadamente a complexidade dos sistemas reais de monitoramento e controle.

Nesse contexto, a aplicação do método TEDA para detectar a presença de *outliers* é de suma importância. A detecção precoce de dados atípicos pode permitir ajustes no processo de treinamento e, conseqüentemente, melhorar a precisão das previsões da RNA. Este cenário está abordado na próxima seção, onde é apresentado como o TEDA pode ser implementado para otimizar o desempenho da rede diante de dados com *outliers*.

5.4.3 UTILIZAÇÃO DA RNA TREINADA SEM OUTLIERS: AVALIAÇÃO DO DESEMPENHO EM DADOS DE TESTE COM OUTLIERS, COM A APLICAÇÃO DO TEDA

Após a realização do treinamento da RNA com e sem *outliers*, além da aplicação da rede em dados de teste que continham *outliers*, observa-se agora a detecção dessas anomalias. Antes de aplicar os dados de teste com *outliers*, foi implementado o método TEDA para identificar as anomalias presentes no conjunto de dados. A resposta gráfica, que ilustra a detecção de *outliers* pelo TEDA, está apresentada na Figura 33, a qual apresenta uma visão clara da distribuição dos dados e dos pontos considerados atípicos.

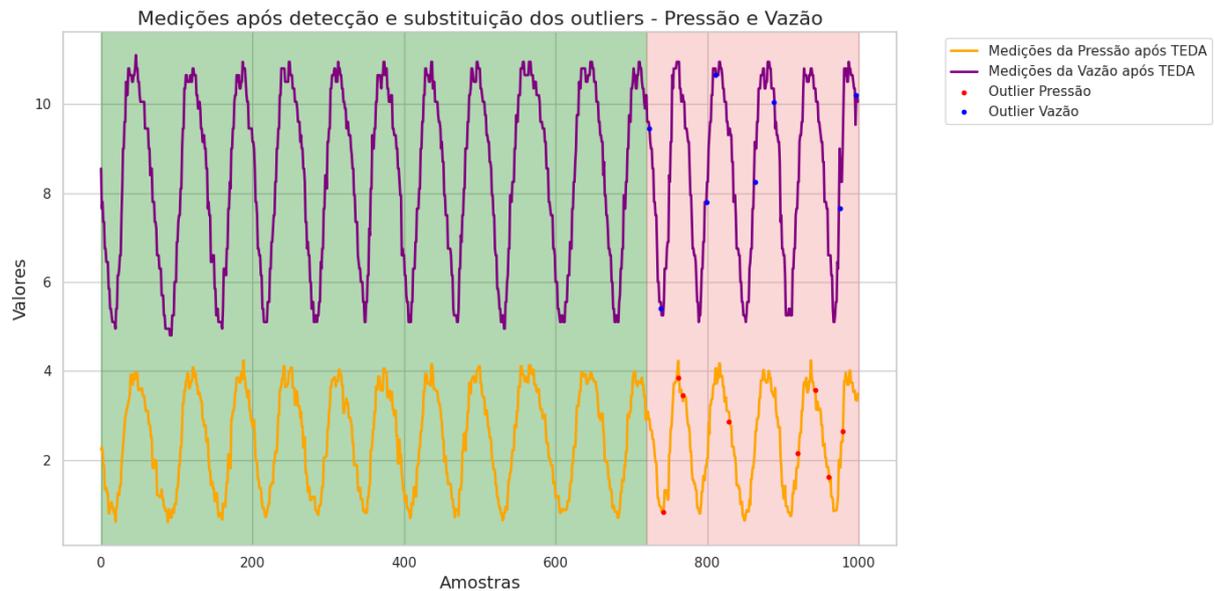
Figura 33 – Grafico do TEDA aplicado na detecção de outliers a uma taxa de 3 % apenas nos dados de teste.



Fonte: Autoria Própria.

Após a detecção dos *outliers* por meio do método TEDA, os dados identificados como atípicos foram removidos do conjunto de dados. Para garantir a continuidade e a integridade das informações, esses valores foram substituídos pela última leitura registrada antes da ocorrência das anomalias. Essa abordagem visa minimizar o impacto das anomalias nas previsões da RNA e melhorar a qualidade do conjunto de dados para análises futuras. A resposta gráfica do novo conjunto de dados, após essa substituição, pode ser visualizada na Figura 34, permitindo uma comparação clara entre os dados originais e os ajustados.

Figura 34 – Outliers substituídos após a detecção do TEDA.

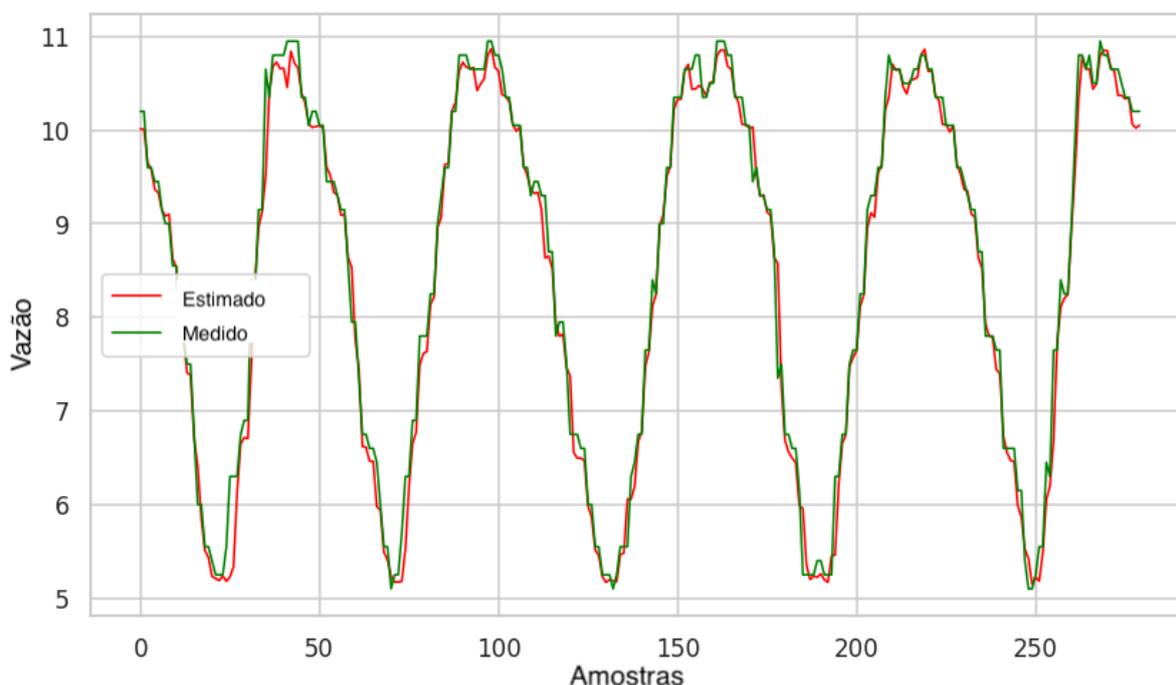


Fonte: Autoria Própria.

Os resultados obtidos com a aplicação da RNA após a remoção dos *outliers* e a substituição pelos últimos valores registrados estão apresentados graficamente na Figura 35. Essa etapa foi crucial para avaliar a eficácia da RNA em prever a demanda de consumo em um conjunto de dados mais limpo e livre de anomalias.

Com a implementação do TEDA e as correções realizadas, a RNA demonstrou um erro quadrático médio (*loss*) de $4.4329e-05$. Esse valor indica uma melhora significativa na performance da rede em relação aos testes anteriores, sugerindo que a RNA agora consegue capturar melhor os padrões dos dados. Além disso, o MAPE obtido foi de 3,04 %, refletindo um desvio médio de apenas 3,04 % em relação aos valores reais. Esse nível de precisão é um indicativo da eficácia do processo de detecção e tratamento de *outliers*, mostrando que as intervenções realizadas contribuíram para um desempenho mais robusto da rede.

Figura 35 – Aplicação da RNA após o uso do TEDA.



Fonte: Autoria Própria.

Esses resultados reforçam a importância da detecção e tratamento de *outliers* na construção de modelos preditivos mais confiáveis, evidenciando que a aplicação do TEDA foi um passo fundamental para otimizar a performance da RNA.

5.5 DETECÇÃO DE OUTLIERS E PREVISÃO DA VAZÃO EM TEMPO REAL

Para emular a aplicação deste trabalho em um sistema de distribuição em tempo de aquisição, foi desenvolvido um fluxograma que ilustra o funcionamento integrado do sistema. Nesse cenário, um ambiente *online* é criado para a aplicação simultânea do TEDA e da RNA.

Utilizando a tecnologia IoT, os dados de pressão e vazão são continuamente monitorados e transmitidos para um *broker* hospedado na AWS. O SCADA LTS desempenha um papel crucial, pois lê essas medições em tempo real e, ao detectar alterações, envia comandos ao inversor responsável pela pressurização da rede de distribuição do laboratório.

A variação nas medições de pressão e vazão nos pontos de medição possibilita a aquisição precisa dessas informações pelo sistema IoT. Quando os dados são enviados

ao *broker*, um processo de verificação é realizado para identificar possíveis *outliers*. Caso um dado atípico seja detectado, ele é automaticamente substituído pela última leitura registrada, garantindo que o sistema mantenha a integridade dos dados.

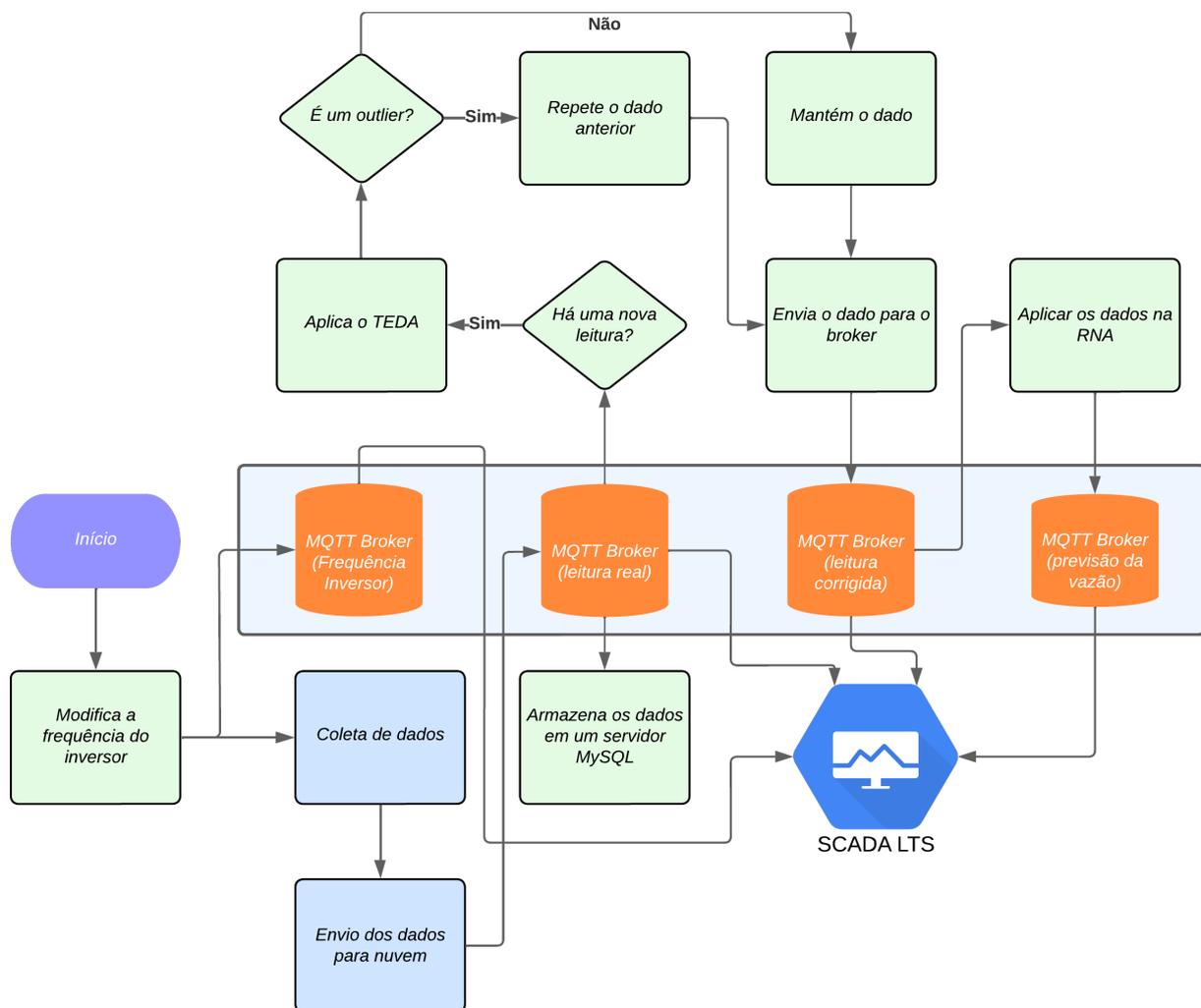
A RNA, por sua vez, está em constante monitoramento, analisando os novos valores disponíveis e realizando previsões para as duas vazões futuras. Essas previsões são armazenadas no *broker* e podem ser visualizadas graficamente no SCADA LTS, proporcionando um *feedback* visual valioso sobre o desempenho do sistema.

A Figura 36 ilustra o fluxograma deste cenário implementado no LENHS, detalhando a interação entre os componentes do sistema. A aplicação do TEDA para a detecção de *outliers*, aliada à capacidade preditiva da RNA, não só aumenta a precisão das leituras, mas também permite ajustes dinâmicos em tempo real, adaptando o sistema a condições variáveis e garantindo uma operação mais confiável.

Esse conjunto de ferramentas e processos evidencia a importância da integração de tecnologias IoT, algoritmos de detecção de anomalias e redes neurais para um monitoramento eficiente e em tempo real em sistemas de distribuição.

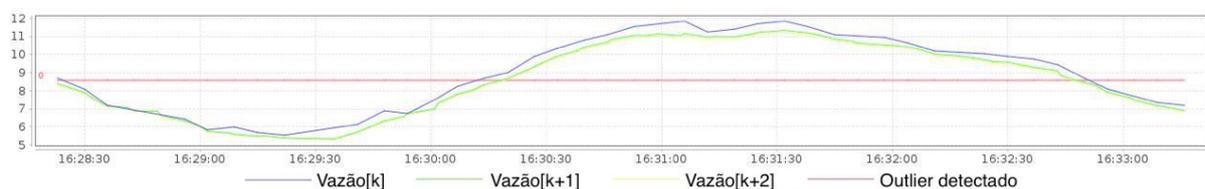
Com a utilização do SCADA LTS é possível observar o comportamento do sistema em tempo real, apresentado na Figura 37. Nesta figura é possível observar as curvas de vazão, sendo a de azul a vazão medida, e a curva de verde a previsão usando a metodologia proposta. Adicionalmente, o algoritmo TEDA encontra-se verificando os *outliers*.

Figura 36 – Fluxograma Experimental do TEDA e RNA executando em tempo real.



Fonte: Autoria Própria.

Figura 37 – TEDA e RNA *online*s.



Fonte: Autoria Própria.

Na Figura 37 é possível observar o comportamento dos sistema sendo executado em tempo real. A linha em vermelho é utilizada para indicar a detecção de *outliers* no sistema. Vazão [k+1] e Vazão [k+2] apresentam as leituras futuras, geradas com a RNA.

6 CONCLUSÕES

Este trabalho abordou a aplicação do TEDA e da RNA no monitoramento de sistemas de distribuição de água. Por meio da implementação e análise dos cenários propostos foi possível observar os efeitos significativos da detecção de *outliers* e das previsões de vazão na eficiência operacional desses sistemas.

Os resultados indicaram que a aplicação do TEDA permitiu a identificação e correção dos *outliers*, melhorando a qualidade dos dados utilizados pela RNA para suas previsões. Quando a RNA foi treinada sem a presença de *outliers*, foi possível obter um MAPE de 3,15 %, evidenciando uma previsão precisa da demanda de consumo. No entanto, ao introduzir *outliers* apenas nos dados de teste, o MAPE aumentou para 5,16 %, indicando uma redução na precisão das previsões. Por outro lado, com a aplicação do TEDA para detectar *outliers*, a RNA conseguiu melhorar seu desempenho, resultando em um MAPE de 3,04 %, demonstrando a eficácia da detecção de *outliers*.

No cenário atual, em que as perdas nos sistemas de distribuição são significativamente altas, a implementação dessas tecnologias se torna ainda mais crucial. Como apresentado no capítulo introdutório, as perdas podem ocorrer devido a vazamentos, falhas de leitura de sensores ou variações inesperadas nas medições. A capacidade do TEDA de detectar e mitigar a influência de *outliers*, combinada com a previsão de vazão da RNA, oferece uma solução para melhorar a gestão e a eficiência de sistemas de distribuição de água. Essa abordagem não só reduz as perdas financeiras, mas também contribui para uma utilização mais sustentável dos recursos hídricos, que são cada vez mais escassos.

Frente ao exposto, com base nos resultados alcançados nessa pesquisa, abre-se um leque de aplicações como a simulação de vazamentos no sistema de distribuição com objetivo, gerando alertas para o operador a partir do erro apresentada pela diferença entre a vazão prevista e medida. Nesse contexto, a utilização do TEDA com dados de vazamentos pode ser investigado. A automação da atuação no inversor por meio de um *script* de controle no SCADA BR também permite estabelecer um sistema de controle fechado a partir da diferença entre a vazão estimada e medida no ponto monitorado, o qual possibilita uma nova pesquisa do uso de um sistema de controle inteligente para monitoramento de sistemas de distribuição. Em síntese, para trabalhos futuros sugere-se:

- Implementar o controle do consumo mediante a abertura de uma válvula no ponto monitorado;
- Avaliar do uso rede neural nos dados adquiridos durante a simulação com vazamentos na bancada experimental;

- Aplicar um sistema de controle automático de atuação no conjunto motor-bomba a partir da diferença entre a vazão estimada e medida;
- Inserir pontos de alerta para o operador com base na identificação de vazamentos no sistema de distribuição.

Por fim, este trabalho contribui de forma direta para a ampliação dos conhecimentos adquiridos ao longo da pós-graduação em engenharia elétrica do autor, sobretudo na área de internet das coisas e aplicação de técnicas de inteligência artificial. Desta maneira o conhecimento utilizado nesse trabalho irá ser de grande utilidade na vida profissional, especialmente no setor público, área de atuação do autor.

REFERÊNCIAS

ALMAS, M. S. et al. Open source scada implementation and pmu integration for power system monitoring and control applications. In: *2014 IEEE PES General Meeting / Conference Exposition*. [S.l.: s.n.], 2014. p. 1–5. Citado na página 39.

ANGELOV, P. *Autonomous learning systems: from data streams to knowledge in real-time*. [S.l.]: John Wiley & Sons, 2012. Citado na página 54.

ANGELOV, P. Outside the box: an alternative data analytics framework. *Journal of Automation Mobile Robotics and Intelligent Systems*, v. 8, n. 2, p. 29–35, 2013. Citado 3 vezes nas páginas 27, 28 e 29.

ARTHI., R.; NISHUTHAN, S.; VIGNESH, L. D. Smart agriculture system using iot and ml. In: *2023 International Conference on Signal Processing, Computation, Electronics, Power and Telecommunication (IConSCEPT)*. [S.l.: s.n.], 2023. p. 1–6. Citado na página 39.

BERNIERI, A.; BETTA, G.; LIGUORI, C. On-line fault detection and diagnosis obtained by implementing neural algorithms on a digital signal processor. *IEEE Transactions on Instrumentation and Measurement*, v. 45, n. 5, p. 894–899, 1996. Citado na página 29.

BRASIL. *Diagnóstico dos Serviços de Água e Esgoto*. 2019. <https://www.gov.br/mdr/pt-br/assuntos/saneamento/snis/diagnosticos-antiores-do-snis/agua-e-esgotos-1/2019/2-Diagnostico_SNIS_AE_2019_Republicacao_31032021.pdf>. [Accessed 08-01-2024]. Citado na página 22.

BRASIL. *Diagnóstico dos Serviços de Água e Esgoto*. 2023. <https://www.gov.br/cidades/pt-br/aceso-a-informacao/acoes-e-programas/saneamento/snis/produtos-do-snis/diagnosticos/DIAGNOSTICO_TEMATICO_VISAO_GERAL_AE_SNIS_2023.pdf>. [Accessed 09-01-2024]. Citado 2 vezes nas páginas 21 e 22.

BRUNI, J. C. A água e a vida. *Tempo Social*, FapUNIFESP (SciELO), v. 5, n. 1–2, p. 53–65, dez. 1993. ISSN 1809-4554. Disponível em: <<http://dx.doi.org/10.1590/ts.v5i1/2.84942>>. Citado na página 14.

BYER, D. Real-time detection of intentional chemical contamination - in the distribution system. *JOURNAL AMERICAN WATER WORKS ASSOCIATION*, WILEY, 111 RIVER ST, HOBOKEN 07030-5774, NJ USA, v. 97, n. 7, p. 130+, JUL 2005. ISSN 2164-4535. Citado na página 18.

CAMINOS, R. A. O camponês. In: DONADONI, S. (Ed.). *O Homem Egípcio*. Lisboa: Editorial Presença, 1994. Citado na página 14.

CASTRO, C. Nunes de. Introduçã - seguranã hídrica, problemas complexos, participaã social: O caso do plano nacional de seguranã hídrica. *Livros*, Instituto de Pesquisa Economica Aplicada - IPEA, n. Água, problemas complexos, p. 11–23, mar. 2022. Disponível em: <<http://dx.doi.org/10.38116/9786556350318intro>>. Citado na página 14.

- CHEN, R.; WANG, M.; LAI, Y. Analysis of the role and robustness of artificial intelligence in commodity image recognition under deep learning neural network. *PLOS ONE*, Public Library of Science (PLoS), v. 15, n. 7, p. e0235783, jul. 2020. Disponível em: <<https://doi.org/10.1371/journal.pone.0235783>>. Citado na página 31.
- COMINATO, C. et al. Water demand profile before and during covid-19 pandemic in a brazilian social housing complex. *Heliyon*, Elsevier BV, v. 8, n. 8, p. e10307, ago. 2022. ISSN 2405-8440. Disponível em: <<http://dx.doi.org/10.1016/j.heliyon.2022.e10307>>. Citado 2 vezes nas páginas 47 e 48.
- COSTA, B. S. J. et al. Online fault detection based on typicality and eccentricity data analytics. In: IEEE. *2015 International Joint Conference on Neural Networks (IJCNN)*. [S.l.], 2015. p. 1–6. Citado na página 29.
- De Oliveira, L. et al. Smart water management: a self-sufficient iot-based application for pressure and flow monitoring in water distribution systems. In: *INSCIT 2023 - 7th International Symposium on Instrumentation Systems, Circuits and Transducers, Proceedings*. [S.l.]: Institute of Electrical and Electronics Engineers Inc., 2023. (INSCIT 2023 - 7th International Symposium on Instrumentation Systems, Circuits and Transducers, Proceedings). Publisher Copyright: © 2023 IEEE.; 7th International Symposium on Instrumentation Systems, Circuits and Transducers, INSCIT 2023 ; Conference date: 28-08-2023 Through 01-09-2023. Citado 3 vezes nas páginas 37, 38 e 50.
- DOVOEDO, Y. H.; CHAKRABORTI, S. Boxplot-based outlier detection for the location-scale family. *Communications in Statistics - Simulation and Computation*, Informa UK Limited, v. 44, n. 6, p. 1492–1513, dez. 2014. ISSN 1532-4141. Disponível em: <<http://dx.doi.org/10.1080/03610918.2013.813037>>. Citado na página 27.
- DUTRA, R. H. d. A.; OLIVEIRA, A. L. d. Utilização de válvulas redutoras de pressão no controle de perdas em redes de abastecimento de água. *Revista DAE*, Revista DAE, v. 65, n. 208, p. 122–134, 2017. ISSN 0101-6040. Disponível em: <<http://dx.doi.org/10.4322/dae.2017.011>>. Citado na página 23.
- EBISI, F. et al. Machine learning schemes for leak detection in iot-enabled water transmission system. In: *2023 International Conference on IT Innovation and Knowledge Discovery (ITIKD)*. [S.l.: s.n.], 2023. p. 1–7. Citado na página 19.
- FABIO, I. S.; MELO, C. L. Silva de; PINA, I. F. B. Scada system to performance index calculation using reliability engineering tools in a modular production system. In: *2015 CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON)*. [S.l.: s.n.], 2015. p. 747–750. Citado na página 39.
- FIGUEIREDO, A. A. d. O. et al. Avaliação e potencial de redução de perdas de água em cidades do estado de pernambuco com escassez hídrica e abastecimento intermitente. *Journal of Environmental Analysis and Progress*, Journal of Environmental Analysis and Progress - JEAP, v. 8, n. 3, p. 212–225, ago. 2023. ISSN 2525-815X. Disponível em: <<http://dx.doi.org/10.24221/jeap.8.3.2023.5465.212-225>>. Citado 4 vezes nas páginas 22, 23, 24 e 26.
- FLORES, T. K. S. *Controle Adaptativo Utilizando o Critério da Máxima Correntropia Aplicado a um Sistema de Bombeamento de Água*. Dissertação

- (Mestrado) — Universidade Federal da Paraíba, 2021. Disponível em: <https://repositorio.ufpb.br/jspui/handle/123456789/22801?locale=pt_BR>. Acesso em: 03 jan. 2024. Citado 2 vezes nas páginas 33 e 34.
- GALAITSI, S. et al. Intermittent domestic water supply: A critical review and analysis of causal-consequential pathways. *Water*, MDPI AG, v. 8, n. 7, p. 274, jun. 2016. ISSN 2073-4441. Disponível em: <<http://dx.doi.org/10.3390/w8070274>>. Citado na página 23.
- GARG, P. et al. Indigenous development of water quality monitoring system for urban areas using IoT ml. In: *2023 International Conference on Computer, Electronics Electrical Engineering their Applications (IC2E3)*. [S.l.: s.n.], 2023. p. 1–6. Citado na página 19.
- GATIAL, E.; BALOGH, Z.; HLUCHY, L. Concept of energy efficient esp32 chip for industrial wireless sensor network. In: *2020 IEEE 24th International Conference on Intelligent Engineering Systems (INES)*. IEEE, 2020. Disponível em: <<http://dx.doi.org/10.1109/INES49302.2020.9147189>>. Citado na página 37.
- KINGMA, D. P.; BA, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. Citado na página 32.
- KRIEGESKORTE, N.; GOLAN, T. Neural network models and deep learning. *Current Biology*, Elsevier BV, v. 29, n. 7, p. R231–R236, abr. 2019. Disponível em: <<https://doi.org/10.1016/j.cub.2019.02.034>>. Citado na página 31.
- LI, S. et al. Review of image classification algorithms based on convolutional neural networks. *Remote Sensing*, MDPI, v. 13, n. 22, p. 4712, 2021. Citado na página 32.
- LIIMATAINEN, K. et al. Convolutional neural network-based artificial intelligence for classification of protein localization patterns. *Biomolecules*, MDPI AG, v. 11, n. 2, p. 264, fev. 2021. Disponível em: <<https://doi.org/10.3390/biom11020264>>. Citado na página 31.
- LUCENA, K.; NASCIMENTO, F. G. D. Análise da operação de um subsistema do sistema de abastecimento de água de campina grande-pb. *Revista Principia - Divulgação Científica e Tecnológica do IFPB*, Instituto Federal de Educacao, Ciencia e Tecnologia da Paraíba, v. 1, n. 45, p. 180, jul. 2019. ISSN 1517-0306. Disponível em: <<http://dx.doi.org/10.18265/1517-03062015v1n45p180-187>>. Citado na página 22.
- LUCENA, K. F. M. d. Aplicação de modelos de otimização na operação de sistemas elevatórios em redes de abastecimento de água. *Revista Principia - Divulgação Científica e Tecnológica do IFPB*, Instituto Federal de Educacao, Ciencia e Tecnologia da Paraíba, v. 1, n. 51, p. 166, set. 2020. ISSN 1517-0306. Disponível em: <<http://dx.doi.org/10.18265/1517-0306a2020v1n51p166-180>>. Citado 2 vezes nas páginas 15 e 20.
- LUNGUINHO, L. L. Domínio territorial do abastecimento de Água na paraíba: municipalização × estadualização. *IV SEMAGEO – Semana de Geografia da UFPB*, p. 585–607, 2007. Disponível em: <<http://www.geociencias.ufpb.br/leppan/gepat/files/gepat010.pdf>>. Citado na página 22.
- MIQUELIN, L. M.; SOLER, E. M. Modelo matemático para otimização da operação de sistemas de abastecimento de água. *C.Q.D. – Revista Eletrônica*

- Paulista de Matemática*, C.Q.D.- Revista Eletronica Paulista de Matematica, v. 7, p. 88–96, dez. 2016. ISSN 2316-9664. Disponível em: <<http://dx.doi.org/10.21167/cqdv07ermac201623169664lmmems8896>>. Citado na página 21.
- MOHAMMADPOUR, R. et al. Prediction of water quality index in constructed wetlands using support vector machine. *Environmental Science and Pollution Research*, v. 22, n. 8, p. 6208–6219, 2015. ISSN 1614-7499. Disponível em: <<https://doi.org/10.1007/s11356-014-3806-7>>. Citado na página 18.
- NETO, A. B.; BONINI, C. D. S. B. Redes neurais artificiais: ApresentaundefinedÃO e utilizaundefinedÃO do algoritmo perceptron em biosistemas. *Revista Brasileira de Engenharia de Biosistemas*, Universidade Estadual Paulista - Campus de Tupa, v. 4, n. 2, p. 87–95, nov. 2010. ISSN 1981-7061. Disponível em: <<http://dx.doi.org/10.18011/bioeng2010v4n2p87-95>>. Citado 2 vezes nas páginas 30 e 31.
- NUNES, L.; TEIXEIRA, R. Xvi congresso brasileiro de águas subterrâneas e xvii encontro nacional de perfuradores de poços. In: . [s.n.], 2010. Disponível em: <<https://api.semanticscholar.org/CorpusID:133643532>>. Citado na página 15.
- OLIVEIRA, L. D. d. Otimização de sistemas de abastecimento de Água utilizando técnicas de inteligência artificial: Sensores iot para o monitoramento de pressão em uma rede de abastecimento de Água. 2023. Citado 3 vezes nas páginas 36, 48 e 51.
- ORELLANA, A. et al. Contribuição ao planejamento de reabilitação de redes de distribuição de água. *Ribagua*, Informa UK Limited, v. 5, n. 2, p. 79–91, jul. 2018. ISSN 2529-8968. Disponível em: <<http://dx.doi.org/10.1080/23863781.2018.1495991>>. Citado na página 20.
- OWENS, M. *The definitive guide to SQLite*. [S.l.: s.n.], 2006. Citado na página 51.
- PATAÇA, C. A internet das coisas: Tipologias, protocolos e aplicações. *Law, State and Telecommunications Review*, Biblioteca Central da UNB, v. 13, n. 2, p. 198–220, set. 2021. ISSN 1984-9729. Disponível em: <<http://dx.doi.org/10.26512/lstr.v13i2.32773>>. Citado 3 vezes nas páginas 15, 24 e 25.
- PATEL, D.; MAITI, C.; MUTHUSWAMY, S. Real-time performance monitoring of a cnc milling machine using ros 2 and aws iot towards industry 4.0. In: *IEEE EUROCON 2023 - 20th International Conference on Smart Technologies*. [S.l.: s.n.], 2023. p. 776–781. Citado na página 38.
- PEREIRA, F. M. d. S.; CUNICO, C. IdentificaundefinedÃO e avaliaundefinedÃO das condiundefinedões de abastecimento de Água e esgotamento sanitário nos diferentes graus de vulnerabilidade socioambiental do município de João pessoa, pb. *Revista de Geociências do Nordeste*, Universidade Federal do Rio Grande do Norte - UFRN, v. 6, n. 1, p. 42–48, jun. 2020. ISSN 2447-3359. Disponível em: <<http://dx.doi.org/10.21680/2447-3359.2020v6n1ID19066>>. Citado na página 22.
- PETERSON, T. J.; WESTERN, A. W.; CHENG, X. The good, the bad and the outliers: automated detection of errors and outliers from groundwater hydrographs. *Hydrogeology journal*, Springer Berlin Heidelberg, Berlin/Heidelberg, v. 26, n. 2, p. 371–380, 2018. ISSN 1431-2174. Citado na página 26.

- Presidência da República. *LEI Nº 9.433, DE 8 DE JANEIRO DE 1997*. 1997. <https://www.planalto.gov.br/ccivil_03/leis/19433.htm>. Acesso em: 12 de fevereiro de 2024. Citado na página 14.
- PUVIYARASI, B. et al. Water quality monitoring and cleaning system. In: *2024 International Conference on Communication, Computing and Internet of Things (IC3IoT)*. [S.l.: s.n.], 2024. p. 1–5. Citado na página 19.
- SALVADOR, H. G.; CUNHA, A. M. D.; CORRÊA, C. S. Vedalogic: um método de verificação de dados climatológicos apoiado em modelos minerados. *Revista brasileira de meteorologia*, FapUNIFESP (SciELO), v. 24, n. 4, p. 448–460, 2009. ISSN 0102-7786. Citado na página 27.
- SARITHA, G. et al. Water quality monitoring system using iot. In: *2023 Eighth International Conference on Science Technology Engineering and Mathematics (ICONSTEM)*. [S.l.: s.n.], 2023. p. 1–5. Citado na página 19.
- SILVA, B. C. O. D.; NÓBREGA, R. S. Dados discrepantes ou outliers: avaliação da quadra chuvosa do semiárido do rn, brasil. *Revista Entre-Lugar*, Universidade Federal de Grande Dourados, v. 12, n. 24, p. 346–366, 2021. ISSN 2176-9559. Citado na página 27.
- SOUSA, M. de. *Desenvolvimento de Sistema de Coleta e Transmissão de Dados utilizando Drones e Sensores IoT*. Trabalho de Conclusão de Curso — Universidade Federal da Paraíba, João Pessoa, PB, oct 2024. Citado na página 34.
- SOUSA, R. A. de; PROFA; BORIN, J. Iot aplicado ao monitoramento inteligente de distribuição de Água. In: . [s.n.], 2019. Disponível em: <<https://api.semanticscholar.org/CorpusID:202740275>>. Citado 3 vezes nas páginas 24, 25 e 26.
- TUCCI, C. E. M. Águas urbanas. *Estudos Avançados*, FapUNIFESP (SciELO), v. 22, n. 63, p. 97–112, 2008. ISSN 0103-4014. Disponível em: <<http://dx.doi.org/10.1590/S0103-40142008000200007>>. Citado na página 20.
- UPADHYAY, D.; SAMPALLI, S. Scada (supervisory control and data acquisition) systems: Vulnerability assessment and security recommendations. *Computers & Security*, v. 89, p. 101666, 2020. ISSN 0167-4048. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0167404819302068>>. Citado na página 39.
- VELAYUDHAN, N. K. et al. Iot-enabled water distribution systems—a comparative technological review. *IEEE Access*, v. 10, p. 101042–101070, 2022. Citado 3 vezes nas páginas 15, 18 e 26.
- VENKATAANUSHA, P. et al. Detecting outliers in high dimensional data sets using z-score methodology. *International Journal of Innovative Technology and Exploring Engineering*, Blue Eyes Intelligence Engineering and Sciences Engineering and Sciences Publication - BEIESP, v. 9, n. 1, p. 48–53, nov. 2019. ISSN 2278-3075. Disponível em: <<http://dx.doi.org/10.35940/ijitee.A3910.119119>>. Citado na página 27.
- VENTURA, T. M. et al. Disponibilização de métodos de tratamento de dados ambientais por meio de um web service. *Revista de Sistemas, Cibernética e Informática*, International Institute of Informatics and Cybernetics, v. 14, n. 2, p. 33–38, 2017. ISSN 1690-8627. Citado 2 vezes nas páginas 26 e 43.

ZANCAN, C.; PASSADOR, J. L.; PASSADOR, C. S. Modelos de inteligência artificial na gestão de consórcios intermunicipais brasileiros. *Revista Gestão e Desenvolvimento*, Associação Pro-Ensino Superior em Novo Hamburgo, v. 20, n. 2, p. 80–123, set. 2023. ISSN 1807-5436. Disponível em: <<http://dx.doi.org/10.25112/rgd.v20i2.3424>>. Citado na página 15.

ZHANG, J. et al. A real-time anomaly detection algorithm for water quality data using dual time-moving windows. In: *2017 Seventh International Conference on Innovative Computing Technology (INTECH)*. [S.l.: s.n.], 2017. p. 36–41. Citado na página 18.

ZUFFO, A. Incorporação de matemática fuzzy em métodos multicriteriais para descrever critérios subjetivos em planejamento de recursos hídricos: Fuzzy - cp e fuzzy - cgt. *Revista Brasileira de Recursos Hídricos*, FapUNIFESP (SciELO), v. 16, n. 4, p. 29–40, 2011. ISSN 2318-0331. Disponível em: <<http://dx.doi.org/10.21168/rbrh.v16n4.p29-40>>. Citado na página 24.
