Análise de Falhas em Rede de Distribuição de Água

Relatório de Estágio

Francisco José Valente Ferreira

2019113494

Orientadores:

Mateus Mendes [mmendes@isec.pt](mailto:mmendes@isec.pt)

José Torres Farinha [tfarinha@isec.pt](mailto:tfarinha@isec.pt)

Samuel Martinho [a21170106@isec.pt](mailto:a21170106@isec.pt)

Licenciatura em Engenharia Informática

Ramo de Desenvolvimento de Aplicações

Instituto Politécnico de Coimbra

Instituto Superior de Engenharia de Coimbra

Março de 2023

# Agradecimentos

…

# Resumo

… este documento pretende apenas servir de linha orientadora, de uso opcional, para a elaboração dos relatórios de estágio… a estrutura e formatação finais dos relatórios deverão ser definidos caso a caso pelos alunos e respetivos orientadores…

***Palavras‑chave:*** *Big Data, Serialização, Spring Boot.*

# Abstract

As redes de distribuição de água potável são infraestruturas essenciais de gestão e transporte de um dos bens essenciais e escassos. A fuga de água das redes constitui uma perda enorme tanto em termos de manutenção como desperdício de água de consumo próprio. Este problema já foi explorado com bastantes abordagens, umas mais avançadas em países mais desenvolvidos através de dispositivos acústicos, mas em países menos desenvolvidos é apenas visível quando a fuga de água aparece na superfície. O uso de software e modelação de dados obtidos dessas redes permite uma abordagem mais barata, com modelos que podem estimar fugas e evitar maiores perdas. Este relatório vai exemplificar com algoritmos de análise e visualização de dados.

***Keywords:***

**Índice**

[Agradecimentos i](#_Toc68777782)

[Resumo iii](#_Toc68777783)

[Abstract v](#_Toc68777784)

[Índice de figuras ix](#_Toc68777785)

[Índice de tabelas xi](#_Toc68777786)

[Acrónimos xiii](#_Toc68777787)

[1 Introdução 1](#_Toc68777788)

[1.1 Entidade de acolhimento 1](#_Toc68777789)

[1.2 Instituto Superior de Engenharia de Coimbra 1](#_Toc68777790)

[1.3 Objetivos e plano de trabalhos 1](#_Toc68777791)

[1.4 Estrutura do relatório 2](#_Toc68777792)

[2 Mais um capítulo 5](#_Toc68777793)

[3 Mais um capítulo 7](#_Toc68777794)

[4 Mais um capítulo 9](#_Toc68777795)

[4.1 Requisitos 9](#_Toc68777796)

[4.2 Arquitetura 9](#_Toc68777797)

[4.3 Tecnologias e ferramentas usadas 9](#_Toc68777798)

[4.4 Implementação 9](#_Toc68777799)

[4.5 Testes e validação 9](#_Toc68777800)

[4.5.1 Método A 9](#_Toc68777801)

[4.5.2 Método B 10](#_Toc68777802)

[5 Mais um capítulo 11](#_Toc68777803)

[6 Conclusões e trabalho futuro 13](#_Toc68777804)

[6.1 Conclusões 13](#_Toc68777805)

[6.2 Trabalho Futuro 13](#_Toc68777806)

[Referências 15](#_Toc68777807)

[Anexos 17](#_Toc68777808)

[Anexo A: Proposta de Estágio A-1](#_Toc68777809)

[Anexo B: Especificação de Requisitos B-1](#_Toc68777810)

# Índice de figuras

[Figura 1.1 Explicação da figura 1](#_Toc68777811)

# Índice de tabelas

[Tabela 4.1 – Tecnologias e ferramentas usadas 9](#_Toc68777812)

# Acrónimos

…

…

# Introdução

As redes de distribuição de água sofrem em todo o mundo perdas água. Sendo este um problema que afeta um bem essencial,

…a referir no texto as figuras, quando existirem, conforme apresentado na Figura 1.1…

[](https://www.google.pt/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=0ahUKEwiW-4ij3_zSAhXB0RoKHYa2AJkQjRwIBw&url=http://mariohenriquemartins.com.br/enquadramento-o-que-voce-destaca-em-seu-argumento/&psig=AFQjCNEBh3yreV0BxghV2mg3qhOIzvt7pA&ust=1490912295373271)

Figura . Explicação da figura

… as subsecções seguintes são apenas exemplificativas…

## Entidade de acolhimento

… a usar referências quando se mencionar algo dito por outra pessoa [1]…

## Instituto Superior de Engenharia de Coimbra

…

## Objetivos e plano de trabalhos

… referir a proposta original, que se encontra em anexo, e apresentar o que realmente aconteceu…

## Estrutura do relatório

…

…

# Estado da Arte

Na análise de sistemas de gestão de água e deteção de falhas existe já imensa literatura, as abordagens mais encontradas foram sistemas de alerta e sistemas preditivos de localização de fugas. Neste capítulo vamos falar sobre as principais causas das fugas de água encontradas na literatura bem como várias abordagens ao problema e os seus resultados.

## Vulnerabilidades principais nas Redes de Distribuição de água

As fugas de água são na maior parte das vezes detetadas nas superfícies das ruas pelos cidadãos que por fim avisam as entidades. Em todo o caso, estas são as que porventura tem menor desperdício visto que rapidamente são arranjadas, as principais perdas de água são as fugas “silenciosas” que vão largando água durante muito tempo, o que leva a um desperdício enorme, estas são as que são mais difíceis de detetar.

## Métodos de deteção de fugas

Os primeiros são os métodos locais, neste grupo as RDA são divididas em várias partes e para cada parte a fuga é procurada usando vários equipamentos, equipamentos acústicos, medições de balanço de massa, radares que penetram no subsolo, gás marcador etc, estes métodos são caros requerem mão de obra complexa e por norma de pouca precisão.

Os métodos globais usam sistemas que recolhem informação sobre caudais, pressões, depois são processados por um algoritmo de análise que encontram zonas críticas. A análise destes sistemas normalmente tem falhas associadas pois o dado obtido tem muitas variáveis associadas como o estado do material, caudais, pressões. Muitos sistemas também ficam prejudicados por falta de sensores nas ligações o que leva a um custo monetário associado. Normalmente são estimados alguns valores o que leva a uma discrepância entre o modelo gerado e o modelo real.

### Métodos de deteção e alerta

No estilo de abordagem mais direta [2], sistemas de direta aplicação que usam um microcontrolador que está sempre a ler dados de caudais da rede, ao detetar que dois sensores adjacentes têm valores diferentes maiores que um determinado valor o microcontrolador emite um sinal para uma válvula solenoide para travar a corrente água. Estes sistemas requerem muitas ligações elétricas o que diminui a capacidade para grandes redes de água, configurações wireless do sistema podem permitir melhorias e aumento das áreas mas implica o aumento do custo do sistema.

### Sistemas de modelação de dados e predição

#### Métodos de Inteligência artificial

Vários artigos publicados revelam que métodos de IA são um bom caminho para o estudo de fugas de água. Por exemplo [3], através do uso do Software EPANET para gerar dados de redes de água e cenários de fugas. Os dados gerados serviram para treinar e testar seis métodos de IA. Inputs e Outputs, foram usados três inputs de caudais cada um de uma localidade diferente e pontos de observação de pressão (5) e o Output é a localização da zona. As principais ligações entre os inputs e a importância na zona de fuga são demonstradas na tabela.

Tabela Relação importância de caudal para respetiva zona

|  |  |  |  |
| --- | --- | --- | --- |
| ****Zona de Fuga**** | ****Caudal 1**** | ****Caudal 2**** | ****Caudal 3**** |
| ****1**** | **Forte** | **Médio** | **Baixo** |
| ****2**** | **Médio** | **Médio** | **Baixo/ Médio** |
| ****3**** | **Baixo** | **Médio** | **Forte** |
| ****4**** | **Baixo** | **Forte** | **Baixo/ Médio** |
| ****5**** | **Médio** | **Médio** | **Baixo/ Médio** |

As capacidades de seis métodos foram comparadas e a sua avaliação foi calculada através de , *Accuracy*, *Precision*, Recall, F1-Score.

O método da *logistic regression* é usado para classificação binária através de determinação de probabilidades para cada classe usando as suas características, a *decision tree* classifica um dataset a partir de ramos de uma arvore e cada nó interno é uma questão que avalia uma determinada característica.

Para avaliar os resultados de cada método são usadas 4 métricas, *Accuracy*, *Precision*, Recall, F1-Score. Na classificação múltiplo, a função de *accuracy* calcula a precisão do subconjunto: o conjunto de rótulos previsto para uma amostra deve corresponder exatamente ao conjunto de rótulos correspondente nos dados reais, mostra o valor de acertos.

A precisão é a razão tp / (tp + fp) onde tp é o número de verdadeiros positivos e fp o número de falsos positivos. A precisão é intuitivamente a capacidade de o classificador não rotular como positiva uma amostra que é negativa. O melhor valor é 1 e o pior valor é 0.

O score F1 pode ser interpretado como uma média harmônica da *precision* e do *recall*, onde um *score* F1 atinge seu melhor valor em 1 e o pior score em 0. A contribuição relativa da *precision* e do *recall* para o escore F1 são iguais.

Tabela Métodos supervisionados

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ****Metodo**** | ****Accuracy**** | ****Precision**** | ****Recall**** | ****F1 - Score**** |
| **Logistic Regression** | **1.0** | **1.0** | **1.0** | **1.0** |
| **Decision Tree** | **0.95** | **0.96** | **0.95** | **0.95** |
| **Random Forest** | **1.0** | **1.0** | **1.0** | **1.0** |

Os métodos não supervisionados, começando pela *hierarquical classification* e a combinação do PCA e K-means mostraram dificuldades por não haver clara distinção dos clusters de dados. Finalmente a rede neuronal com *multilayer backpropagation* mostrou precisão 100% em todos os parâmetros de métricas. Sendo indiferente se apenas usando dados de caudais, pressão ou ambos.

A investigação permitiu perceber a capacidade destes 6 métodos no problema das fugas de água, os métodos com melhores capacidades foram a *logistic regression, random forest* e a *Artificial Neural Net* (ANN).

Um outro autor usando o mesmo [4] software EPANET para treino da rede neuronal, obteve bons resultados com o método *Radial Basis Function*. O referido estudo usou *RBF* com três *layers*, *input* , *hidden layer* e *output* *layer* com 2 *neurons*, a determinação do numero de *neurons* na *hidden layer* e ajustada por tentativa erro, até obter o menor erro e o melhor tempo de treino. Os dados usados foram a pressão em cada junção/interceção de cada tubo.

Dados foram obtidos quando uma fuga está em acontecimento, estes dados não foram usados no treino, pois seriam poucos dados, mas sim dados simulados no software EPANET. Os dados usam tamanho e largura nas tubagens, rugosidade, elevação e caudal base. Foram testados 1328 modelas de tubos, com 98% de *accuracy*, e 1921 com 95%. É expectável que o referido método detete magnitude e localização. Este autor usou também um microcontrolador para obter dados em tempo real.

Outro artigo recente do autor Xiaoxuan Wu [5] descreve o uso de uma rede neuronal integrada com *deep learning*, que combina uma *Covulocional Neural Network* com LSTM para localizar pontos de fuga. Dados recolhidos incluem comprimento das tubagens, diâmetro, e também o caudal base requerido.

Devido custo de medição de pressão ou caudal em cada ligação de uma rede este artigo demonstra se será possível avaliar a rede em geral monitorizando a rede a partir de um numero finito de locais com medições. A ideia deste autor foi encontrar similaridades entre as ligações e classificá-las como sendo da mesma classe. Ao recolher dados de caudais/pressão de diferentes períodos, (alto, baixo, médio) usa uma técnica de *clustering* chamada *fuzzy C-means* que gerou várias versões de *clustering* que depois agrupou por uma matriz de pesos.

Para a técnica de deteção de fugas e visto que os dados não estacionários (pressão ou caudal) têm características de serie temporal LSTM é o indicado para este problema. A avaliação inicial foi retirada com dados estacionários. Técnicas de avaliação de series temporais foram testadas como Augmented Dickey Fuller(ADF), Empirical Mode Decomposition(EDM).

Segue-se os dados estacionários acima referidos que foram trabalhados com uma *CNN* que têm vantagens de avaliar *features* realçadas pela própria *CNN.*

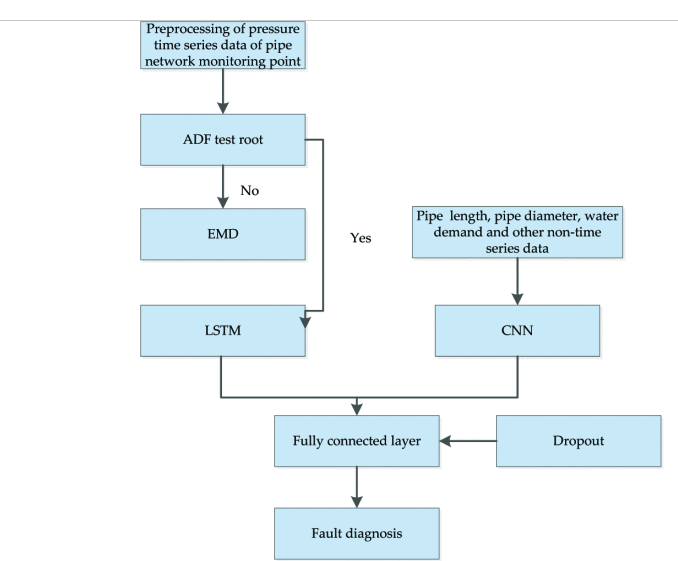


Figure Rede LSTM com CNN

O algoritmo e modelo foram aplicados num projeto real numa rede de água na região de Langxi County, Xuancheng City, Anhui Province. Considerando apenas um local de fuga o modelo apresentou uma *accuracy* de 90.25%. O autor refere também que o uso redes neuronais no problema das fugas usando métodos como o SVM ou *BPNN* que requerem mais processamento e mais dados a precisão do modelo irá reduzir e que a complexidade temporal vai aumentar, consolidando a escolha de CNN e LSTM pois teem características similares em que o aumento de dados aumenta a *accuracy* ajustando também a questão de *overfitting.*

Joseph Kembra [6] usando *Support Vector Machine*(SVM) e *Radial Basis Function(* RBF*)*, SVM é uma Técnica de IA para reconhecimento de padrões capazes de obter resultados superiores a outras técnicas, tem capacidade operar com sucesso com número elevado de inputs, são capazes de lidar com poucas amostras de dados para treino e teste.

SVMs podem ser usadas para regressão ou classificação, a *Support Vector Regression* usa várias *kernel* funções tais como RBF, *Linear* e *Polynomial Kernel*. O autor usou o RBF com o *Gamma* e o C, seguindo sequéncias exponenciais do *Gamma* e do C para encontrar o melhor par de parâmetros, avaliando através do *root mean squared error* (RMSE), altos valores significam baixa qualidade de predição, baixos valores significam boas predições.

Foram usados a partir do EPANET dados hidráulicos simulados que depois foram dividídos que simulam a perda de água em que 540 desses dados usados para treino e 60 como para teste.

Os melhores cinco pares de parametrização obtiveram exatidões acima dos 85% sendo a melhor com 90%.

## Panorama geral

Este capítulo resumiu brevemente as principais abordagens de métodos tecnológicos e globais no problema das fugas de água.

O problema principal indica falta de dados destes sistemas que provém da dificuldade tanto monetária como tecnológica do uso de sensores de pressão e leitores de caudais ao longo das redes de água sendo este um fator fundamental para o aumento da precisão destes algoritmos, outro fator passa pela variabilidade dos dados o que já se pode melhor com a escolha adequada do método com testagem requerida. As *features* principais do problema das fugas são os valores de pressão e o caudal.

Sendo as técnicas de IA as mais indicadas para previsão de localização de falhas ou deteção de ligações críticas, bem como o estado geral da rede. No ramo da IA artificial destacam se métodos supervisionados como *logistic regression, random forest* e métodos de *deep learning*, *RBF, SVMs, LSTM*.

Tabela Métodos Selecionados

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ****Método**** | ****Autor**** | ****Quantidade de dados**** | ****Implementação**** | ****Complexidade**** |
| **Random Forest** | **Mashhadi, Neda** | **Baixa** | **Fácil** | **Baixa complexidade temporal** |
| **Logistic Regression** | **Mashhadi, Neda** | **Baixa** | **Fácil** | **Baixa complexidade temporal** |
| **Radial Basis Function** | Joseph Kembra | **Alta** | **Complexa** | **Elevado tempo Processamento e de espaço** |
| **Support Vector Machine** | Joseph Kembra | **Alta** | **Complexa** | **Elevado tempo Processamento e de espaço** |
| **Long Short Time Memory** | Xiaoxuan Wu | **Alta** | **Complexa** | **Elevado tempo Processamento e de espaço** |

A tabela 3 apresenta alguns dos métodos mais promissores na resolução do problema tendo em conta a quantidade de dados para treino e predição, na generalidade a sua implementação tendo em conta os dados do problema, e a sua complexidade em termos de treino e processamento tanto como de memória. Devido a complexidade dos métodos de *deep learning* a abordagem inicial será com os métodos clássicos de *machine learning.*

# Análise Exploratória

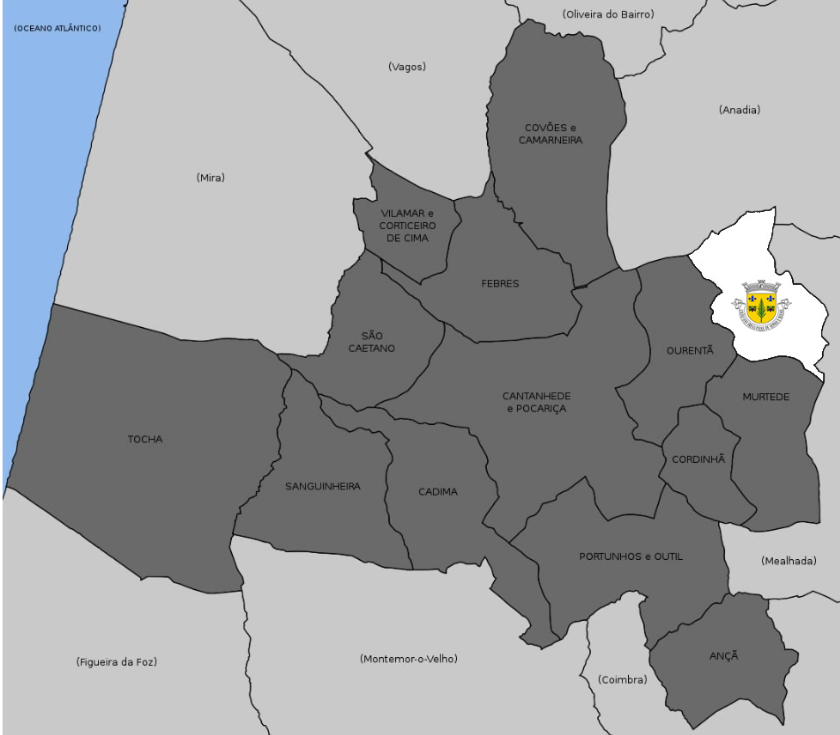
Neste capítulo vamos falar sobre a abordagem na exploração dos dados da rede da cidade de Cantanhede, onde foram inicialmente obtidos dados de caudal de uma zona de medição e controlo (ZMC), registos de intervenções desde 2018, tubagens, para a análise do problema das fugas de água e avaliação geral da rede de águas. O processo de análise foi abordado pela metodologia CRISP-DM sendo este capítulo dedicado a parte de *data understanding*.

## Ferramentas de análise de dados

Durante o processo de análise foi preciso procurar ferramentas que permitissem a fácil gestão e processamento de dados, uma das ferramentas mais conhecidas para análise de *datasets* é o *jupiter notebook* com python, tanto na capacidade de visualização (usando o *jupiter notebook*) como processamento e agilidade de programática usando librarias clássicas como o panda, pyplot.

## Caudal da ZMC do bolho

A freguesia do bolho no município de Cantanhede caracteriza-se por uma planície que faz fronteira com o distrito de Aveiro.



Os dados obtidos de caudal são medições em m3 por hora por razões de poupança de medições e de rendimento de baterias associadas aos sensores de medição. Os inícios das medições obtidas foram registadas em 2020 até agosto de 2022.

A análise mensal do caudal mostra uma variação de sinal mensal ao longo do ano claramente similar ao longo dos anos. Da para ver que durante os meses de Inverno e primavera até ao inicio do verão há um estabilização do valor do caudal, pouco crescente que, mas em que em maio há realmente um aumento significativo do caudal.

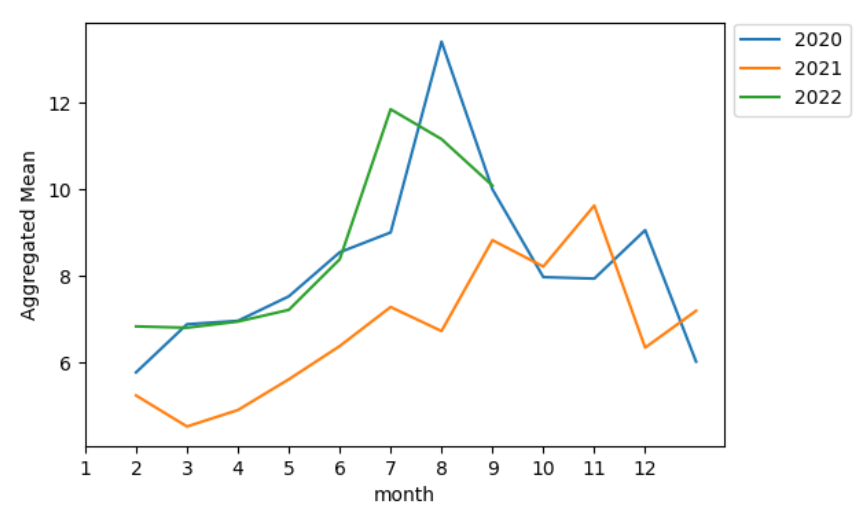


Figura Caudal Média Mensal

A figura 2 mostra a variação a longo do ano com aumentos significativos nos meses de verão e onde há variações mais acentuadas que será mais propenso a fugas nas ligações. O ano de 2021 mostra uma redução que pode ser explicada pela redução do consumo como podemos comprovar pela figura 3 que mostra que o ano 2021 foi bastante chuvoso logo o uso da água armazenada em poços e furos garante a pouca utilização da água da ZMC do bolho. Sendo assim possível concluir que o principal fator do caudal anualmente é a precipitação.

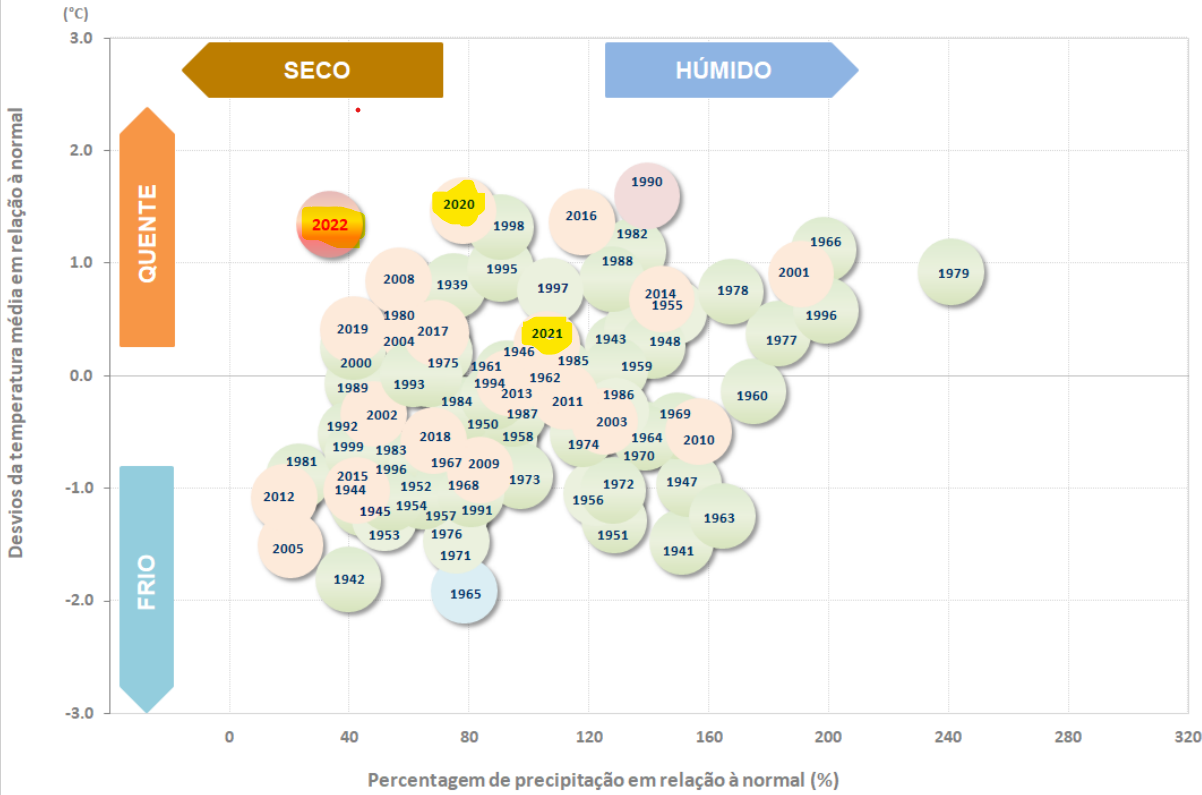


Figura Boletim Climatológico Inverno 2022 Portugal Continental

# Mais um capítulo

… as subsecções seguintes são apenas exemplificativas…

## Requisitos

…

## Arquitetura

…

## Tecnologias e ferramentas usadas

… que se encontram descritas na Tabela 4.1.

Tabela . – Tecnologias e ferramentas usadas

|  |  |  |
| --- | --- | --- |
| **Nome** | **Descrição** | **Utilização no projeto** |
| Picasso [1] | Biblioteca *open source* que permite fazer *download* e *caching* de imagens de forma simples. | Utilizada para fazer o *download* das imagens de perfil do utilizador. |
|  |  |  |

## Implementação

…

## Testes e validação

…

### Método A

…

### Método B

…

# Mais um capítulo

…

# Conclusões e trabalho futuro

...

## Conclusões

...

## Trabalho Futuro

…

# Referências

|  |  |
| --- | --- |
| [1] | “ISEC - Instituto Superior de Engenharia de Coimbra | Instituto Politécnico de Coimbra,” [Online]. Available: https://www.isec.pt/PT/Default.aspx. [Acedido em 8 4 2021]. |
| [2] | G. P, International Journal of Applied Engineering Research ISSN 0973-4562, 2017. [Online]. Available: https://www.ripublication.com/ijaer17/ijaerv12n16\_19.pdf. [Acedido em 12 11 2022]. |
| [3] | N. Mashhadi, “Use of Machine Learning for Leak Detection and Localization,” 28 Setembro 2021. [Online]. Available: https://www.mdpi.com/2624-6511/4/4/69/htm. [Acedido em 11 8 2022]. |
| [4] | A. U. Salam, “ON-LINE MONITORING SYSTEM OF WATER LEAKAGE DETECTION IN,” 10 october 2014. [Online]. Available: http://www.arpnjournals.com/jeas/research\_papers/rp\_2014/jeas\_1014\_1258.pdf. [Acedido em 13 11 2022]. |
| [5] | X. Wu, Leakage Location Method of Water Supply Pipe, Dalian, China, 2022. |
| [6] | J. Kemba, “Leakage detection in Tsumeb east water distribution network using EPANET and support vector regression,” 30 Maio 2017. [Online]. Available: https://ieeexplore.ieee.org/document/8102401. [Acedido em 11 8 2022]. |
| [7] | K. B. Adedeji, “Leakage Detection and Estimation Algorithm for Loss Reduction in Water Piping Networks,” 11 October 2017. [Online]. Available: https://www.mdpi.com/2073-4441/9/10/773/htm. [Acedido em 13 11 2022]. |
| [8] | “ISEC - Licenciaturas - Engenharia Informática,” [Online]. Available: https://www.isec.pt/PT/estudar/licenciaturas/EngInfor/. [Acedido em 8 4 2021]. |

# 

# Anexos

Proposta de Estágio

… conteúdo Anexo A

… conteúdo Anexo A

… conteúdo Anexo A

Especificação de Requisitos

… conteúdo Anexo B

… conteúdo Anexo B

… conteúdo Anexo B