



FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

MESTRADO INTEGRADO EM ENGENHARIA DO AMBIENTE 2019/2020

**MINIMIZAÇÃO DAS PERDAS APARENTES DOS SISTEMAS DE
ABASTECIMENTO DE ÁGUA COM RECURSO A ALGORITMOS DE
*MACHINE LEARNING***

ANA CAROLINA NUNES LOUREIRO

Dissertação submetida para obtenção do grau de

MESTRE EM ENGENHARIA DO AMBIENTE

Presidente do Júri: Cidália Maria de Sousa Botelho

(Professora Auxiliar do Departamento de Engenharia Química da Faculdade de
Engenharia da Universidade do Porto)

Orientador académico: Joaquim Manuel Veloso Poças Martins

(Professor Associado do Departamento de Engenharia Civil da Faculdade de
Engenharia da Universidade do Porto)

Orientador na empresa: Helena Maria Pinto Tomaz

(Diretora da Área Comercial e Atendimento da Empresa Águas de Gaia, EM, S.A)

setembro, 2020

Agradecimentos

Esta dissertação é uma exposição escrita de um projeto que exigiu muito trabalho e ajuda de várias pessoas e, desta forma, cabe-me agradecer a cada pessoa que me apoiou para que isto fosse possível.

Antes de mais quero agradecer à Faculdade de Engenharia da Universidade do Porto por me ter fornecido as ferramentas necessárias para conseguir realizar todo este projeto.

Ao professor Joaquim Poças Martins, pelo seu profissionalismo, conhecimentos transmitidos e por me ter ajudado a encontrar uma empresa onde fosse possível o desenvolvimento deste projeto.

Na empresa Águas de Gaia, EM, SA, são vários os nomes que tenho de agradecer e que sem eles certamente as coisas teriam sido muito mais complicadas. Desde já quero agradecer ao Dr. Miguel Lemos, administrador executivo da empresa Águas de Gaia, EM,S.A pela oportunidade dada em desenvolver a minha dissertação na empresa.

À Dra. Helena Tomaz toda a disponibilidade, acompanhamento e ajuda nesta experiência, que se tornou muito mais fácil graças a toda a sua dedicação.

Ao Sr. Amaral, ao Sr. Manuel, ao Sr. Fernando, à Lara e à Ju, por toda a integração, boa disposição, ensinamentos e apoio que me deram todos os dias, sem dúvida que tudo foi muito mais fácil porque vos tive sempre por perto.

Ao Eng. Carlos Ranginha toda a ajuda no funcionamento e compreensão da linguagem *Python* nos momentos em que a esperança já escasseava.

A todas as outras pessoas da empresa Águas de Gaia, EM, SA que não mencionei, mas que sempre se disponibilizaram para me ajudar em todo este projeto.

À minha família, em especial à minha mãe e aos meus irmãos, por todo o apoio não só nesta fase como em toda a minha vida, um obrigado é pouco por tudo o que SEMPRE fizeram por mim.

À Mó, às Lau's, à Inês, às minhas colegas de casa, Nat, Kika, Lili e Verónica, mas em especial à Lu, por todas as conversas, por toda a ajuda e dedicação, sem ela as coisas teriam sido sem dúvida mais complicadas. E a todos os que não referi, mas de algum modo também foram uma peça fundamental em todo este processo, sem o vosso apoio este projeto não teria sido a mesma coisa. O meu MUITO obrigada a cada um de vocês!

Resumo

Este trabalho teve como objetivo minimizar as perdas aparentes do sistema de abastecimento de águas do município de Vila Nova de Gaia através da deteção dos consumos não autorizados, recorrendo para isso a algoritmos de *Machine Learning*.

Este é um tópico bastante importante para as empresas desta área, uma vez que a ineficiência das mesmas está relacionada com a percentagem de perdas de água nos sistemas de abastecimento. As perdas de água dividem-se em perdas reais e perdas aparentes, sendo que estas últimas têm uma especial importância para as EG uma vez que afeta as receitas das mesmas.

No sentido de atuar na minimização deste problema, o presente projeto foi dividido inicialmente em três partes. Numa primeira parte começou-se por recolher os dados de todos os clientes domésticos do ano 2019, todos os clientes que rescindiram contrato nesse mesmo ano, bem como todos os clientes que apresentavam dívidas e que foram detetadas contraordenações. Em segundo lugar, compilaram-se todos os dados recolhidos num único ficheiro *Excel*, recorrendo-se a algumas ferramentas como *Tabelas Dinâmicas* e *PROCV*. Por último, organizaram-se os clientes em dois grupos: os clientes ativos e os casos ilícitos conhecidos, correspondendo a clientes a quem foram detetadas contraordenações.

Na segunda parte procedeu-se ao desenvolvimento de um código em linguagem *Python*, desenvolvido no programa *Anaconda Navigator*, em que se utilizou alguns algoritmos de *Machine Learning* (KNN, RFC, LRN, SVR) para prever casos ilícitos que atualmente se encontram a consumir água de forma ilegal.

A terceira parte deste projeto passou pela análise dos resultados obtidos pelo código desenvolvido, de forma a verificar-se presencialmente cada um dos casos encontrados e se correspondem, de facto, a casos de ilicitude.

Em suma, o algoritmo onde se obteve os melhores resultados foi o KNN, a partir do qual foi possível prever, no total 48 clientes, 28 casos ilícitos já conhecidos e 20 que seriam possíveis ilícitos desconhecidos para a empresa. Porém, dos 20 ilícitos desconhecidos, todos eles foram verificados no local e nenhum se encontrava a praticar ilicitudes. Isto não significa que o modelo criado não esteja apto para prever ilícitos - apesar de não se ter confirmado nenhum caso, o mesmo conseguiu detetar 28 casos ilícitos conhecidos. A principal razão prende-se com o facto de que os ilícitos conhecidos fornecidos ao modelo correspondiam apenas a ilícitos aos quais foi instruído um processo jurídico, sendo poucos comparativamente aos clientes ativos estudados,

pois os algoritmos utilizados neste modelo não conseguem prever valores com elevada eficiência nestas condições.

Palavras-chave: Perdas aparentes, Água não faturada, Consumos não autorizados, Ilícitos, *Machine Learning*.

Abstract

This work aimed to minimize the apparent losses of Vila Nova de Gaia's water supply system through the detection of unauthorized consumption resorting to Machine Learning algorithms.

This is an extremely important topic for water companies, since their inefficiency is related to the percentage of water losses in the water supply systems. Water losses are divided into real losses and apparent losses, the latter having a special importance for EG once it affects their revenues.

In order to minimize this problem, the present project was initially divided into three parts. The first part started by collecting data of all domestic customers in the year 2019, all customers who terminated the contract in the same year, as well as all customers who were in debt and were detected infractions. In the second step, it was compiled the collected data to create a single Excel file, using the aid of some features, such as Dynamic Tables and PROCV. Lastly, the clients were organized into two groups: known active and illicit client groups, corresponding to customers to whom breaches were detected.

The second part consisted in the development of a Python language code, developed in the Anaconda Navigator program, using Machine Learning algorithms (KNN, RFC, LRN, SVR) to predict illicit cases that are currently consuming water illegally.

The third part of this work passed by the analysis of results obtained by the developed code, in order to verify each of the cases found in person and if they are practicing any illegality.

In conclusion, the algorithm where the best results were obtained was the KNN, from which it was possible to predict a total of 48 clients, 28 known illicit cases and 20 unknown to the company. However, of the 20 unknown illicit cases, all of them were verified on site and none were found to be practicing illegality. This does not mean that the model created is not able to predict illicit events - although no case has been confirmed, it was able to detect 28 known illicit cases. The main reason is due to the fact that the known illicit cases supplied to the model corresponded only to those which a legal process was instructed, being few when compared to the studied active clients, since the algorithms used in this model cannot predict values with high efficiency in this conditions.

Keywords: Apparent losses, Unbilled water, Unauthorized consumption, Illicit, Machine Learning.

Índice

Agradecimentos	iii
Resumo	v
Abstract.....	viii
Índice	x
Lista de Figuras.....	xiii
Lista de Tabelas	xvi
Nomenclatura.....	xviii
1 Introdução.....	1
1.1 Apresentação do Projeto	1
1.2 Objetivos da Dissertação	2
1.3 Organização da Dissertação	2
2 Contexto e Estado da Arte	5
2.1 Caracterização do Sistema de Abastecimento de Água	5
2.1.1 Introdução.....	5
2.1.2 Etapas do Sistema de Abastecimento de Água	5
2.2 Perdas de Água.....	10
2.2.1 Introdução.....	10
2.2.2 Caracterização das Perdas de Água	12
2.3 Dimensões do problema	22
2.3.1 Dimensão Económico-Financeira.....	22
2.3.2 Dimensão Técnica	23
2.3.3 Dimensão Ambiental	23
2.3.4 Dimensão de Saúde Pública	23
2.3.5 Dimensão Social	24
2.4 Balanço Hídrico	24

2.5 <i>Machine Learning</i>	27
2.5.1 Rede Neuronal Humana e Rede Neuronal Artificial	30
2.5.2 Criação de um Modelo	32
2.5.3 Modelos de Classificação	35
2.5.4 <i>Anaconda Navigator</i>	37
3 A Empresa	40
3.1 Águas de Gaia EM,S.A.	40
3.1.1 Caracterização da Entidade Gestora	40
3.1.2 Balanço Hídrico da Empresa	46
4 Trabalho Desenvolvido	51
4.1 Ficheiro de <i>Excel</i>	51
4.1.1 Variáveis.....	51
4.2 <i>Machine Learning</i>	53
5 Análise dos Dados	58
5.1 Dados de <i>Excel</i>	58
5.2 Dados do <i>Machine Learning</i>	61
5.3 Análise Final.....	65
6 Conclusões	68
7 Avaliação do Trabalho Realizado	70
7.1 Objetivos Realizados	70
7.2 Recomendações Futuras	70
7.3 Apreciação Final	71
8 Referências	73
Apêndices	i
Apêndice A - Dados fornecidos pela empresa	iii
A.1 Clientes domésticos ativos	iii
A.2 Clientes que rescindiram contrato.....	vi
A.3 Clientes com dívida	viii

A.4 Casos Ilícitos por Contraordenação da Água.....	x
A.5 Casos Ilícitos por Contraordenação dos Contadores.....	xi
Apêndice B - Gráficos do <i>Machine Learning</i>	xiv

Lista de Figuras

Figura 1: Esquema das Etapas do Sistema de Abastecimento de Água.	6
Figura 2: Esquema da rede de distribuição entre o sistema público e as habitações. 10	
Figura 3: Evolução do indicador água segura.....	11
Figura 4: Esquema do tempo em função do volume de água perdido por uma fuga de difícil deteção.	13
Figura 5: Esquema do tempo em função do volume de água perdido por uma fuga de fácil deteção.....	13
Figura 6: Esquema do tempo em função do volume de água perdido através de uma perda de base.....	14
Figura 7: Correlador de ruído de fugas das Águas de Gaia, EM, S.A.	15
Figura 8: Geofone das Águas de Gaia, EM, S.A.....	16
Figura 9: Esquema de uma ligação "bypass".	18
Figura 10: Esquema de uma ilicitude do tipo derivação do ramal.	19
Figura 11: Exemplo de uma ligação direta fornecida pela empresa Águas de Gaia, EM, S.A.	19
Figura 12: Exemplo de uma ilicitude encontrada pela empresa Águas de Gaia, EM,S.A com marcos de incêndio.	20
Figura 13: Balanço Hídrico	26
Figura 14: Ramos da Inteligência Artificial.	27
Figura 15: Classificação do ML.	29
Figura 16: Constituição de um neurónio biológico.	31
Figura 17: Constituição de um neurónio artificial.	31
Figura 18: Modelo de classificação do KNN com k=5.	36
Figura 19: Modelo de classificação SVM.	37
Figura 20: Anaconda Navigator.	38
Figura 21: Mapa da rede de distribuição de água potável do município de Vila Nova de Gaia.	41
Figura 22: Sistema de Telegestão da empresa Águas de Gaia, EM, S.A.	42

Figura 23: Gráfico do nível de um reservatório ao longo de um dia.	43
Figura 24: Gráfico do nível de cloro ao longo de um dia.	43
Figura 25: Gráfico do consumo diário de um reservatório.	44
Figura 26: Imagem das zonas onde existe uma perda de água.	45
Figura 27: Imagem da localização exata da perda de água.	45
Figura 28: Evolução do número de clientes na empresa Águas de Gaia, EM, S.A.	47
Figura 29: Evolução do volume de água que entrada no sistema, volume de água faturada e volume de água não faturada na empresa Águas de Gaia, EM, S.A.....	48
Figura 30: Evolução dos valores relativos aos erros de medição e aos consumos não autorizados na empresa Águas de Gaia, EM, S.A.	49
Figura 31: Código utilizado para importar todas as bibliotecas fundamentais para o projeto.....	53
Figura 32: Treino do modelo com os algoritmos de Machine Learning.	54
Figura 33: Código utilizado para calcular as variáveis confusion matrix, accuracy score e classification report.	55
Figura 34: Resultados obtidos com o código da confusion matrix e da accuracy score.	55
Figura 35: Código utilizado para com o auxílio dos algoritmos identificar os clientes associados.	55
Figura 36: Exportação dos resultados.	56
Figura 37: Representação gráfica dos casos ativos e casos ilícitos referentes à a) quantidade de água consumida b) valor em dívida e c) número de faturas em atraso.	59
Figura 38: Relação entre o valor em dívida e quantidade de água consumida para os clientes ativos e para os casos ilícitos.	60
Figura 39: Relação entre o número de rescisões e o a) valor em dividia b) quantidade de água consumida e c) número de faturas em atraso.	60
Figura 40: Representação gráfica dos resultados obtidos pelos 4 algoritmos relativos ao a) número de rescisões b) número de faturas em atraso e c) quantidade de água consumida.	64

Figura B. 1: Representação gráfica do número de cliente em função a) número de cliente, b) número de rescisões, c) valor em dívida, d) número de faturas em atraso, e) quantidade de água consumida. xiv

Figura B. 2: Representação gráfica do número de rescisões em função a) número de cliente, b) número de rescisões, c) valor em dívida, d) número de faturas em atraso, e) quantidade de água consumida. xiv

Figura B. 3: Representação gráfica do valor em dívida em função a) número de cliente, b) número de rescisões, c) valor em dívida, d) número de faturas em atraso, e) quantidade de água consumida. xv

Figura B. 4: Representação gráfica do número de faturas em atraso em função a) número de cliente, b) número de rescisões, c) valor em dívida, d) número de faturas em atraso, e) quantidade de água consumida. xv

Figura B. 5: Representação gráfica da quantidade de água consumida em função a) número de cliente, b) número de rescisões, c) valor em dívida, d) número de faturas em atraso, e) quantidade de água consumida. xv

Lista de Tabelas

Tabela 1: Valores dos parâmetros mais relevantes do balanço hídrico da empresa Águas de Gaia, EM, S.A.	47
Tabela 2: Tabela final com todas as variáveis (apenas para alguns casos).	51
Tabela 3: Resultados obtidos pelo algoritmo KNN.	61
Tabela 4: Resultados obtidos pelo algoritmo RFC.	62
Tabela 5: Resultados obtidos pelo algoritmo LRN.	63
Tabela 6: Resultados obtidos pelo algoritmo SVM.	63
Tabela 7: Resultados obtidos pelos 4 algoritmos.	65
Tabela A.1. 1: Dados referentes aos clientes domésticos ativos de Janeiro a Maio. ...	iii
Tabela A.1. 2: Valores referentes aos clientes domésticos ativos de Junho a Setembro.	iv
Tabela A.2. 1: Clientes domésticos que rescindiriam contrato	vi
Tabela A.3. 1: Valores referentes aos clientes com dívida.	viii
Tabela A.4. 1: Casos ilícitos por contraordenação ao nível da água.	x
Tabela A.5. 1: Casos ilícitos por contraordenação dos contadores.	xi

Nomenclatura

Adp	Águas de Portugal
CAP	Controlo Ativo de Perdas
DECO	Associação Portuguesa para a Defesa do Consumidor
EG	Entidade Gestora
ERSAR	Entidade Reguladora dos Serviços de Águas e Resíduos
ETA	Estação de Tratamento de Água
IA	Inteligência Artificial
IWA	<i>International Water Association</i>
KNN	<i>K Nearest Neighbors</i>
LRN	<i>Logistic Regression</i>
ML	<i>Machine Learning</i>
ONU	Organização das Nações Unidas
RFC	<i>Random Forest Classifier</i>
SAA	Sistemas de Abastecimentos de Água
SIG	Sistema de Informação Geográfica
SVM	<i>Support Vector Machine</i>
VRP	Válvulas Redutoras de Pressão
ZMC	Zona de Monitorização de Controlo

1 Introdução

1.1 Apresentação do Projeto

A Organização das Nações Unidas (ONU) alertou, no 8º Fórum Mundial da água em 2018, para a falta de água que virá a afetar cerca de 9,7 bilhões de pessoas até ao ano de 2050. (Época Negócios, 2018) Desta forma, a minimização das perdas de água é uma das metodologias a ser aplicada para se mitigar este problema que está a preocupar todo o mundo. À escala nacional, a Associação Portuguesa para a Defesa do Consumidor (DECO) prevê que as perdas de água poderão chegar anualmente a mais de 180 milhões de m³ de água em 258 municípios, sendo que dos restantes 20 municípios em falta não existe informação sobre o volume perdido. (Observador, 2019)

O crescimento populacional e o aumento da intensificação das atividades económicas têm levado a uma utilização excessiva dos recursos hídricos e, conseqüentemente, à degradação das origens da água, resultando num aumento dos custos de serviços do abastecimento de água. (Teixeira et al., 2014)

A percentagem de água não faturada é um parâmetro bastante relevante para as entidades gestoras envolvidas neste ramo, pois quanto maior for a percentagem de água não faturada menos eficiente será a empresa. Os valores de água não faturada dependem essencialmente das perdas de água totais e do consumo autorizado não faturado. (Cardoso et al., 2013)

O cálculo do balanço hídrico é um sistema muito utilizado pelas entidades gestoras para avaliarem estes parâmetros de uma forma mais simplificada e discriminada.

As perdas de água subdividem-se em perdas reais e perdas aparentes. As primeiras são caracterizadas por todas as roturas e fugas que existem no sistema de abastecimento de água. Já as perdas aparentes referem-se aos erros de medição e a casos de ilicitude.

São várias as medidas que têm vindo a ser implementadas ao longo dos anos no que se refere à diminuição das perdas reais. Em contrapartida, as medidas aplicadas para a minimização das perdas aparentes não têm evoluído muito ao longo dos anos, por esta razão a presente dissertação será realizada no âmbito da minimização das perdas aparentes dos Sistemas de Abastecimentos de Água recorrendo a metodologias de deteção com algoritmos de *Machine Learning*.

A empresa Águas de Gaia, EM, S.A será a entidade gestora na qual será aplicado este método através do qual se desenvolvem todos os objetivos propostos.

É importante continuar a manter a evolução dos últimos anos no setor do abastecimento de água e dos sistemas de recolha de saneamento, para isso, deve-se manter uma gestão eficiente dos recursos com a mesma qualidade e quantidade. Este é um dos principais objetivos do Plano Estratégico de Abastecimento de Água e Saneamento de Águas Residuais para se cumprir até ao ano 2020 (PENSAAR 2020). (Frade et al., 2015)

1.2 Objetivos da Dissertação

O presente trabalho tem como principal objetivo a minimização das perdas aparentes dos sistemas de abastecimento de água com recurso a algoritmos de ML. Por esse motivo, a utilização da linguagem *Python* como forma de tratamento dos dados fornecidos é uma opção bastante satisfatória.

Com a evolução da IA é possível desenvolver-se modelos que sejam capazes de detetar casos de ilícitos desconhecidos para as EG. Assim, o avanço da IA será bastante importante para a diminuir deste tipo de perdas de água.

1.3 Organização da Dissertação

Esta dissertação encontra-se dividida em sete capítulos: Introdução; Contexto e Estado da Arte; A Empresa; Trabalho Desenvolvido; Análise dos Dados; Conclusões e Avaliação do trabalho realizado.

Capítulo 1: Na Introdução é possível encontrar uma pequena contextualização da situação e apresentação do trabalho desenvolvido.

Capítulo 2: O Contexto e Estado da Arte engloba toda a informação de base teórica essencial para a compreensão e desenvolvimento deste projeto. Na parte final, é possível encontrar uma previsão dos próximos anos relativamente ao número total de clientes, água que entra no sistema e consumos de água faturada e não faturada para a empresa Águas de Gaia EM, SA.

Capítulo 3: A Empresa, aqui é possível encontrar uma breve descrição da empresa Águas de Gaia EM,SA, bem como toda a metodologia aplicada relativamente às perdas reais e às perdas aparentes.

Capítulo 4: Trabalho Desenvolvido, neste capítulo é analisado e descrito todo o processo que foi realizado até se obterem os resultados finais e, ainda, uma breve contextualização teórica do *Machine Learning*.

Capítulo 5: Análise dos Dados, aqui serão apresentados todos os resultados obtidos, bem como a respetiva discussão dos mesmos.

Capítulo 6: Conclusões, aqui serão descritas todas as conclusões retiradas deste projeto.

Capítulo 7: Avaliação do trabalho realizado, neste capítulo irão ser analisados os objetivos deste trabalho, bem como as limitações encontradas. Para além disso, apresenta recomendações para futuros trabalhos nesta área e termina com uma breve apreciação pessoal do trabalho desenvolvido nesta dissertação.

2 Contexto e Estado da Arte

2.1 Caracterização do Sistema de Abastecimento de Água

2.1.1 Introdução

O Sistema de Abastecimento de Água (SAA) e o Sistema de Saneamento de Águas Residuais são dois sistemas extremamente importantes para a saúde e bem-estar da população. Em Portugal, nos anos 90, cerca de 50% da água que era utilizada para consumo humano apresentava as condições necessárias para o seu consumo, enquanto que apenas $\frac{1}{4}$ da população se encontrava a libertar as águas residuais para os sistemas de tratamento de águas residuais. A restante população encontrava-se a libertar toda a água usada diretamente para o meio ambiente sem qualquer tipo de tratamento. (AdP - Águas de Portugal, 2016)

A evolução destes sistemas é fundamental para se conseguir garantir às gerações futuras a qualidade e a disponibilidade necessária destes serviços. Desta forma, um dos objetivos futuros destes sistemas passa por conseguirem assegurar, mesmo com um aumento da população, a sustentabilidade financeira, social, técnica e ambiental das Entidades Gestoras (EG). (AdP - Águas de Portugal, 2016)

2.1.2 Etapas do Sistema de Abastecimento de Água

Desde a captação em linhas de água até à fase em que é consumida, parte da água é perdida ao longo do percurso e não chega ao destinatário. Assim sendo, esta quota não será faturada pela empresa nem utilizada para outros usos autorizados. (Cardoso et al., 2013)

Todo esse percurso, designado por SAA, encontra-se representado na Figura 1 sendo o responsável pela distribuição de água a todos os clientes.

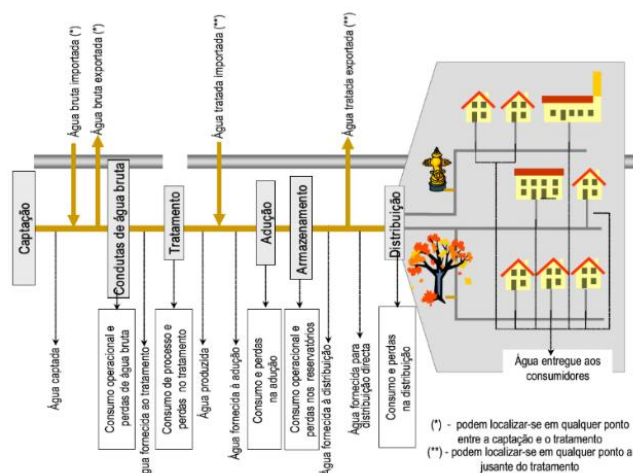


Figura 1: Esquema das Etapas do Sistema de Abastecimento de Água. (Alegre et al., 2005)

2.1.2.1 Captação da Água

A primeira etapa do sistema de abastecimento de água é a sua captação, que consiste na recolha de água com parâmetros adequados para que no final do tratamento esta esteja apta para ser consumida. A origem dos pontos de captação podem ser: água subterrânea, água superficial e água pluvial. A escolha é feita com base no levantamento de todas as bacias hidrográficas das zonas em estudo, pois há locais com maior capacidade para as águas subterrâneas e outros para as águas superficiais. Para além disso, a qualidade da água é também um fator preponderante, uma vez que existem locais que são direcionados para a agricultura e, nestes casos, os pesticidas podem muitas vezes atingir os lençóis de água que existem no subsolo. Existem outros locais em que o setor industrial é mais intenso e em que as descargas nas linhas de águas mais próximas sem qualquer tratamento são mais frequentes. Assim, tanto a disponibilidade como a qualidade da água são fatores decisivos na escolha da melhor origem de água para um determinado local. (VANGENTE, Argentino; GASPAS, 2018)

Depois de a água ser captada é direccionada para um reservatório de água bruta, que deverá estar a uma cota suficiente para que todo o processo de tratamento seja feito por gravidade.

Águas Subterrâneas

Este tipo de origem de água localiza-se abaixo da superfície do solo, encontrando-se em contacto com o solo ou com o subsolo. É nesta zona que se podem encontrar os lençóis de água, sendo que se deve avaliar previamente a sua qualidade a fim de

estudar a viabilidade da sua exploração. A quantidade de água existente nos lençóis de água está relacionada com a precipitação que irá posteriormente infiltrar-se nos solos, aumentando a quantidade de água nos aquíferos e, conseqüentemente, as águas subterrâneas. (República, 2005; VANGENTE, Argentino; GASPAR, 2018)

Águas Superficiais

As águas superficiais aparecem na natureza sob a forma de nascentes, rios, lagos ou lagoas. Destes tipos de águas superficiais as que apresentam melhor qualidade para abastecimento público de água são as nascentes. Contudo, com a elevada procura de água, estas fontes começaram a ser insuficientes, bem como todos os outros tipos de água superficial. Atualmente, devido a toda esta procura tornou-se necessário utilizar água do mar devidamente tratada. Os processos de tratamento da água do mar para consumo humano são bastante dispendiosos, acabando por interferir negativamente na distribuição deste bem tão valioso. (República, 2005; VANGENTE, Argentino; GASPAR, 2018)

Água Pluvial

Segundo o decreto regulamentar nº 23/95 “as águas pluviais ou também designadas como águas residuais pluviais consistem na precipitação atmosférica apresentando normalmente menores quantidades de matéria poluente. São também designadas de águas pluviais toda a água proveniente das regas dos jardins e espaços verdes, da lavagem de passeios, pátios e parques de estacionamento, normalmente recolhidos por sarjetas. A descarga final das águas pluviais deve ser feita para a linha de água mais próxima, tendo a necessidade de se garantir que a qualidade dessas águas pluviais seja compatível com a das águas das linhas de água”. (Decreto-Lei nº 23/95, 1995)

Em Portugal Continental, segundo os dados mais recentes validados pela Entidade Reguladora dos Serviços de Água e Resíduos (ERSAR), no ano de 2015 existiam cerca de 6016 captações de água subterrânea e 267 captações de água superficial. (ERSAR - Entidade Reguladora dos Serviços de Águas e Resíduos, 2015)

Para o caso das Águas de Gaia a empresa responsável por este processo é as Águas do Douro e Paiva, mais concretamente a ETA do Lever. Esta é constituída por três pontos de captação em profundidade e oito bombas de captação superficial na albufeira de Crestuma-Lever, sendo aqui onde se dá o início de todo o procedimento. A água é captada e posteriormente segue caminho para a ETA, onde sofre vários processos de tratamento a fim de que seja distribuída com a qualidade exigida por lei e atendendo a todos os requisitos necessários para o consumo humano. (Águas do Douro e Paiva, 2017)

2.1.2.2 Elevação

A água depois de tratada encontra-se apta para ser distribuída. Quando não é possível utilizar o escoamento gravítico é necessário introduzir energia ao sistema a fim de que a água possa chegar aos consumidores finais. Em Portugal Continental, existem cerca de 2313 estações elevatórias, sendo que na ETA do Lever há 3 estações elevatórias a partir das quais a água tratada será elevada para o Reservatório de Jovim, o Reservatório de Lagoa e o Reservatório de Seixo Alvo e, a partir deles, para a próxima etapa. (Águas de Gaia, EM, 2020a; ERSAR - Entidade Reguladora dos Serviços de Águas e Resíduos, 2015)

2.1.2.3 Transporte

A água após ser elevada fica armazenada em reservatórios para posteriormente ser transportada. Este transporte é realizado por um sistema de condutas adutoras que tem como principal objetivo o transporte da água desde o local de tratamento até aos reservatórios de armazenamento, onde seguidamente serão distribuídos até ao consumidor final. (ERSAR - Entidade Reguladora dos Serviços de Águas e Resíduos, 2015) Estas condutas adutoras apresentam elevadas dimensões pois têm como principal papel transportar uma maior quantidade de água do que as condutas de distribuição.

2.1.2.4 Armazenamento

A água depois de ser transportada por condutas adutoras é encaminhada para os reservatórios que existem ao longo de toda a rede de abastecimento de água. Este processo é de elevada importância, pois irá permitir controlar as irregularidades de consumo dos consumidores finais; permite também constituir uma reserva de emergência para eventuais incêndios, ou para o caso de haver uma interrupção mais prolongada do abastecimento do sistema de condutas adutoras; é também importante para se regular as pressões na rede de distribuição, bem como o funcionamento das bombagens. (ERSAR - Entidade Reguladora dos Serviços de Águas e Resíduos, 2015) A entidade gestora é que tem como função armazenar a água, desta forma existe uma preocupação por parte da mesma de maneira a manter a qualidade pretendida, para

isso, é necessário a injeção do cloro para que a água chegue ao consumidor final com a qualidade desejada.

Em Portugal Continental existem aproximadamente 8732 reservatórios com capacidade total de 5.070.452 m³, o que significa que, caso o abastecimento de água falhe, existe uma capacidade de reserva de 1,35 dias. Em relação ao município de Vila Nova de Gaia, existem cerca de 32 reservatórios que armazenam a água necessária para abastecer toda a rede de distribuição. (Águas de Gaia, EM, 2020a; ERSAR - Entidade Reguladora dos Serviços de Águas e Resíduos, 2015)

2.1.2.5 Distribuição

O sistema de distribuição de água tem por objetivo o fornecimento de água na quantidade e qualidade suficientes a todas as habitações da rede de distribuição. Este sistema irá funcionar com base na ligação entre o sistema público de distribuição de água e a rede de utilização, (Figura 2) sendo que este último é constituída por: (Decreto-Lei nº 23/95, 1995)

- Ramal de Ligação: Canalização entre a rede pública e o limite da propriedade a servir;
- Ramal de introdução coletivo: Canalização entre o limite da propriedade e os ramais de introdução individuais dos utentes;
- Ramal de introdução individual: Canalização entre o ramal de introdução coletivo e os contadores individuais dos utentes ou entre o limite da propriedade e o contador, no caso de edifício unifamiliar;
- Ramal de distribuição: canalização entre os contadores individuais e os ramais de alimentação;
- Ramal de alimentação: canalização para alimentar os dispositivos de utilização;
- Coluna: troço de canalização de prumada de um ramal de introdução ou de um ramal de distribuição.

Em Portugal Continental a rede de distribuição tem cerca de 108.757 km de condutas de água. Em relação ao concelho de Vila Nova de Gaia, a rede de distribuição de água tem uma extensão de 1494,2 km, o que corresponde a uma taxa de cobertura efetiva de aproximadamente 87%. Daqui a água segue para a etapa final de todo o processo de abastecimento de água: a utilização pelos clientes. Neste momento a água chega ao consumidor final com qualidade e em quantidade necessária, existindo ainda a

garantia de que o serviço de abastecimento de água funciona ininterruptamente para todos os clientes desta EG. (Águas de Gaia, EM, 2020a; ERSAR - Entidade Reguladora dos Serviços de Águas e Resíduos, 2015)

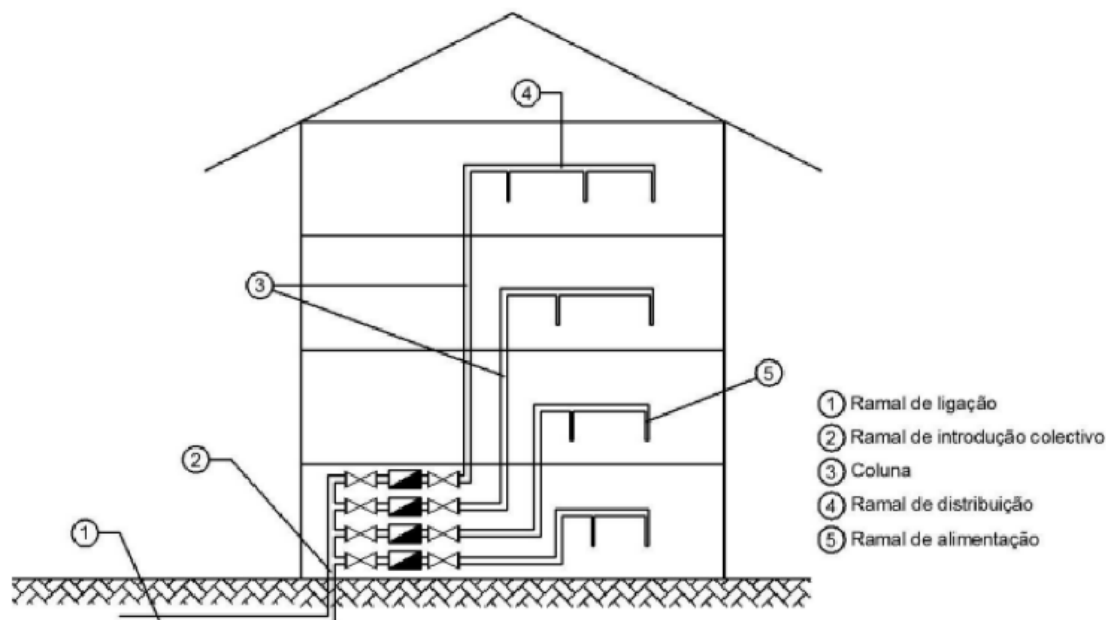


Figura 2: Esquema da rede de distribuição entre o sistema público e as habitações.
(Pedroso, 2007)

2.2 Perdas de Água

2.2.1 Introdução

As perdas de água são uma parte integrante de todas as redes de abastecimento de água e constituem uma das principais fontes de ineficiência das entidades gestoras dos serviços de abastecimento de água. Os pontos mais importantes para o aumento da eficiência do setor prendem-se com uma boa prática de gestão combinada com uma correta aplicação das técnicas mais eficazes. (Cardoso et al., 2013)

A ineficiência associada às perdas de água leva a que na empresa exista uma destabilização financeira e uma deficiente qualidade do serviço prestado aos consumidores. As perdas de água correspondem à água que não é faturada nem utilizada para outros usos autorizados, mas que é captada, tratada e transportada e, por isso, tem custos de operação e manutenção relevantes. A água não faturada

engloba as perdas de água totais e o consumo autorizado não faturado. (Alegre et al., 2005)

Em Portugal Continental entra no sistema anualmente aproximadamente 811 milhões de m³, e os volumes de água faturado e não faturado são de aproximadamente 563 milhões de m³/ano e 239 milhões de m³/ano, respetivamente. É de notar que isto representa uma percentagem de água não faturada em Portugal de 29,4%. (Entidade Reguladora dos Serviços de Águas e Resíduos, 2019b) A diminuição das perdas de águas é considerada uma alternativa económica à exploração de novas barragens, poços profundos ou dessalinização. (Teixeira et al., 2014)

Para o caso em particular da entidade gestora que vai ser estudada neste trabalho, Águas de Gaia, EM, S.A, alguns dos parâmetros importantes são: (Entidade Reguladora dos Serviços de Águas e Resíduos, 2019b)

- Água Segura: 100%
- Adesão ao Serviço de água: 96,1%
- Água não faturada: 27,3%
- Adesão ao serviço de saneamento: 95,6%

Em Portugal Continental o abastecimento de água segura tem sofrido algumas alterações ao longo dos anos, sendo que essa evolução pode ser observada na Figura 3. Atualmente, a percentagem de água segura é, em média, 98,63% sendo que, no ano de 1993, este valor era de 50%. Já para o caso do município de Vila Nova de Gaia desde o ano de 2014 a percentagem de água segura é de 100%. Assim, é possível concluir-se que, ao longo dos anos a qualidade da água dos seus clientes tornou-se mais controlada e de melhor qualidade. (Entidade Reguladora dos Serviços de Águas e Resíduos, 2019a)

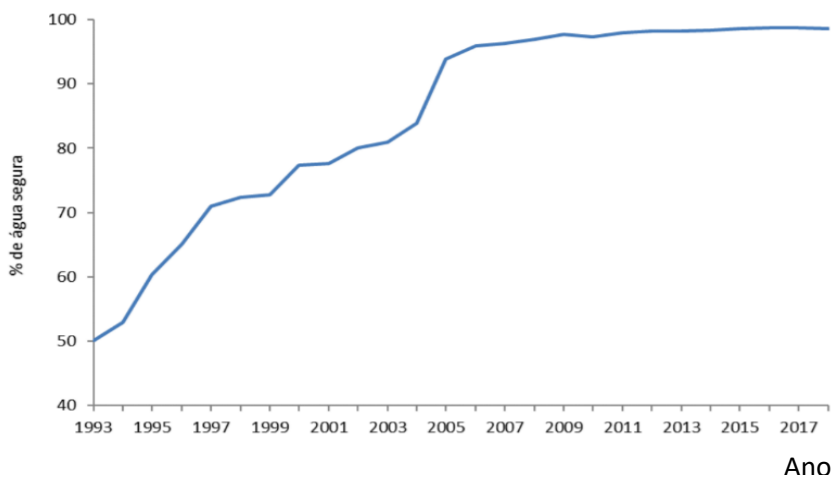


Figura 3: Evolução do indicador água segura.

No que diz respeito às águas residuais, segundo o relatório de contas intercalar do 1º semestre do ano de 2019 das Águas de Portugal, em Portugal Continental o volume de água tratado foi de cerca de 249 milhões de m³ e o volume faturado de 244 milhões de m³. (AdP - Águas de Portugal, 2019)

2.2.2 Caracterização das Perdas de Água

Como referido anteriormente uma das principais fontes de ineficiência das entidades gestoras de abastecimento de água são as perdas de água. Estas dividem-se em perdas reais e perdas aparentes.

2.2.2.1 Perdas Reais

As perdas reais relacionam o volume de água correspondente às perdas físicas desde o início do tratamento até ao contador do cliente. O tipo de fissuras, roturas e extravasamento depende do caudal, e da duração média de cada fuga, o que irá influenciar o volume anual de água perdida ao longo de toda a rede de distribuição. (Alegre et al., 2005) Este volume de água perdida está diretamente relacionado com as características e estado de degradação da rede de distribuição, bem como da rapidez e operacionalidade das reparações, e da utilização de tecnologia e conhecimento especializado para as devidas reparações na rede de distribuição. (Cardoso et al., 2013)

Alguns dos fatores que influenciam as perdas reais são: o estado das condutas; o seu material; a pressão de serviço média; a densidade e comprimento médio dos ramais; o comprimento total das condutas; a tipologia do solo e as condições do terreno. (Alegre et al., 2005)

As perdas reais encontram-se relacionadas com as perdas existentes nas condutas de distribuição e nos ramais domiciliários e com as perdas existentes devido a fugas, roturas e extravasamentos em reservatórios. Relativamente às perdas por fugas e roturas nas condutas de distribuição e nos ramais domiciliários estas podem ser divididas em três tipos de roturas: as roturas de difícil deteção, as roturas de fácil deteção e as perdas de base. (Cardoso et al., 2013)

As roturas de difícil deteção são caracterizadas por fugas que não surgem à superfície, encontram-se normalmente no subsolo e apenas são detetadas pelos equipamentos de

deteção. Este tipo de roturas é normalmente definido por caudais baixos levando a que o tempo de duração da fuga seja maior e haja um volume de água perdida mais elevado. (Figura 4) Contudo, este último parâmetro vai estar em parte dependente das estratégias implementadas pelo Controlo Ativo de Perdas (CAP). (Cardoso et al., 2013)



Figura 4: Esquema do tempo em função do volume de água perdido por uma fuga de difícil deteção. (adaptada de Farley et al., 2008)

As roturas de fácil deteção apresentam dimensões elevadas e, por isso, caudais mais elevados que o caso anteriormente mencionado. Desta forma este tipo de fugas aflora à superfície permitindo serem facilmente visualizadas pela população que de seguida informam a entidade responsável. Estas fugas definem-se por terem uma duração mais curta resultando num volume perdido mais baixo (Figura 5). (Oliveira, 2013)



Figura 5: Esquema do tempo em função do volume de água perdido por uma fuga de fácil deteção. (adaptada de Farley et al., 2008)

As perdas de base estão relacionadas com a existência de microrroturas e de fugas nas zonas de ligação da conduta de distribuição aos ramais domiciliários. Devido à reduzida dimensão deste tipo de rotura e fuga são dificilmente detetadas pelos equipamentos de deteção, por esse motivo são também designadas como perdas inevitáveis da rede de distribuição. Este tipo de rotura apenas será possível de detetar com o auxílio de equipamentos mais específicos. Contudo, a sua utilização não é eficiente, pois, a quantidade de água recuperada não compensaria os custos de utilização deste tipo de equipamento. São caracterizadas por caudais muito baixos levando a que o tempo de duração da fuga seja maior e o volume de água perdida muito elevado, é possível observar essa relação na Figura 6. (Oliveira, 2013)

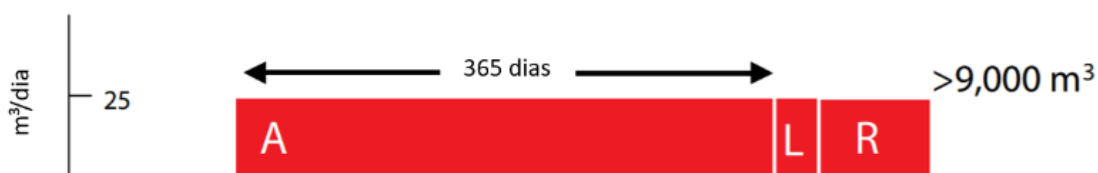


Figura 6: Esquema do tempo em função do volume de água perdido através de uma perda de base.
(adaptada de Farley et al., 2008)

A localização de uma perda de água por fuga ou rotura deve levar à realização de dois processos em conjunto: primeiramente a identificação do local, recorrendo ao subzonamento, e em seguida descobrir qual a localização exata da perda de água, através de equipamentos de deteção de fugas.

O subzonamento é um método que auxilia na deteção exata de fugas, e baseia-se na divisão da Zona de Monitorização de Controlo (ZMC) em subzonas. Dentro delas existem válvulas estrategicamente colocadas para que depois de fechadas se distingam as subzonas da ZMC separadamente. São também implementados estrategicamente medidores de caudal nas fronteiras de forma a quantificar a água que entra em cada subzona de forma a monitorizar onde acontecem as perdas. (Alegre et al., 2005; Carvalho, 2014)

Contudo, é imperativo identificar-se a localização exata das perdas de água, sejam elas visíveis ou não. Para tal recorre-se a equipamentos acústicos de forma a identificar o ponto que apresenta uma maior intensidade do som da água, pois esse local deverá eventualmente corresponder à zona onde se encontra a fuga. (Alegre et al., 2005; Carvalho, 2014) Na empresa Águas de Gaia, EM, S.A, os equipamentos de deteção mais

usados para as fugas de água são: satélite, correlador de ruído de fugas, *loggers acústicos*, vareta eletrónica de escuta e geofone.

O satélite foi a mais recente aquisição desta entidade gestora em parceria com uma empresa Israelita. Este equipamento tem por base a redução da quantidade de água perdida por fugas e roturas através da deteção do cloro existente no solo; este processo é possível através da sobreposição dos dados fornecidos por este equipamento em conjunto com o Sistema de Informação Geográfica (SIG) sendo assim possível identificar os potenciais locais detentores de fugas e roturas de água.

O correlador de ruído de fugas consiste na implementação à superfície de dois sensores em locais diferentes que através da medição do ruído e pelo conhecimento do tempo que demorará até chegar ao outro sensor, é possível determinar o local exato da fuga. É importante ter por base o tipo de material da conduta, o seu diâmetro e a distância a que se encontram os dois sensores. (Figura 7) Este tipo de método é o mais utilizado quando existe bastante ruído no local que se pretende analisar.



Figura 7: Correlador de ruído de fugas das Águas de Gaia, EM, S.A.

Quanto aos *loggers acústicos*, estes são colocados em diferentes hidrantes de forma a medir o nível de ruído e assim ser possível detetar a fuga. O local exato só é possível de se identificar com o auxílio de outros equipamentos.

A vareta eletrónica de escuta funciona apenas para casos em que a fuga é visualmente e audivelmente detetada, onde com a ajuda da vareta e dos auscultadores será possível de se ouvir o som na conduta que para o ouvido humano seria impercetível.

O geofone é constituído por um sensor, por um amplificador e pelos auscultadores. Este equipamento baseia-se na colocação dos auscultadores e com o auxílio de um sensor à superfície do solo é possível de se localizar o local onde o ruído se apresenta com maior intensidade, isso significa que esse será o local exato da fuga. (Figura 8)



Figura 8: Geofone das Águas de Gaia, EM, S.A.

O controlo das perdas reais é possível através da aplicação de algumas metodologias, desde a redução da pressão na rede através da simulação das válvulas reductoras de pressão (VRP); da eficácia e rapidez das reparações de fugas e roturas e, da reabilitação e substituição das infraestruturas mais antigas. (Cardoso et al., 2013)

2.2.2.2 Perdas Aparentes

As perdas aparentes, também denominadas de perdas comerciais, contabilizam todos os tipos de imprecisões associadas às medições da água consumida, e ainda ao consumo não autorizado por furto ou uso ilícito. (Alegre et al., 2005)

Conforme supramencionado, este tipo de perdas depende essencialmente de consumos não autorizados e de erros de medição, porém, ainda se pode subdividir em: erros humanos; erros informáticos e erros de procedimentos. Contudo, neste trabalho apenas será focado as principais causas envolvidas neste tipo de perdas, que são os consumos não autorizados e os erros de medição. (Alegre et al., 2005)

Os consumos não autorizados englobam o estabelecimento e uso de ligações ilícitas e utilização fraudulenta de marcos e bocas-de-incêndio, localizadas quer em locais públicos quer em locais privados. (Teixeira et al., 2014)

Não existem ainda muitas medidas de redução das perdas aparentes, o que acaba por ter um impacto económico negativo para muitas empresas. Contudo, a diminuição das perdas aparentes é uma vantagem financeira para a empresa, levando a um aumento da receita da entidade gestora. Já a redução das perdas reais implica uma diminuição da quantidade de água tratada comprada. (Teixeira et al., 2014)

Erros de medição

São um tipo de perdas aparentes que corresponde a erros de medição dos contadores que se apresentam incorretamente instalados; erros de leitura ou registo; erros de medição por avaria (violação do equipamento) e leituras em falta por dificuldades de acesso aos contadores por se encontrarem dentro das habitações. (Alegre et al., 2005)

Uma forma de se minimizar os erros por dificuldade de acesso aos contadores é através da implementação de um sistema de autoleituras em conjunto com os clientes em determinados horários a combinar, havendo apenas alteração deste compromisso em caso de realização de obras na habitação e apenas se estas influenciarem os sistemas prediais de água. (Alegre et al., 2005)

Consumo não autorizado

Este tipo de perda refere-se a todos os clientes que consomem água sem pagar a partir do estabelecimento de ligações ilícitas, religações não autorizadas, uso de ligações do tipo “bypass”, derivação de ramal, ligação clandestina, adulteração dos contadores e a utilização fraudulenta de marcos e bocas-de-incêndio, localizadas quer em locais públicos quer privados. No entanto, uma análise do historial dos clientes e da evolução do comportamento dos seus consumos ajuda na diminuição dos consumos não autorizados.

As ligações “bypass” são um tipo de ligação fraudulenta que consiste na instalação de uma conduta adicional que contorna o contador, onde apenas um pequeno volume de água é que passa pelo mesmo. Normalmente este tipo de fraude é cometida pelas instalações industriais e comerciais, sendo dos casos mais difícil de deteção. Este tipo de ligação encontra-se ilustrado na Figura 9. (Teixeira et al., 2014)

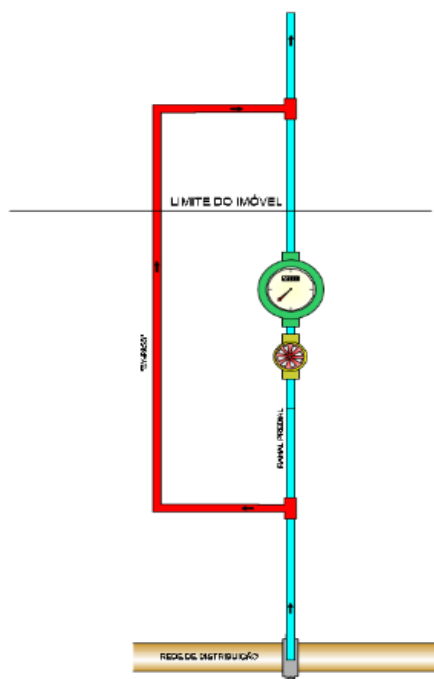


Figura 9: Esquema de uma ligação “bypass”.
(Lédo, 1999)

A derivação do ramal é outro tipo de ligação ilícita muito semelhante ao caso das ligações “bypass”. Existe um desvio do sentido natural da água, mas desta vez não se volta a conectar novamente ao ramal predial. Este tipo de ligação encontra-se representado na Figura 10. (Teixeira et al., 2014)

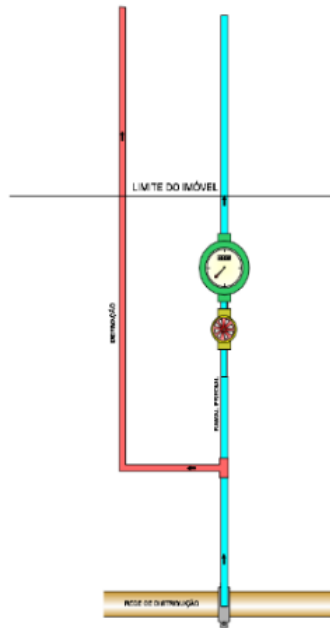


Figura 10: Esquema de uma ilicitude do tipo derivação do ramal.

(Lédo, 1999)

As ligações diretas são outro tipo de consumo não autorizado, que se caracteriza pela existência de um ramal de ligação conectado diretamente à rede de distribuição, o que significa que a água não irá passar pelo contador e o volume de água consumido não é contabilizado. Esta ligação encontra-se apresentada na Figura 11 e 12. (Teixeira et al., 2014)

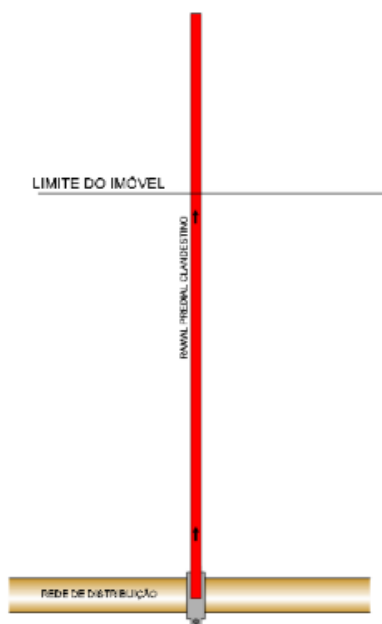


Figura 11: Esquema de uma ligação direta.

(Lédo, 1999)



Figura 11: Exemplo de uma ligação direta fornecida pela empresa Águas de Gaia, EM, S.A.

A adulteração do contador é outro tipo de consumo não autorizado, este tipo de ilicitude baseia-se essencialmente na manipulação do contador de forma a alterar as suas características. Fala-se, portanto, da quebra do lacre do contador, manipulação da relojoaria, inversão do sentido de escoamento, furo da cúpula e ainda a introdução de uma agulha que imobilize o movimento da relojoaria do contador. (Teixeira et al., 2014)

Outro tipo de consumo não autorizado é a utilização fraudulenta de marcos e bocas-de-incêndio na via pública, (Figura 12) este tipo de sistemas não tem contadores instalados o que leva a um maior consumo não autorizado da água nestes locais. A não colocação de contadores está relacionado com o facto de se evitar a ocorrência das respetivas perdas de carga localizadas. Em alguns casos é frequente o enchimento de tanques de veículos para rega ou lavagem de ruas, onde supostamente os marcos de incêndio apenas poderiam ser utilizados para serviços específicos pelas entidades detentoras que neste caso seriam os bombeiros, onde posteriormente essas mesmas situações teriam de ser reportados à EG. (Alegre et al., 2005)



Figura 12: Exemplo de uma ilicitude encontrada pela empresa Águas de Gaia, EM,S.A com marcos de incêndio.

Outra situação muitas vezes recorrente é o consumo de água após a suspensão do abastecimento da mesma, havendo uma posterior religação à rede de distribuição por parte dos clientes sem autorização, ou mesmo a violação do selo e posterior restabelecimento ilegal. (Teixeira et al., 2014) Uma dificuldade que também é muitas vezes encontrada é a impossibilidade de aceder ao interior das habitações quando o

contador se encontra no interior. Por isso mesmo, uma das medidas que devem ser implementadas é que nas novas habitações os contadores sejam colocados na parte exterior das habitações.

Algumas formas de minimizar o consumo não autorizado é através da inspeção da rede predial por parte das equipas de leituras e a implementação de contadores com telecontagem também designados de contadores “inteligentes”. (Teixeira et al., 2014)

A implementação de algumas medidas para o controlo das perdas aparentes são essencialmente no sentido de minimizar o consumo ilegal de água. Como tal, algumas dessas medidas são: a utilização do Sistema de Informação Geográfica (SIG) para comparar todos os dados de consumo de água que uma determinada tipologia habitacional com as mesmas características normalmente consome, pois cada cliente tem um consumo de água que irá depender da tipologia habitacional e do número de moradores, permitindo assim a identificação dos casos anómalos. A partir da determinação do consumo médio mensal por pessoa e com informação do número de moradores é possível prever o consumo para uma determinada habitação e, assim, identificar os consumos de água que possam ser superiores ou inferiores ao que seria de esperar. (Malheiro et al., 2011)

De forma a quantificar o consumo de uma habitação realizam-se leituras reais através do sistema de telemetria, que auxiliam também na monitorização de casos ilícitos. Por exemplo, quando é feito um corte de água e posteriormente se detetam leituras acima de zero, sabe-se que houve incumprimento por parte do cliente, uma vez que houve religação à rede de distribuição sem autorização. Implementam-se também contadores totalizadores à entrada dos edifícios para se comparar a quantidade de água que é fornecida com a que é realmente faturada individualmente por cada cliente, averiguando-se a existência de casos irregulares. (Malheiro et al., 2011)

Estes contadores com telecontagem são uma das formas de quantificar o consumo de água por via telemetria. Caracterizam-se por ser um sistema de controlo das perdas aparentes que permite a recolha de todas as leituras, da gestão dos parques de contadores e da atuação em caso de acidentes domiciliários à distância, permitindo o controlo em tempo real dos consumos de cada cliente, permitindo assim a eliminação do consumo com base em estimativas. Este tipo de contadores permite também quantificar o volume de água perdido, através da diferença entre o volume de água que entra no sistema e o volume efetivamente consumido e faturado, podendo-se aferir acerca da existência de perdas aparentes ou reais. Em suma, este sistema é capaz de transmitir alertas de anormalidades, tais como consumo excessivos ou baixos,

avarias ou bloqueios dos contadores, aumentando assim o êxito do combate das perdas. (Malheiro et al., 2011)

Ainda referente à telemetria, para que este sistema funcione é necessário ser instalado um concentrador que terá como principal função armazenar as leituras enviadas por todos os contadores e, posteriormente, transmiti-las. A distância a que deve estar o concentrador dos contadores deve ser o mais reduzida possível, pois a quantidade de objetos físicos que se encontram entre essa distância irá fazer com que a informação não seja bem recebida. Este sistema irá funcionar com a instalação de um contador totalizador instalado à entrada da zona de abastecimento de água, permitindo analisar a quantidade de água que entrou com a que foi realmente consumida. Assim, torna-se mais fácil o trabalho dos leitores, pois não terão a necessidade de entrarem nas casas dos clientes para recolherem as leituras. (Malheiro et al., 2011) Com os dados obtidos por este sistema e com o auxílio do balanço hídrico é possível calcular-se o valor de água não faturada.

Uma das medidas que a empresa Águas de Gaia, EM tem implementada para a minimização das perdas aparentes passa pela substituição dos contadores mais antigos, atualmente, por ano são substituídos por esta EG cerca de 8 a 10 mil contadores.

2.3 Dimensões do problema

2.3.1 Dimensão Económico-Financeira

A água que se encontra na rede de distribuição passa por diversas etapas, desde a captação, tratamento, elevação e transporte até ao consumidor final. Todas essas etapas acarretam custos para a entidade gestora e, por isso mesmo, a dimensão económica das perdas de água torna-se bastante importante para a empresa. O aumento do interesse por parte das entidades gestora por este parâmetro deve-se sobretudo ao facto destas entidades terem começado a comprar serviços a empresas exteriores. (Alegre et al., 2005)

Em Portugal Continental, no ano de 2004, o volume de perdas aparentes foi de aproximadamente 110 milhões de m³ e das perdas reais rondaram os 164 milhões de m³, sendo que nesse ano obteve-se um valor aproximado de 70 milhões de euros. (Alegre et al., 2005)

2.3.2 Dimensão Técnica

É impossível a existência de redes irredutíveis, podendo sempre ocorrer algumas roturas e fugas na rede. Por vezes, o tamanho da fuga ou da rotura não compensa a intervenção por parte da empresa devido aos custos que acarretam, contudo a empresa deve manter a atenção nestes casos de quebra na rede. De forma a controlar estas situações, em relação a um nível mais técnico é possível interceder em dois principais pontos: (Alegre et al., 2005)

- Numa vertente mais física, através de uma reparação pontual;
- Em termos de operação, é importante intervir no sentido de diminuir a probabilidade de contaminações exteriores e da gestão das pressões de forma a não interferir na boa qualidade de serviço prestado aos consumidores.

2.3.3 Dimensão Ambiental

A dimensão ambiental das perdas de água tem maior importância em zonas com maior falta de água para consumo humano, tornando-se mais significativa com a implementação da Diretiva Quadro da Água. Com a sua entrada em vigor limitou-se a construção de novas captações, não permitindo ainda reforçar as captações já existentes se a jusante das captações as perdas de água forem significativas. Assim, é importante para a empresa saber qual é o volume de água que entra no sistema e qual o volume de água que é perdido por roturas e fugas, de forma a quantificar a quantidade de água que é efetivamente perdida e, assim, avaliar se é necessário a criação de novos pontos de captação. (Alegre et al., 2005)

2.3.4 Dimensão de Saúde Pública

Os locais de fugas e roturas são pontos que se encontram mais expostos às contaminações existentes no solo, havendo uma maior probabilidade de contaminação da água que é distribuída aos clientes. Assim, a entidade gestora terá de corrigir estes problemas, mesmo que muitas vezes não compense o investimento, devido ao impacto que poderá ter na saúde pública. Apesar da adição de cloro (desinfetante) ser uma técnica utilizada para tentar reduzir o problema, esta não é totalmente confiável, sendo preferível atuar no foco do problema. (Alegre et al., 2005)

2.3.5 Dimensão Social

Para uma boa qualidade da água o preço dela será mais elevado, pois acarreta maiores custos de tratamento. Se a população estiver corretamente informada tem também um papel fulcral na dimensão social, pois poderão alertar para a localização de possíveis fugas e roturas visíveis e controlar o consumo de água das suas habitações, que levará à diminuição das perdas de água, desta forma, o aumento do valor da água pode ser impedido se os problemas forem impedidos antecipadamente. (Alegre et al., 2005)

2.4 Balanço Hídrico

A *International Water Association* (IWA) desenvolveu uma metodologia para o cálculo do balanço hídrico que consiste num conjunto de conceitos, sendo que os mais relevantes são: volume de água que entra no sistema, consumo autorizado faturado e não faturado, perdas de água reais e aparentes, água faturada e água não faturada, esta metodologia é possível de se visualizar na Figura 13. (Alegre et al., 2005; Carvalho, 2014)

- **Volume Entrado no Sistema:** corresponde à quantidade de água que entra no sistema de abastecimento de água e que irá ser fundamental para o cálculo do balanço hídrico; trata-se assim de toda a água que é captada e tratada.

- **Consumo Autorizado Faturado:** este consumo está relacionado com a quantidade de água gasta por parte dos clientes que estão ligados à rede, sendo o principal parâmetro responsável pela receita da entidade gestora.

- **Consumo Autorizado Não Faturado:** diz respeito a todo o consumo que esteja autorizado pela entidade gestora mas que não é faturado pela mesma, como é o caso da água utilizada no combate aos incêndios, para a lavagem de arruamentos, para a rega dos espaços verdes municipais, toda a água que seja necessária para as obras, a água utilizada para o enchimento dos fontanários e lavadouros, por fim, toda a água que seja utilizada para a lavagem das condutas e dos coletores da entidade gestora.

- **Perdas Reais:** correspondem a toda a água que foi perdida ao longo da rede por roturas e fissuras. O volume de água perdido por estas fugas irá depender do tamanho,

da duração e do caudal de cada fuga, portanto este parâmetro será influenciado pelas fugas dos reservatórios, fugas na rede, fugas nas ligações e roturas.

- **Perdas Aparentes:** estão relacionadas com ilicitudes do tipo: erros de medição da água que é consumida e por consumo não autorizado, como são exemplo, as ligações sem autorização à rede de distribuição.

- **Água Faturada:** corresponde ao mesmo valor do consumo autorizado faturado, valor esse que é relativo à receita da entidade gestora.

- **Água Não Faturada:** refere-se à diferença entre o volume de água que entra no sistema e o consumo autorizado faturado. Este valor será influenciado pelas perdas de água totais (perdas reais e perdas aparentes) e pelo consumo autorizado não faturado. Este valor irá corresponder à quantidade de água que não é contabilizada e, por isso, não entrará para os valores da receita da entidade gestora. Por este motivo, este é um dos parâmetros que a empresa deverá ter em consideração porque influencia a eficiência da mesma.

Água Entrada no Sistema [m ³ /ano]	Consumo Autorizado [m ³ /ano]	Consumo Autorizado Faturado [m ³ /ano]	Consumo Faturado Medido [m ³ /ano]	Água Faturada [m ³ /ano]
		Consumo Autorizado Não Faturado [m ³ /ano]	Consumo Autorizado Não Faturado Medido [m ³ /ano]	
			Consumo Autorizado Não Faturado Não Medido [m ³ /ano]	
		Perdas de Água [m ³ /ano]	Perdas Aparentes [m ³ /ano]	Consumo Não Autorizado [m ³ /ano]
	Erros de Medição [m ³ /ano]			
	Perdas Reais [m ³ /ano]		Fugas nas Conduções de Adução e Distribuição [m ³ /ano]	
			Fugas e Extravasamentos nos Reservatórios [m ³ /ano]	
	Fugas nos Ramais de Distribuição a Montante do Contador [m ³ /ano]			

Figura 13: Balanço Hídrico

2.5 *Machine Learning*

O presente trabalho tem como objetivo de prever os casos com ilicitudes desconhecidos, para isso irá necessitar do tratamento e análise de uma quantidade considerável de dados. Por esse motivo, a utilização da linguagem *Python* como forma de tratamento dos dados fornecidos é uma opção satisfatória. Esta linguagem é um tipo de programação que se caracteriza essencialmente por ser uma linguagem dinâmica e de código aberto.

Segundo Arthur Samuel, um dos primeiros criadores americanos dos jogos de computadores, da inteligência artificial e do ML define este último parâmetro como sendo “o campo de estudo que fornece aos computadores a capacidade de aprenderem. Pode-se também definir o *machine learning* como sendo um conjunto de algoritmos e técnicas que são usadas para a criação sistemas computacionais que consigam através dos dados fornecidos fazer previsões e, desta forma, solucionar os problemas existentes”. (Swamynathan, 2019)

Quanto à inteligência artificial (IA) esta ramifica-se em diversas áreas, desde a robótica, otimização, ML, entre outras que se encontram representadas na Figura 14.

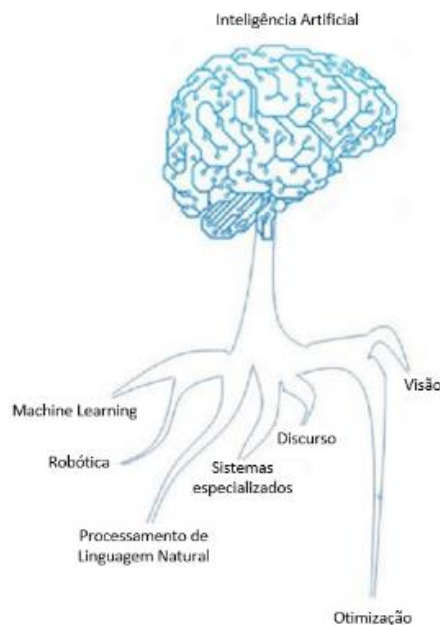


Figura 14: Ramos da Inteligência Artificial.

(adaptada de Swamynathan, 2019)

A IA é uma das áreas que, ao longo dos anos, tem apresentado um notável crescimento e o ML é um dos ramos que acompanha esse desenvolvimento. O ML assenta na ideia de que um sistema, através do fornecimento de um conjunto de dados, tem a capacidade de aprender a relacionar padrões e, posteriormente, permite resolver problemas das mais diversas áreas. (Swamynathan, 2019)

Assim sendo, aplicar esta ferramenta ao presente trabalho poderá ser uma forma simples, rápida e com reduzida necessidade de recursos, para detetar os casos de consumo não autorizado na entidade gestora em estudo. Portanto, é possível através da utilização de alguns algoritmos de ML, com o fornecimento de dados relativos aos clientes estabelecer alguns padrões de forma a detetar casos ilícitos que até ao momento eram desconhecidos.

A IA ao longo dos anos tem vindo a sofrer uma evolução ganhando aos poucos um lugar de destaque na sociedade atual. Antigamente, a inteligência artificial de uma máquina era igual ou superior à inteligência do cérebro humano conseguindo ambos solucionar um determinado problema. Atualmente, a IA consegue não apenas solucionar, mas, também resolver qualquer problema de uma determinada área, em vez de solucionar apenas um problema de uma área em específico. Futuramente, espera-se atingir inteligências acima da média, o que significa que a inteligência artificial será bem mais evoluída e desenvolvida em todas as áreas do que o cérebro humano. (Swamynathan, 2019)

Hoje em dia, as empresas usam a análise de dados para que seja possível analisar a eficiência da mesma de forma a aumentar o retorno do investimento. Desta forma, é necessário escolher o tipo de análise mais adequada. Para o caso do trabalho que se vai estudar a análise preditiva será o tipo de análise mais apropriada, pois um dos exemplos que este tipo de análise trata é da deteção de fraudes e está relacionado com a procura de padrões passados para se conseguir prever futuros acontecimentos.

Para isso será necessário utilizar a IA e o ML para avaliar no caso em questão os dados atuais da empresa e conseguir prever sobre os dados futuros.

O ML consiste na análise dos dados fornecidos inicialmente pela empresa e irá ser possível aplicar um algoritmo de forma a encontrar a relação que existe entre eles e assim ser possível prever os ilícitos desconhecidos.

Um modelo de ML deve basear-se em dois principais princípios: (Swamynathan, 2019)

- Encontrar os principais padrões no conjunto de dados inicialmente fornecidos;

- Ser o mais generalista possível de forma a conseguir funcionar com qualquer conjunto de dados que seja fornecido;

O ML subdivide-se em três principais grupos: aprendizagem supervisionada, aprendizagem não supervisionada e aprendizagem por reforço, sendo que estes tipos são diferenciados com base nos valores de saída que se deseja. (Figura 15)

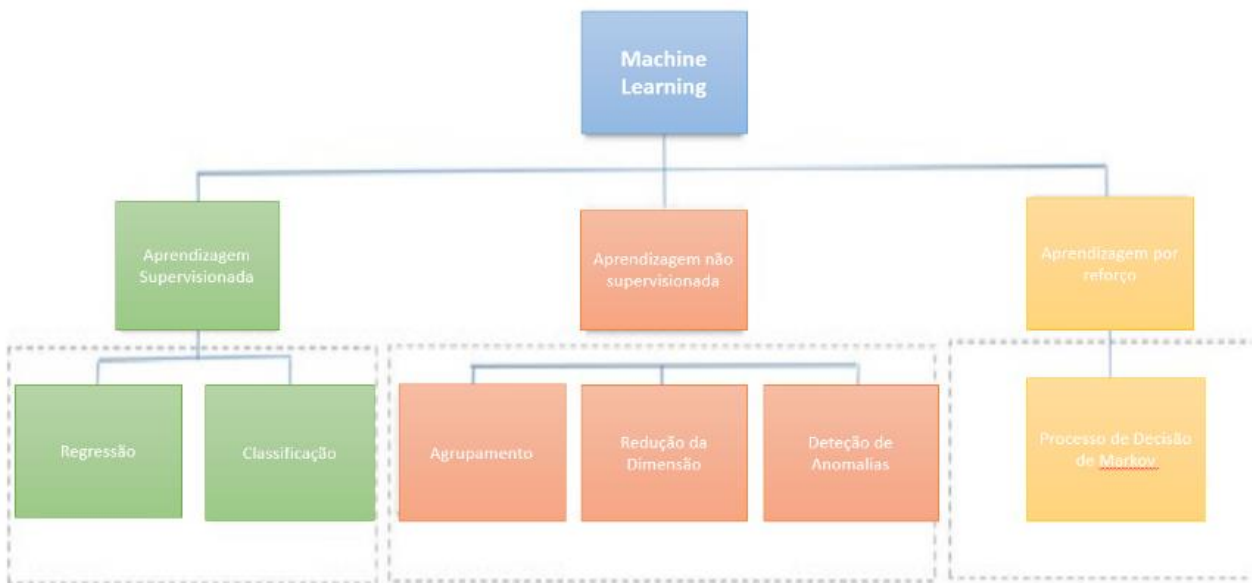


Figura 15: Classificação do ML.

(adaptada de Swamynathan, 2019)

A aprendizagem supervisionada consiste num algoritmo de *ML* que tem bastantes dados de entrada. O objetivo deste algoritmo é aprender com os valores padrão. Este tipo subdivide-se em regressão e classificação, sendo que o primeiro consiste na saída de dados consoante a quantidade de dados de entrada. Já a classificação baseia-se no número de classes que devem ser previstas podendo ser duas ou mais. O algoritmo tem por base os valores padrão de entrada mais relevantes de cada classe a partir dos dados antigos e, assim, é capaz de prever a classe ou dados futuros. (*Supervised Learning: Predicting an Output Variable from High-Dimensional Observations*, 2020; Swamynathan, 2019)

A construção de um modelo em ML do tipo aprendizagem supervisionada possui três fases: (Swamynathan, 2019)

- Treinar: O algoritmo baseia-se nos dados de entrada padrão e prevê cada valor de saída dando origem ao modelo.
- Teste e validação: Nesta fase é avaliado o desempenho do modelo proposto, aplicando para isso um conjunto de dados que não foi utilizado para a criação do modelo em teste.
- Previsão: Nesta etapa realiza-se o mesmo que na etapa anteriormente mencionada contudo, aqui, é necessário verificar-se o funcionamento normal para se poder aplicar o modelo.

Já a aprendizagem não supervisionada consiste em situações que não se consegue prever os dados de saída, estudando-se o padrão dos dados de entrada e tentando-se encontrar uma relação para os dados de saída. Este tipo de aprendizagem divide-se em: agrupamento, redução da dimensão e deteção anómala.

O agrupamento consiste em casos em que não se sabe as classes para um determinado conjunto de dados. Como tal, o principal objetivo será dividir o conjunto de dados de entrada em classes relacionadas. Já a redução da dimensão baseia-se em simplificar um grande conjunto de dados de entrada para um espaço dimensional inferior. Por último, a deteção de anomalias baseia-se na identificação de valores que não estão em conformidade com o comportamento que seria de esperar em relação a outros valores.

O sistema que terá aplicabilidade para o caso de estudo é a aprendizagem supervisionada, pois a partir dos dados fornecidos ao sistema é possível estabelecer um padrão com os dados de entrada e, assim detetar outros possíveis casos.

2.5.1 Rede Neuronal Humana e Rede Neuronal Artificial

O cérebro humano consiste numa complexa rede neuronal conectada, onde cada região do cérebro está associada a uma tarefa específica, essas mesmas zonas recebem os sinais para que seguidamente sejam executadas as ações necessárias. Toda esta informação será enviada através dos neurónios até ao cérebro onde toda a informação será decodificada para que depois seja possível processar a resposta mais adequada. (Swamynathan, 2019)

O funcionamento dos neurónios biológicos (Figura 16) inicia-se com a receção da informação através das dendrites; o corpo celular será o responsável pelo

processamento da informação e por fim, o terminal axônio é que se encarregará de transferir a informação aos outros neurónios.

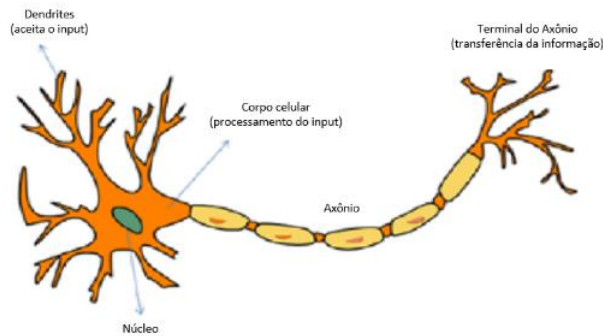


Figura 16: Constituição de um neurônio biológico.

(adaptada de Swamynathan, 2019)

A IA funciona com base em *perceptron* designação dada aos neurónios artificiais base da rede neuronal artificial. Este tipo de neurónios foi inspirado nos neurónios biológicos.

Os neurónios artificiais são constituídos por várias camadas de entrada que irão ser as responsáveis por aceitar as amostras representadas. A etapa seguinte baseia-se no processamento para que os pesos (w) sejam ajustados de modo a que os erros de saída sejam minimizados como pode ser observado na Figura 17. Por fim, o resultado será encaminhado para uma função de ativação que irá produzir um valor de saída. (Swamynathan, 2019)

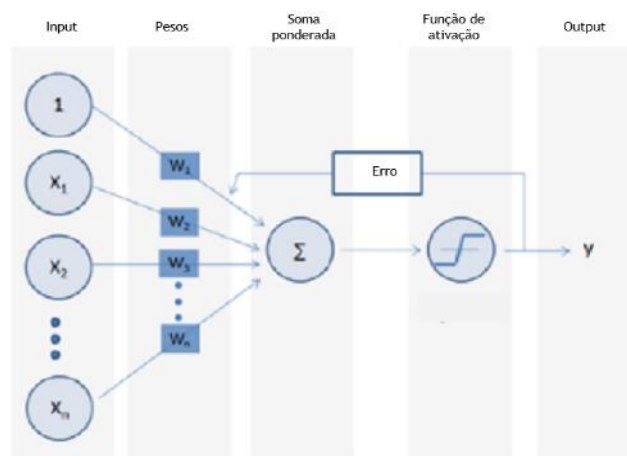


Figura 17: Constituição de um neurônio artificial.

(adaptada de Swamynathan, 2019)

Quando o conjunto de dados é complexo o uso de poucos neurónios pode resultar num modelo mal ajustado, contudo a utilização de muitos neurónios poderá resultar num demasiado ajuste do modelo. Desta forma, para se construir uma rede neuronal eficiente é necessário definir primeiramente quantas camadas ocultas e quantos neurónios deverá ter cada uma delas, de forma a criar-se um modelo ideal. Normalmente, o número de neurónios na camada oculta pode idealmente ser a média dos neurónios que se encontram nas camadas de entrada e saída. (Swamynathan, 2019)

2.5.2 Criação de um Modelo

A criação de um modelo que seja o mais eficiente possível para a análise dos dados em estudo baseia-se em 5 principais etapas: (Swamynathan, 2019)

Amostra: Consiste na seleção de um conjunto de dados dentro da amostra geral que foi introduzida para a criação do modelo. Nesta fase, os dados são também divididos em dados de teste e em dados de treino.

Exploração: Nesta etapa são realizados testes de forma a perceber quais são os dados que apresentam lacunas. Para essa análise são usados dois métodos, a análise univariada e a análise multivariada. Na primeira, cada variável procura apenas compreender individualmente a sua distribuição; já a análise multivariada consiste na exploração da relação que existe entre as diferentes variáveis.

Modificar: Nesta fase as variáveis tanto podem ser eliminadas como podem ser transformadas. O resultado desta fase é um conjunto de dados que se encontram preparados e que podem ser passados para o algoritmo de ML para a construção do modelo.

Modelo: Nesta etapa várias técnicas de modelagem de dados são aplicadas aos dados pré-processados para ser possível comparar o seu desempenho com os resultados desejados.

Avaliar: Esta última etapa baseia-se na avaliação do modelo em comparação com os dados de teste para garantir a veracidade do modelo.

Bibliotecas Python

Existem algumas bibliotecas de código aberto que já estão inseridas na linguagem *Python* e que são fundamentais para o bom desenvolvimento do tratamento dos dados, sendo designadas principalmente por bibliotecas científicas de *Python*. Estas bibliotecas podem ser divididas em bibliotecas para a análise de dados e bibliotecas principais de ML. (Swamynathan, 2019)

As bibliotecas principais de ML correspondem a um conjunto de bibliotecas que fornecem todos os algoritmos e funcionalidades necessários para o ML, normalmente são aplicadas num determinado conjunto de dados de forma a extrair o padrão dos dados. (Swamynathan, 2019)

As bibliotecas responsáveis pela análise de dados consistem num conjunto de bibliotecas que fornecem as funcionalidades matemáticas e científicas essenciais para ser possível executar o pré-processamento e transformação de dados. As 4 principais bibliotecas usadas para a análise de dados são: *NumPy*, *SciPy*, *Matplotlib*, *Pandas*. O *Pandas*, *NumPy* e o *Matplotlib* desempenham o papel mais importante para a análise dos dados, já o *SciPy* irá complementar a biblioteca de *NumPy*. (Swamynathan, 2019) Estas três bibliotecas serão as mesmas a serem utilizadas para a análise dos dados do caso de estudo.

- NumPy

O *NumPy* é a principal biblioteca de computação científica em *Python*. Fornece um objeto de matriz multidimensional de alto desempenho e as ferramentas necessárias para trabalhar com estas matrizes. A matriz *NumPy* baseia-se num conjunto de valores todos com números semelhantes, e é classificada segundo o número de dimensões e a forma da mesma. O *NumPy* tem ainda muitas funções internas que irão ajudar na criação de matrizes.

- Pandas

O *Pandas* é uma biblioteca *Python* de código aberto que fornece estruturas de dados rápidos, flexíveis e expressivos tendo sido projetados para facilitar o trabalho. O *Pandas* é ainda uma biblioteca adequada para dados em colunas heterogêneas, como é o caso das tabelas de SQL e de Excel.

- Matplotlib

O Matplotlib é uma extensão numérica da matemática NumPy e uma excelente biblioteca para exibir ou apresentar dados num formato gráfico. Este tipo de análise irá permitir que os analistas vejam os dados apresentados graficamente, para que assim seja mais fácil entender conceitos mais difíceis e também ser mais fácil identificar novos padrões.

Alguns dos gráficos mais usados são:

- plt.bar: gráfico de barras
- plt.scatter: gráfico de dispersão
- plt.boxplot: gráfico de caixa
- plt.hist: histograma
- plt.plot: gráfico de linhas

O *Python* possui uma infinidade de bibliotecas de ML de código aberto. A biblioteca de ML *Python* que vai ser abordada neste trabalho e que também é das mais conhecidas e comumente usadas é a biblioteca *Scikit Learn*, ou também designada por *sklearn*. Esta biblioteca é construída sobre o *SciPy* (Scientific Python) e apresenta um elevado número de algoritmos de aprendizagem supervisionada e não supervisionada. Também fornece várias ferramentas para ajustar o modelo, para a previsão, processamento de dados, para a seleção e avaliação do modelo. (Swamynathan, 2019)

- Scikit Learn

O *scikit learn* é uma biblioteca de ML de código aberto que suporta a aprendizagem supervisionada e não supervisionada, é das melhores bibliotecas em *python* para se criar e treinar modelos de ML, por esse motivo vamos trabalhar com ela para o modelo que vai ser desenvolvido. Tem como principais funções ajudar no ajuste do modelo, pré-processamento de dados, seleção e avaliação do modelo. (Swamynathan, 2019)

Esta biblioteca fornece vários algoritmos e modelos de ML, cada algoritmo (KNN, RFC, LRN, SVM) podem ser ajustados a alguns dados usando para isso o seu método de ajuste. Uma vez ajustado o algoritmo ele pode ser usado para prever valores de novos dados, para se verificar se esses valores foram bem avaliados, será necessário avaliar cada algoritmo.

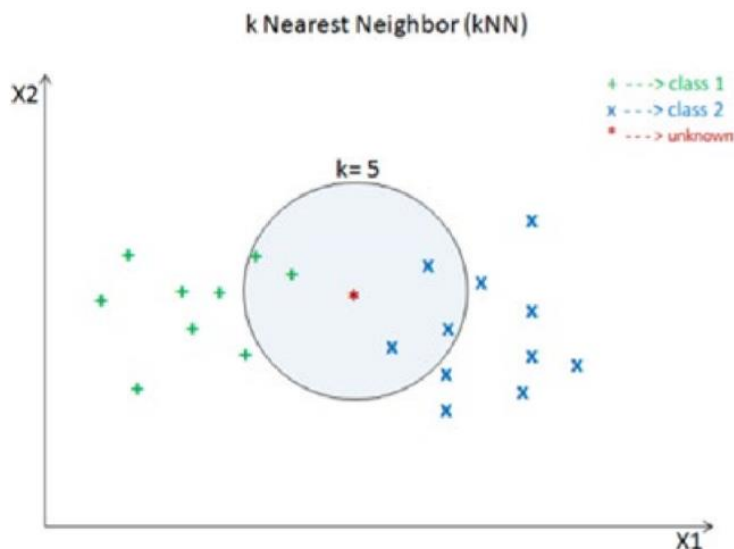
2.5.3 Modelos de Classificação

Para se avaliar o conjunto de dados fornecidos pela empresa é necessário inicialmente treinar e validar um modelo de aprendizagem supervisionada aplicando técnicas de ML a dados antigos. Em seguida, a esse modelo aplica-se o novo conjunto de dados para assim se prever os valores futuros.

Os modelos de classificação de aprendizagem supervisionada supramencionados são: o K Nearest Neighbors (KNN), o Logistic Regression (LRN), o Support Vector Machine (SVM), e o Random Forest Classifier (RFC).

O primeiro modelo a ser estudado é o KNN, que foi desenvolvido a partir da necessidade de realizar análises mais pormenorizadas quando os valores estimados pretendidos são desconhecidos ou de difícil determinação. Este método é conhecido como a “a regra do vizinho mais próximo” uma vez que consiste na classificação de padrões. Este algoritmo irá funcionar com base nos dados (“vizinhos”) que se encontram mais próximos do valor a estimar, designando essa classe por k igual ao número de dados que se encontram em torno do valor a estimar. Como é possível de se observar na Figura 18, onde $k=5$ porque é o número de dados que se encontram em torno do valor desconhecido (ponto a cor vermelho), esta distância do valor desconhecido aos valores conhecidos é estimada por nós. O valor desconhecido será classificado consoante a classe de valores conhecidos que se encontra mais representada em torno do mesmo, que neste caso seria a classe 2. (*Scikit-Learn: KNeighbors Classifier*, 2020; Swamynathan, 2019)

A principal desvantagem do modelo KNN é a dificuldade em encontrar os dados mais próximos para cada amostra. Dois pontos que também se deve ter em atenção quando se trabalha com este modelo é escolher um valor ímpar para o parâmetro k , quando o problema tem duas classes e o valor de k não deve ser múltiplo do número de classes. (Swamynathan, 2019)

Figura 18: Modelo de classificação do KNN com $k=5$.

(Swamynathan, 2019)

O modelo LRN pode ser usado para prever e ajustar os dados, como também pode ser usada para prever a variável dependente. Quando existe um aumento do número de variáveis, a probabilidade de excesso de ajuste também aumenta, contudo, é importante evitar-se os ajustes excessivos. (*Scikit-Learn: Logistic Regression*, 2020; Swamynathan, 2019)

Já o modelo SVM é caracterizado pelo seu principal objetivo de desenhar um hiperplano que separe as duas classes de forma ideal, de maneira a que a margem seja máxima entre o hiperplano e as observações. Como é possível de se observar na Figura 19, existe a hipótese de existirem diferentes hiperplanos. Contudo, o objetivo deste modelo é mesmo encontrar o hiperplano que nos dá uma margem mais alta. (*Scikit-Learn: Support Vector Classification*, 2020; Swamynathan, 2019)

O modelo SVM comparativamente ao modelo LRN é menos propício a uma disparidade de valores, pois o SVM apenas se preocupa com os pontos que estão mais próximos do limite de decisão. (Figura 22)

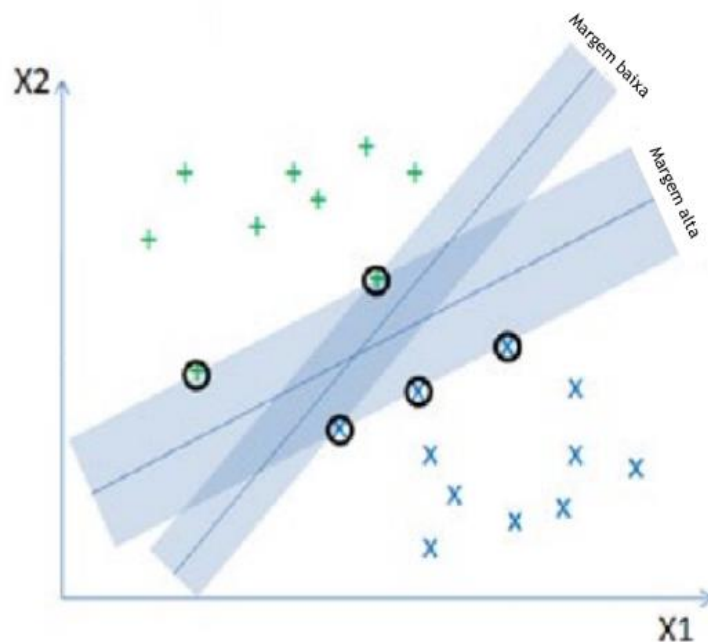


Figura 19: Modelo de classificação SVM.

(adaptada de Swamynathan, 2019)

Por último, o modelo RFC é caracterizado por um subconjunto de observações e um subconjunto de variáveis que são selecionadas aleatoriamente de forma a criar vários modelos independentes. (*Scikit-Learn: Random Forest Classifier*, 2020)

Assim, de forma a se criar um modelo generalizado robusto deve-se combinar modelos que não sejam semelhantes.

2.5.4 Anaconda Navigator

O *Anaconda Navigator* é um *software* de código aberto mais conhecido de todo mundo das linguagens de programação *Python* e *R* com uma vertente para o ML que vai ser utilizado para elaborar o nosso modelo.

O *Spyder* (*Scientific Python Development Environment*) é um ambiente de desenvolvimento integrado escrito em *Python* que se encontra incluído no *Anaconda*, portanto, depois de instalado o *Anaconda* inicia-se o *spyder* executando para isso o comando *spyder*, como se pode observar na Figura 20. (*Anaconda Documentation: Spyder*, 2020)

Este ambiente científico apresenta uma combinação de funcionalidades avançadas de edição, análise, execução interativa e bons recursos de visualização. Para além de todas estas funcionalidades o spyder apresenta várias bibliotecas científicas muito conhecidas, das quais: o *Numpy*, o *Pandas* e o *Matplotlib*. (Anaconda: *Spyder*, 2020)

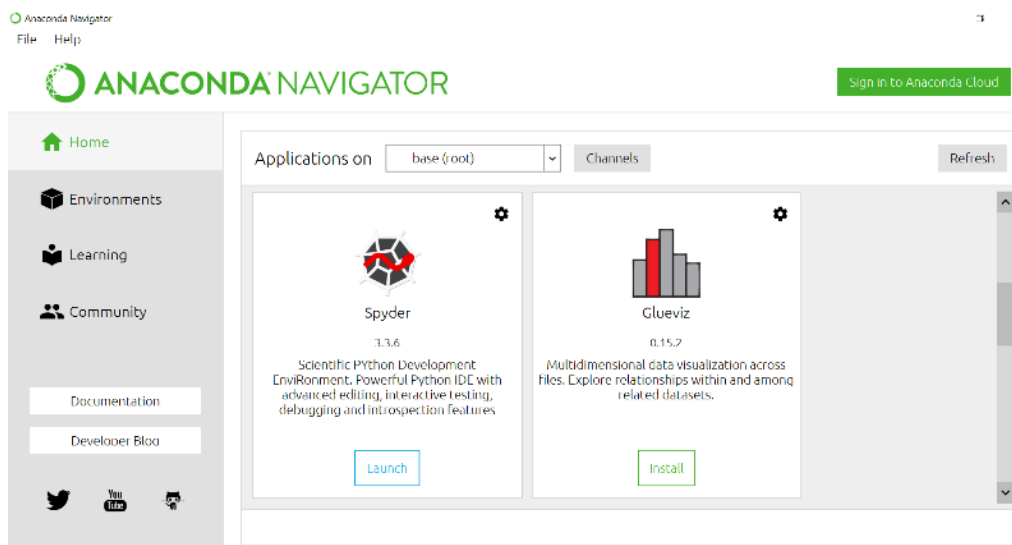


Figura 20: *Anaconda Navigator*.

3 A Empresa

3.1 Águas de Gaia EM,S.A.

3.1.1 Caracterização da Entidade Gestora

O nível de excelência de qualquer empresa deste setor só é possível caso o sistema de gestão seja economicamente sustentável e os valores de água não faturada sejam o mais reduzido possível.

As empresas de águas além de terem como principal função garantir os serviços públicos de abastecimento de água potável e da rede de saneamento, deve também ter em atenção a devolução da água residual às Estações de Tratamento de Águas Residuais (ETAR). Contudo, atualmente, o objetivo é realizar um ciclo da água residual, onde a água tratada pela ETAR, em vez de ser devolvida aos rios, seja diretamente direcionada novamente nas redes de distribuição.

No dia 12 de Abril de 1999, os então Serviços Municipalizados de Água e Saneamento deram origem à empresa hoje em dia designada como Águas de Gaia, EM, S.A.

A empresa municipal Águas de Gaia, EM, S.A. é uma entidade gestora que tem como principal missão o abastecimento público de água; saneamento de águas residuais urbanas, gestão da drenagem de águas pluviais, contribuindo para a despoluição do rio Douro e das ribeiras do concelho; gestão de resíduos sólidos urbanos; promoção da melhor utilização das praias e zonas balneares do concelho e o contributo para a qualidade de vida da população, estes são os principais objetivos desta empresa.

Atualmente, a empresa Águas de Gaia, EM,S.A. trabalha em conjunto com duas outras empresas, a Simdouro e as Águas do Douro e Paiva. A primeira responsabiliza-se pelo tratamento das águas residuais proveniente do saneamento dos habitantes de Vila Nova de Gaia. Já a empresa Águas do Douro e Paiva, mais precisamente a ETA de Lever, encarrega-se da captação e tratamento da água potável para ser distribuída por toda a população. Na Figura 21 é possível observar toda a rede de abastecimento de água potável do município de Vila Nova de Gaia que se encontra ao encargo da empresa Águas de Gaia, EM, S.A.



Figura 21: Mapa da rede de distribuição de água potável do município de Vila Nova de Gaia.

(Águas de Gaia, EM, 2020a)

As empresas de água no que diz respeito à gestão dos sistemas de abastecimento têm como um dos principais objetivos a redução da quantidade de água não faturada. No que diz respeito à empresa Águas de Gaia, EM, no ano de 2018, segundo os dados do balanço hídrico do relatório de contas esse valor era de 29,3%, já no ano de 2019 este valor foi de 27,4%. (Águas de Gaia, EM, 2018, 2020b)

O caso das Águas de Gaia foi estudado no ano de 1998 quando a percentagem de água não faturada foi de 47,2%, tendo sido necessária a aplicação de algumas medidas de forma a reduzir as perdas reais e as perdas aparentes. O modo de funcionamento do controlo de perdas e dos setores envolvidos da entidade gestora são os seguintes: Centro de Atendimento Permanente; Direção de Águas de Abastecimento (DAA); Equipa de deteção de fugas (EDF) e Laboratório de contadores. Deste modo, a deteção das perdas reais é realizada da seguinte forma: a DAA analisa diariamente, através de um sistema de telegestão que se baseia na recolha de valores relativamente aos parâmetros, níveis dos reservatórios, níveis de pressão da rede de distribuição, valores

mínimos e máximos do caudal de adução, caudal de distribuição bem como os níveis de cloro na rede, onde posteriormente serão enviados para as salas de comando da entidade gestora. Todos estes valores são apresentados de forma gráfica para que seja possível uma melhor avaliação dos mesmos. (Figuras 23, 24 e 25) Quando algum destes níveis ultrapassam os limites admissíveis é lançado um alerta via mensagem telefónica do reservatório ou do local que tem a possível fuga onde, com o auxílio dos equipamentos específicos, descobrem a localização exata das fugas na rede de distribuição para a posterior reparação. As fugas normalmente são detetadas durante o período noturno onde o consumo de água é mais baixo.



Figura 22: Sistema de Telegestão da empresa Águas de Gaia, EM, S.A.

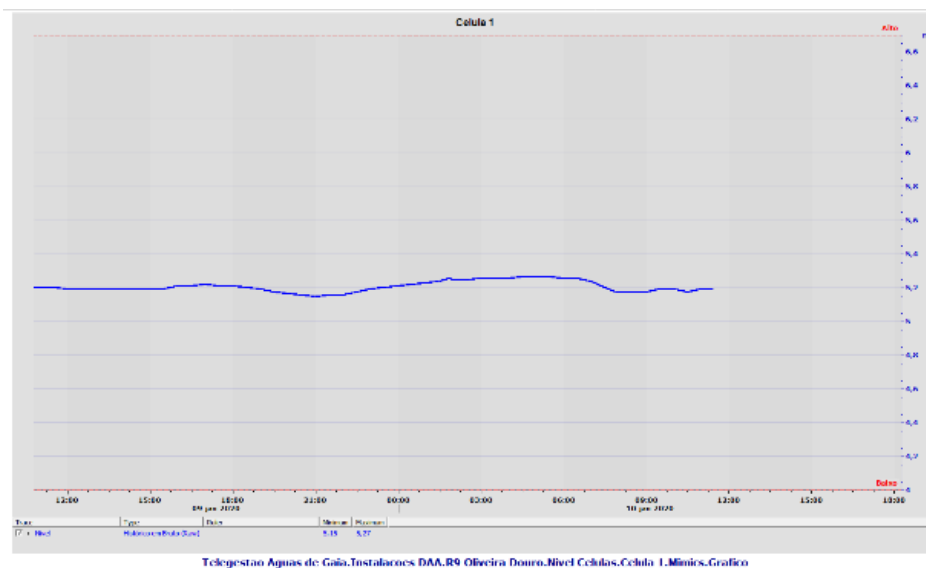


Figura 23: Gráfico do nível de um reservatório ao longo de um dia.

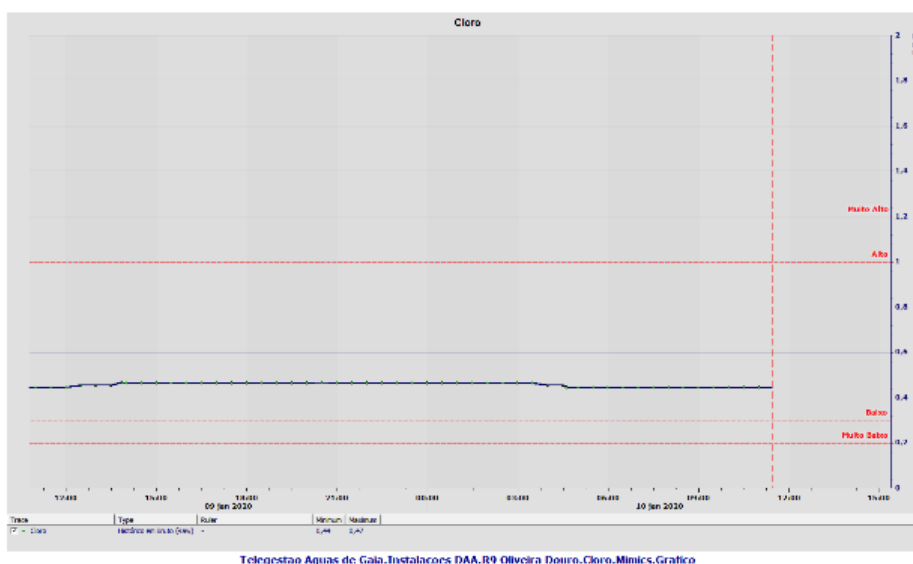


Figura 24: Gráfico do nível de cloro ao longo de um dia.

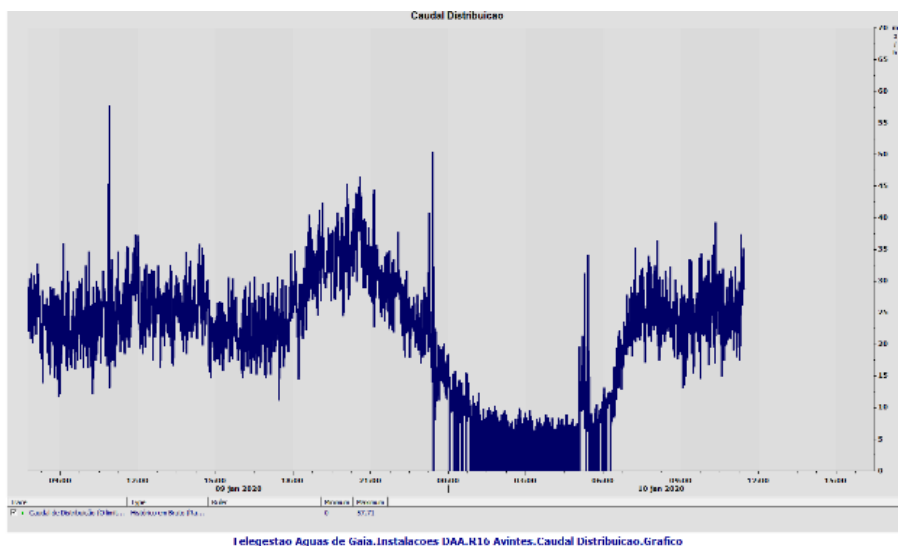


Figura 25: Gráfico do consumo diário de um reservatório.

Atualmente, esta entidade gestora foi a empresa pioneira de um projeto piloto que foi implementado em Novembro do ano 2019 para a deteção de fugas e roturas na rede de distribuição, mais precisamente para a diminuição das perdas reais principalmente das que não são visualmente detetadas. Este projeto baseia-se num modelo semelhante ao utilizado em Marte para a deteção de água. Quando aplicado ao caso das empresas de água, este modelo consiste na deteção de cloro por um satélite onde em seguida com o auxílio do SIG é possível analisar com maior pormenor os casos detetados e averiguar se não se tratam de piscinas, sistemas de rega ou fontanários. (Figura 26) Por fim, com a ajuda dos equipamentos acústicos deteta-se com maior rigor a localização exata das fugas de água, diminuindo desta forma a percentagem de água não faturada. (Figura 27)



Figura 26: Imagem das zonas onde existe uma perda de água.

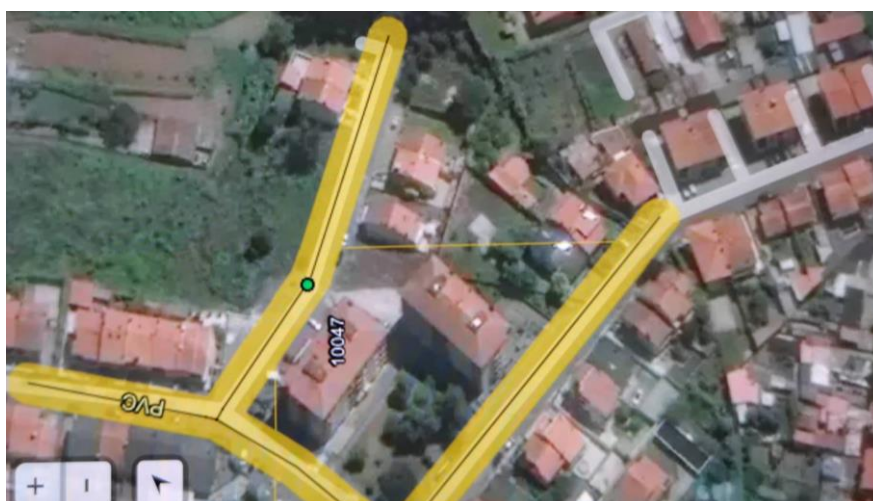


Figura 27: Imagem da localização exata da perda de água.

Quanto às perdas aparentes, atualmente a empresa Águas de Gaia adota um sistema computacional de deteção do consumo ilegal de água através de ligações ilícitas pelos consumos anómalos, isto é, que tem consumos nulos, que sofreram um decréscimo ou um aumento mais acentuado comparativamente aos consumos anteriores. Este sistema permite um alerta de forma a ser analisado cada caso em particular para seguidamente passar para a fase de vistoria e, assim, verificar no local a existência ou não de alguma anomalia.

Uma outra forma de detetar os casos ilícitos ocorre aquando da substituição de um contador, em que é avaliado no terreno se houve alguma adulteração do equipamento.

No entanto, apesar de todas as medidas tomadas existem muitas ilicitudes que são desconhecidas e, por isso, surge a necessidade de desenvolver novas metodologias que permitam estreitar o espectro de detecção de casos ilícitos.

As perdas aparentes podem ainda ser minimizadas pela realização de ensaios a contadores usados pelo laboratório de contadores, sabendo o fabricante e a idade dos contadores que entram no laboratório, conseguem estimar qual o erro associado à sua medição. Assim, quando se encontrar na rede de distribuição contadores do mesmo tipo sabe-se qual será o erro de medição associado às leituras. Desta forma, consegue-se averiguar se é necessária a instalação de outro contador, pois com o passar dos anos os contadores vão perdendo precisão, deixando de contabilizar de forma correta a água consumida.

Algumas das medidas que podem ser implementadas para a redução das perdas aparentes é a diminuição das leituras por estimativa, substituição dos contadores antigos, substituição das bocas-de-incêndio por marcos de incêndio, visto que estes últimos são de mais difícil acesso, diminuindo assim os consumos não autorizados.

É de notar mais uma vez que poucas são as medidas implementadas para a diminuição das perdas aparentes. Desta forma, torna-se fulcral a implementação de mais metodologias de forma a diminuir este tipo de perdas. Uma possível solução para a minimização deste problema será estudada seguidamente, e baseia-se na aplicação de algoritmos de ML.

3.1.2 Balanço Hídrico da Empresa

O número de clientes com consumo de água potável tem vindo a aumentar ao longo dos anos. Esta evolução deve-se essencialmente ao crescimento populacional, bem como ao aumento da preocupação das populações pela saúde e bem-estar, como se pode analisar pela Figura 28.

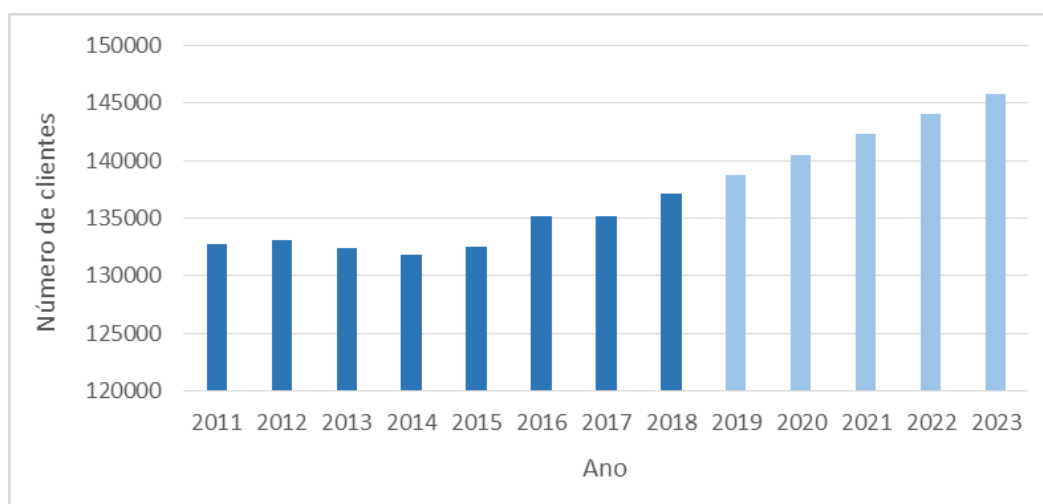


Figura 28: Evolução do número de clientes na empresa Águas de Gaia, EM, S.A.

Seguidamente, e de forma a analisar a evolução da empresa Águas de Gaia, EM, S.A., entre os anos 2011 e 2019, compilou-se na Tabela 1 os conceitos mais relevantes para o cálculo de um balanço hídrico para uma entidade gestora, de forma a analisar como foi evoluindo cada um desses parâmetros ao longo dos anos. (Águas de Gaia, EM, 2015, 2018, 2020b)

Tabela 1: Valores dos parâmetros mais relevantes do balanço hídrico da empresa Águas de Gaia, EM, S.A.

	2011	2012	2013	2014	2015	2016	2017	2018	2019
Volume entrado no sistema (m ³)	18717756 (100%)	18126786 (100%)	18396207 (100%)	18297752 (100%)	18585181 (100%)	18677179 (100%)	19169081 (100%)	18875876 (100%)	19563715 (100%)
Água Faturada									
Consumo autorizado faturado (%)	76,6	72,8	71,1	68,4	70,9	71,3	70,7	72,7	72,6
Consumo autorizado não faturado (%)	8,5	8,9	8,7	9,3	7,6	7,3	7,9	6,0	6,5
Água Não Faturada									
Perdas Aparentes (%)	2,0	3,3	5,0	6,4	5,5	5,6	5,6	5,7	5,7
Perdas Reais (%)	12,9	15,1	15,2	15,9	15,9	15,8	15,7	15,6	15,2
Total (Água Não faturada) (%)	23,4	27,2	28,9	31,6	29,1	29,1	28,7	29,3	27,4

Com a utilização dos dados anteriormente referidos é possível de se representar graficamente a quantidade de água que entra no sistema, bem como a quantidade de água que é faturada e não faturada. Assim, é possível fazer uma análise da evolução desta entidade gestora ao longo dos anos e em que sentido devem ser implementadas as medidas de forma a se tornarem o mais eficiente possível. (Figura 29)

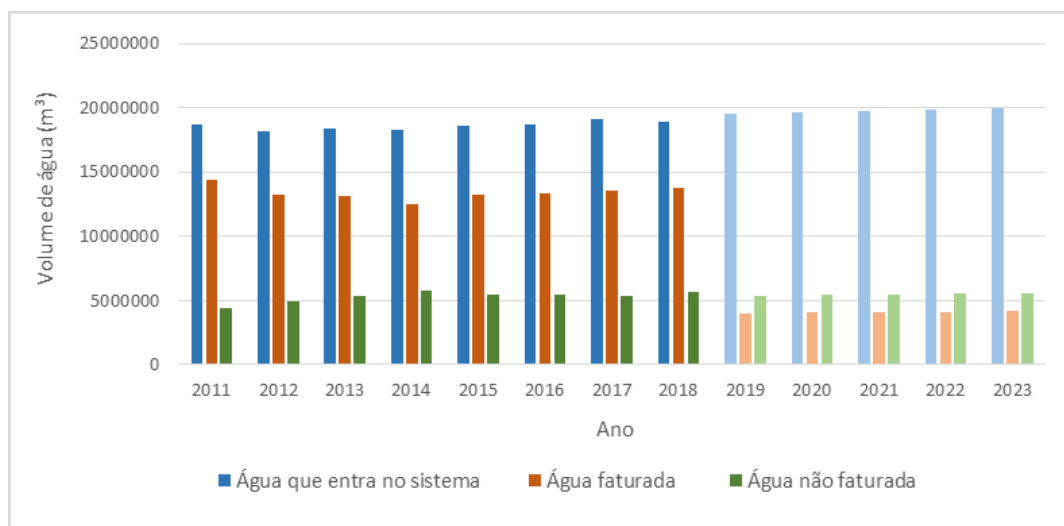


Figura 29: Evolução do volume de água que entrada no sistema, volume de água faturada e volume de água não faturada na empresa Águas de Gaia, EM, S.A.

Pela análise da Figura 29 é possível concluir que o volume de água que entrou no sistema foi superior no ano 2017 relativamente aos restantes anos. Quanto ao volume de água faturada, verifica-se uma diminuição desde o ano de 2011 até ao ano 2014 e, consequentemente, a percentagem de água não faturada foi superior nestes anos. Quanto aos valores das perdas, as perdas reais apresentam percentagens mais elevadas do que as perdas aparentes, e isto está relacionado com o facto de o volume de água perdida por roturas e fugas quer nas ligações, quer na rede ou mesmo nas adutoras ser maior e, por isso, a percentagem de perdas reais é superior às perdas aparentes.

Como foi referido inicialmente a água não faturada divide-se em perdas de água totais e consumo autorizado não faturado, sendo que as primeiras se ramificam em perdas aparentes e perdas reais. Como referido anteriormente, as perdas aparentes são as mais importantes a minimizar, uma vez que têm maior importância a nível financeiro para a EG. Por esse motivo, e de forma a aplicar algumas medidas de minimização deste tipo de perdas é apresentado na Figura 30, onde se encontram discriminadas as perdas aparentes em erros de medição e consumos não autorizados. Assim, é possível

observar-se como evolui este tipo de perdas e quais as medidas mais adequadas a serem aplicadas de forma a minimizar estes tipos de ilicitude.

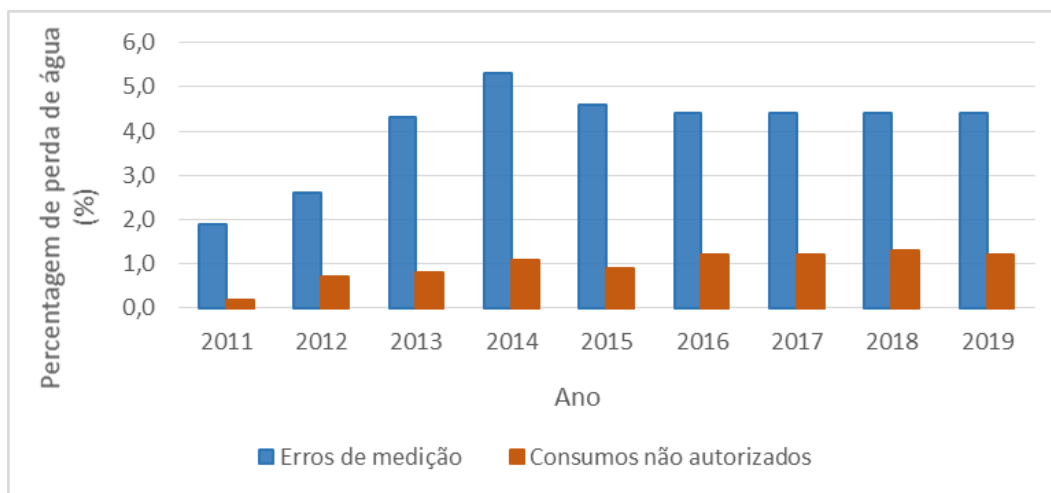


Figura 30: Evolução dos valores relativos aos erros de medição e aos consumos não autorizados na empresa Águas de Gaia, EM, S.A.

Pela análise da Figura 30, onde se encontram todos os tipos de perdas aparentes discriminadas, os erros de medição tiveram um aumento mais significativo a partir do ano de 2012 até 2014. Já a partir do ano de 2015 os valores diminuíram até estabilizar em 2018. Por sua vez, os consumos não autorizados tiveram um aumento mais considerável também a partir do ano 2012, tendo vindo a aumentar até ao ano de 2018, exceto no ano de 2015 em que houve um pequeno decréscimo deste valor. Conclui-se então que foi no ano 2012 que houve a revolução no aumento da percentagem de perdas aparentes, sendo que o maior pico foi no ano de 2014. Contudo, o ano de 2018 foi um bom ano para a empresa Águas de Gaia, uma vez que se comprou menos água à empresa Águas do Douro e Paiva relativamente ao ano de 2017, e conseguiu-se faturar mais do que nesse mesmo ano. Por essa razão, o balanço foi bastante positivo para a empresa.

4 Trabalho Desenvolvido

4.1 Ficheiro de *Excel*

O presente projeto iniciou-se com a recolha de uma amostra de dados relativos ao ano de 2019 junto da empresa Águas de Gaia, EM, S.A. Os dados recolhidos foram relativos a clientes domésticos ativos, clientes que rescindiram contrato, clientes com dívidas e clientes que realizaram contraordenações.

4.1.1 Variáveis

As variáveis que foram utilizadas para a realização deste projeto são: clientes ativos; clientes que rescindiram contrato; clientes com dívida; clientes que realizaram contraordenações, todas estas variáveis são referentes aos clientes domésticos do ano 2019.

De uma forma mais descritiva, com os dados dos clientes ativos foi possível avaliar o número de cliente, número da instalação e os consumos que cada cliente realizou ao longo do ano; com o ficheiro dos contratos rescindidos foi possível avaliar quantas rescisões cada cliente realizou ao longo do ano; aos dados referentes à soma do valor em dívida por cada cliente juntou-se o número de faturas em dívida, por último, recolheu-se os dados dos clientes que realizaram contraordenações, quer por adulteração dos contadores quer por ligações ilícitas, todas estas informações fornecidas pela empresa encontram-se apresentadas no Apêndice A.1, A.2, A.3 e A.4.

Depois de recolhidos e analisados todos os dados foi possível com o apoio de algumas ferramentas do *Excel* (*Tabelas Dinâmicas* e *PROCV*) agregar todos esses dados numa única tabela para que a informação ficasse mais organizada, a Tabela 2 representa uma parte da tabela geral.

Tabela 2: Tabela final com todas as variáveis (apenas para alguns casos).

Nº de Cliente	Tipo de Cliente	Nº Rescisões	Valor total em dívida (euros)	Nº de faturas em atraso	Consumo de água (m3/ano)
1 186 575	1	0	0	0	58
1 186 585	1	0	0	0	35
1 186 593	1	0	0	0	0
1 186 597	1	0	30	1	61

1 186 606	1	0	53	1	168
1 186 608	1	0	0	0	4
1 186 621	1	0	0	0	51
1 186 629	1	0	33	1	308
1 186 641	1	0	0	0	8
1 186 655	1	0	42	1	191
1 186 659	1	0	0	0	30
1 186 665	1	0	265	3	199
1 186 672	1	0	79	2	100
1 186 673	1	0	0	0	18
1 186 674	1	1	0	0	0
1 186 680	1	0	0	0	0
1 186 683	1	0	0	0	65
1 186 694	1	0	0	0	62
1 186 705	1	0	47	1	105
1 186 706	1	0	1	1	55
...					
1 190 624	3	1	0	0	70
1 252 908	3	0	163	2	223
1 258 795	3	0	11	1	42
1 276 681	3	0	23	1	54
1 290 024	3	0	36	1	110
1 297 081	3	0	9	2	3
1 307 176	3	1	0	0	53
1 312 427	3	1	0	0	8
1 340 156	3	1	0	0	84
1 340 714	3	1	0	0	57
1 379 544	3	1	0	0	52

Avaliaram-se dois tipos de clientes: os clientes ativos- categoria 1, e ilícitos conhecidos- categoria 3, isto é, clientes a quem foi detetada uma contraordenação e irão corresponder aos casos irregulares.

Uma forma de melhorar o ficheiro de *Excel* passa por adicionar mais variáveis ao mesmo. Uma variável que poderia ser adicionada é a estimativa do consumo de água associada a cada cliente. Assim, seria possível analisar algumas irregularidades pela relação desta informação com a quantidade de água contabilizada pela EG.

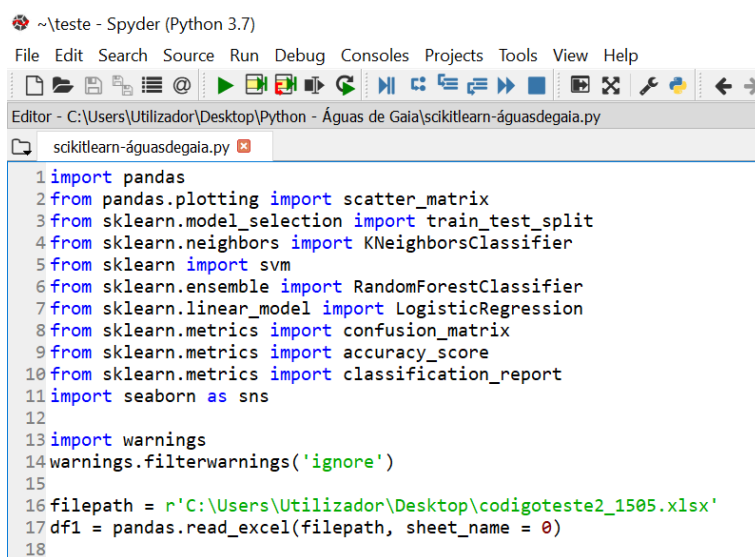
4.2 *Machine Learning*

Tal como supramencionado, um modelo ML baseia-se em encontrar padrões no conjunto de dados inicialmente fornecidos e, por esta razão, construiu-se uma tabela de *Excel* com os dados bem estruturados para que a avaliação dos padrões no conjunto de dados seja mais clara. Utilizou-se a aprendizagem supervisionada, uma vez que a partir dos dados fornecidos ao sistema é possível estabelecer um padrão com os dados de entrada e, assim, detetar novos casos ilícitos que até ao momento eram desconhecidos.

O código que foi elaborado para este projeto tem como principal finalidade encontrar novos casos ilícitos do consumo de água do sistema de abastecimento de água do município de Vila Nova de Gaia.

Através do ML pretendeu-se treinar o modelo de forma a transformar o *input* - dados fornecidos no início- em *output*- dados de saída. Depois desse passo é necessário que o modelo preveja novos dados, designando-se esses novos valores de *predicted*. (Pires, 2020)

Começou-se pela importação de todas as bibliotecas que são necessárias para construir o modelo e, de seguida, importou-se a folha de *Excel* contendo os dados fornecidos pela empresa (*input*) que foram previamente tratados e analisados. Estes passos podem ser visualizados na Figura 31. (Pires, 2020)



```
~\teste - Spyder (Python 3.7)
File Edit Search Source Run Debug Consoles Projects Tools View Help
Editor - C:\Users\Utilizador\Desktop\Python - Águas de Gaia\scikitlearn-águasdegaia.py
scikitlearn-águasdegaia.py
1 import pandas
2 from pandas.plotting import scatter_matrix
3 from sklearn.model_selection import train_test_split
4 from sklearn.neighbors import KNeighborsClassifier
5 from sklearn import svm
6 from sklearn.ensemble import RandomForestClassifier
7 from sklearn.linear_model import LogisticRegression
8 from sklearn.metrics import confusion_matrix
9 from sklearn.metrics import accuracy_score
10 from sklearn.metrics import classification_report
11 import seaborn as sns
12
13 import warnings
14 warnings.filterwarnings('ignore')
15
16 filepath = r'C:\Users\Utilizador\Desktop\codigoteste2_1505.xlsx'
17 df1 = pandas.read_excel(filepath, sheet_name = 0)
18
```

Figura 31: Código utilizado para importar todas as bibliotecas fundamentais para o projeto.

Seguidamente passou-se para a fase de treino do modelo e, para isso, os valores foram divididos em dados de teste 30% (que corresponde a 37505 casos) e dados de treino 70% (que corresponde a 87510 casos). Uma vez concluído este passo, obtiveram-se os valores de *output*, valores reais que vão ser classificados pelos algoritmos de ML (KNN, RFC, LRN e SVM), sendo que todo este processo está descrito na Figura 32. (Pires, 2020)

```
46
47 print("estou aqui 1")
48 KNN = KNeighborsClassifier(n_neighbors=1).fit(X_train, y_train)
49 round(KNN.score(X_test, y_test), 4)
50 print("estou aqui 2")
51
52 RFC = RandomForestClassifier(n_estimators=100).fit(X_train, y_train)
53 round(RFC.score(X_test, y_test), 4)
54 print("estou aqui 3")
55
56 LRN = LogisticRegression(solver='newton-cg', multi_class='ovr').fit(X_train, y_train)
57 round(LRN.score(X_test, y_test), 4)
58 print("estou aqui 4")
59
60 SVM = svm.SVC(decision_function_shape='ovr', kernel='rbf').fit(X_train, y_train)
61 round(SVM.score(X_test, y_test), 4)
62
```

Figura 32: Treino do modelo com os algoritmos de Machine Learning.

Depois de classificado o *output* por cada algoritmo analisou-se qual deles obteve melhores resultados que, para o caso em estudo, foi o KNN. Para avaliar esta classificação utiliza-se três variáveis que são essenciais para este processo: *confusion matrix*, *accuracy score* e *classification report*, sendo que na Figura 33 é possível observar-se a forma como foram utilizadas.

A *confusion matrix*- Figura 34- é essencial para avaliar a precisão de uma classificação, em que a primeira coluna corresponde ao número de casos conhecidos- *output*, e a segunda coluna ao número de casos previstos- *predicted*. Assim sendo, havendo um total de 124972 casos ativos, o modelo conseguiu detetar 20 casos como sendo ilícitos desconhecidos; e dos 43 casos conhecidos de ilícitos o modelo previu 28 casos desses já existentes. (Pires, 2020; *Scikit-Learn: Confusion Matrix*, 2020)

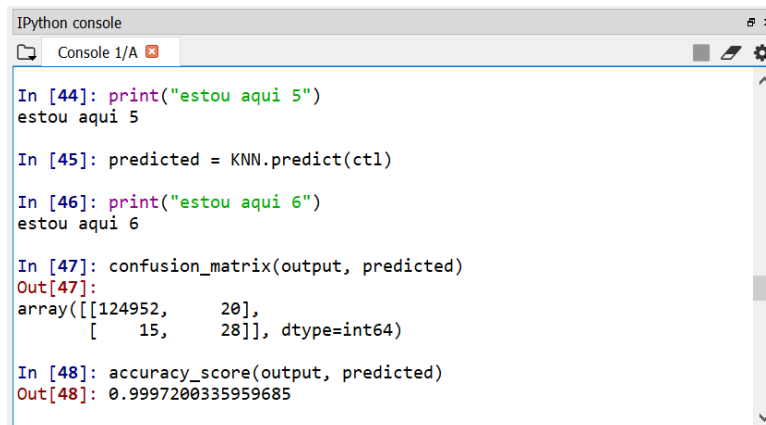
A variável *accuracy score*- Figura 34- é utilizada para avaliar a precisão, ou seja, na classificação dos vários *outputs* esta função irá analisar o conjunto de *outputs* previstos. (Pires, 2020; *Scikit-Learn: Accuracy Score*, 2020) Para os valores previstos pelo algoritmo KNN obteve-se um valor de aproximadamente 0,99972.

```

65 print("estou aqui 5")
66 predicted = KNN.predict(ct1)
67 print("estou aqui 6")
68 confusion_matrix(output, predicted)
69 accuracy_score(output, predicted)
70 print("estou aqui 7")
71 print(classification_report(output, predicted, target_names=['class 0', 'class 1']))
72

```

Figura 33: Código utilizado para calcular as variáveis confusion matrix, accuracy score e classification report.



```

IPython console
Console 1/A

In [44]: print("estou aqui 5")
estou aqui 5

In [45]: predicted = KNN.predict(ct1)

In [46]: print("estou aqui 6")
estou aqui 6

In [47]: confusion_matrix(output, predicted)
Out[47]:
array([[124952,  20],
       [  15,  28]], dtype=int64)

In [48]: accuracy_score(output, predicted)
Out[48]: 0.9997200335959685

```

Figura 34: Resultados obtidos com o código da confusion matrix e da accuracy score.

Uma vez quantificados os casos ilícitos, torna-se importante identificar os clientes associados, em que se teve por base o código representado na Figura 35.

```

74 df1sel = df1[df2]
75 df1sel.drop(['GAIA'], axis = 1, inplace = True)
76 df1sel.dropna(inplace=True)
77 print("estou aqui 8")
78 dfpred = df1sel.copy()
79
80 dfpred['KNN'] = KNN.predict(df1sel)
81 test_var = dfpred['KNN']
82 dfpred['RFC'] = RFC.predict(df1sel)
83 test_var = dfpred ['RFC']
84 dfpred['LRN'] = LRN.predict(df1sel)
85 test_var = dfpred ['LRN']
86 dfpred['SVM'] = SVM.predict(df1sel)
87 test_var = dfpred ['SVM']
88 print("estou aqui 9")
89

```

Figura 35: Código utilizado para com o auxílio dos algoritmos identificar os clientes associados.

Por último, extrai-se o ficheiro final para que posteriormente seja mais simples analisar todos os valores obtidos (Figura 36).

```
91 dfpred.KNN.sum()  
92 dfpred.RFC.sum()  
93 dfpred.LRN.sum()  
94 dfpred.SVM.sum()  
95 print("estou aqui 10")  
96 dfpred.to_csv('tabelacomresultadosfinaiss.txt')  
97 print("estou aqui 11")  
98
```

Figura 36: Exportação dos resultados.

5 Análise dos Dados

5.1 Dados de *Excel*

Os gráficos apresentados na Figura 37 foram elaborados pelo modelo de ML com base no ficheiro de *Excel* fornecido ao programa, onde os marcadores a azul representam os clientes ativos e os marcadores a laranja representam os ilícitos conhecidos.

Analisando a Figura 37 a) é possível aferir que os casos ilícitos apresentam consumos mais baixos quando comparados com os clientes ativos, uma vez que, tal como mencionado, não contabilizam o consumo total de água.

Por sua vez, o número de faturas em atraso e o valor em dívida encontram-se relacionados, uma vez que os casos ilícitos apresentam baixos valores em dívida e poucas faturas em atraso em relação aos clientes ativos- Figuras 37 b) e c). Esta relação está associada à não contabilização total do consumo de água por parte dos casos ilícitos, daí as dividas que apresentam serem de valores mais baixos. Por norma, espera-se que os clientes ativos apresentem os consumos mais elevados e os ilícitos os consumos mais baixos pela razão acima referida.

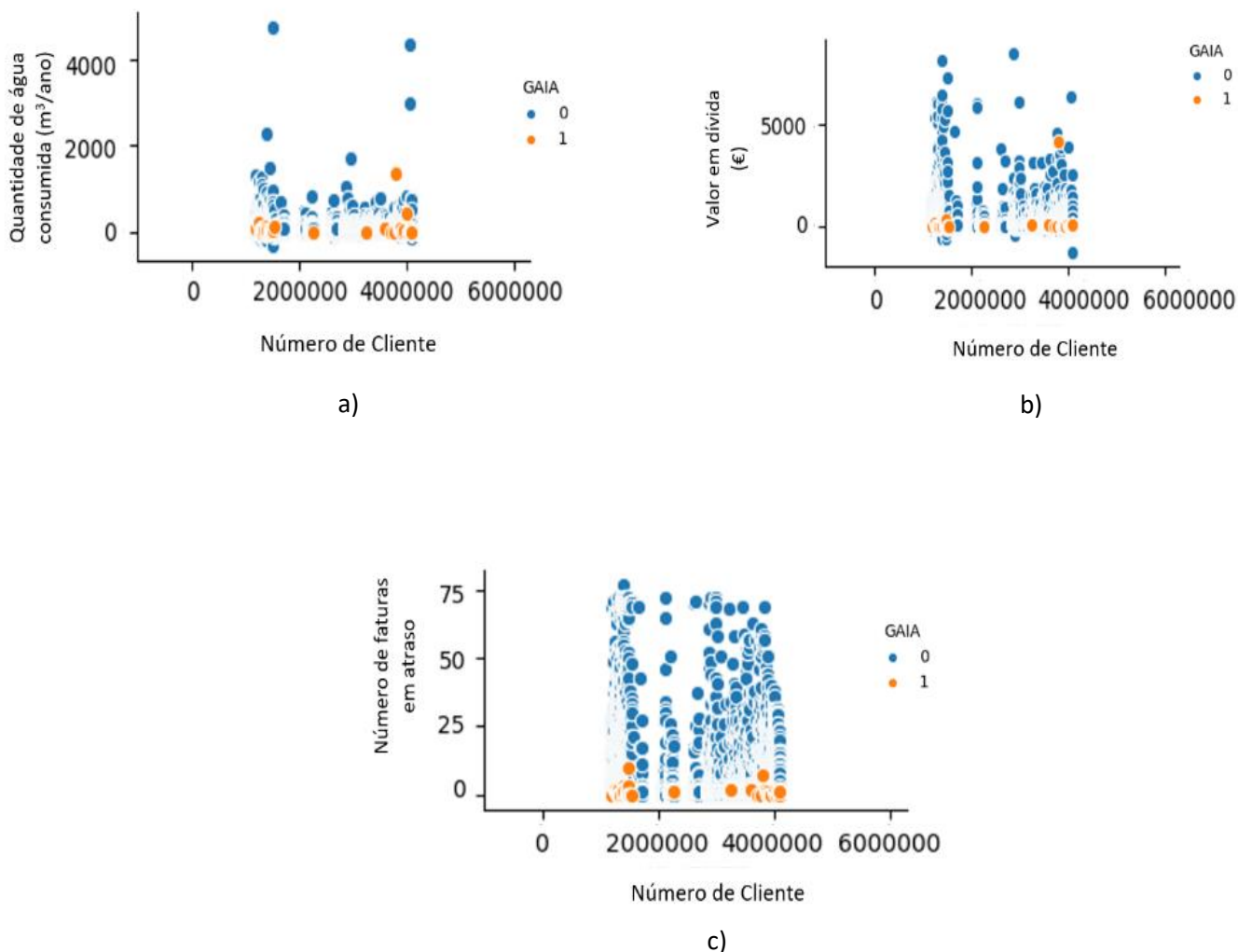


Figura 37: Representação gráfica dos casos ativos e casos ilícitos referentes à a) quantidade de água consumida b) valor em dívida e c) número de faturas em atraso.

O número de faturas em atraso encontra-se, ainda, relacionado com o consumo de água. Desta forma, pela análise da Figura 38 é possível concluir que os casos ilícitos apresentam menor número de faturas em atraso e uma menor quantidade de água consumida como foi supramencionado. Como os ilícitos não pagam o consumo total de água também não irão ter consumos elevados, valores em dívida exorbitantes, nem um elevado número de faturas em atraso.

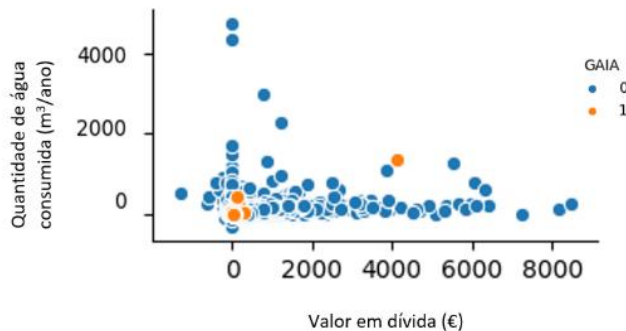
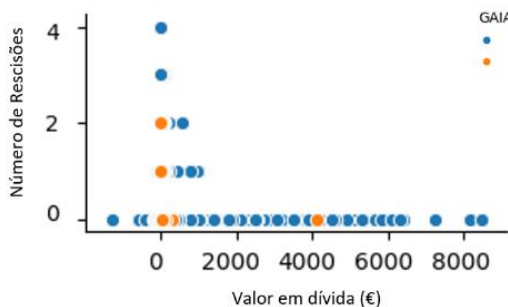
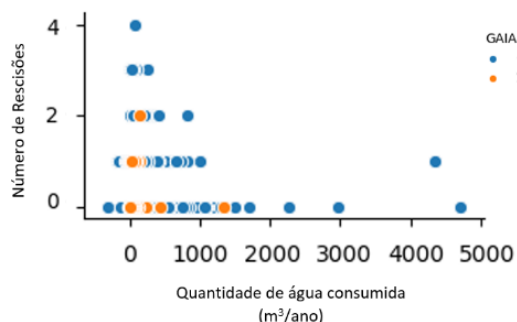


Figura 38: Relação entre o valor em dívida e quantidade de água consumida para os clientes ativos e para os casos ilícitos.

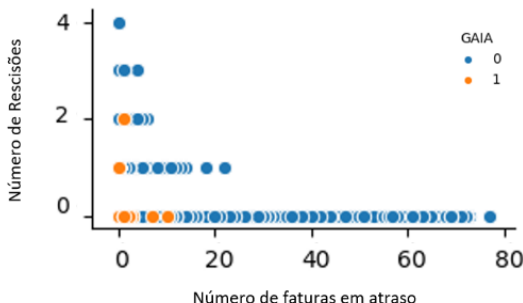
É possível observar pela análise dos gráficos supramencionados que baixos valores em dívida, baixas quantidades de consumo de água e poucas faturas em atraso correspondem, normalmente, a casos ilícitos. Contudo, o número de rescisões varia, para valores mais baixos do valor em dívida, da quantidade de água consumida e do número de faturas em atraso, observa-se uma maior variação no número de rescisões - Figura 39 a), b) e c).



a)



b)



c)

Figura 39: Relação entre o número de rescisões e o a) valor em dívida b) quantidade de água consumida e c) número de faturas em atraso.

5.2 Dados do *Machine Learning*

A Tabela 3 apresenta 48 casos ilícitos previstos pelo algoritmo KNN, bem como os diversos parâmetros associados a cada um deles. Nas Tabelas 4, 5 e 6 encontram-se apresentados todos os resultados obtidos pelos algoritmos RFC, LRN e SVM, respetivamente.

Tabela 3: Resultados obtidos pelo algoritmo KNN.

Algoritmo	Nº Cliente	Nº Total de Rescisões	Valor em Dívida (€)	Nº de Faturas Atrasadas	Quantidade de água consumida (m ³ /ano)
KNN	1290035	0	32.87	1	103
	1312434	0	0.0	0	10
	1340165	0	0.0	0	89
	1381077	0	13.17	2	1
	1381091	0	0.0	0	18
	1394523	0	0.0	0	132
	1394552	0	36.1	1	132
	1436787	0	0.0	0	3
	1436789	0	0.0	0	2
	1436809	0	0.0	0	2
	1452709	0	50.91	2	50
	1475140	0	0.0	0	89
	1475163	0	0.0	0	1
	1475165	0	0.0	0	4
	3579174	0	13.13	1	-6
	3868615	0	20.18	1	36
	3868649	0	0.0	0	30
	3868697	1	13.82	1	30
	3868719	0	22.28	1	28
	4079404	0	0.0	0	26
	1190624	1	0.0	0	70
	1252908	0	163.37	2	223
	1258795	0	11.17	1	42
	1276681	0	22.55	1	54
	1290024	0	35.85	1	110
	1312427	1	0.0	0	8
	1340156	1	0.0	0	84
	1340714	1	0.0	0	57
	1379544	1	0.0	0	52
	1381075	1	0.0	0	27
	1386786	0	123.55	2	68
	1394542	2	14.19	1	130
	1436794	1	0.0	0	0

1452695	0	39.86	1	30
1475117	0	0.0	0	18
1529495	1	0.0	0	136
2240900	0	25.5	1	0
3579163	0	60.21	2	75
3701863	1	0.0	0	11
3856503	0	16.66	2	71
3868642	0	12.69	1	26
3871287	1	0.0	0	88
3907298	1	0.0	0	14
3984348	0	113.78	1	437
4034801	1	0.0	0	21
4055582	0	17.87	1	38
4079406	1	0.0	0	27
4086546	0	58.24	1	1

Tabela 4: Resultados obtidos pelo algoritmo RFC.

Algoritmo	Nº Cliente	Nº Total de Rescisões	Valor em Dívida (€)	Nº de Faturas Atrasadas	Quantidade de água consumida (m ³ /ano)
RFC	1190624	1	0.0	0	70
	1252908	0	163.37	2	223
	1258795	0	11.17	1	42
	1276681	0	22.55	1	54
	1290024	0	35.85	1	110
	1312427	1	0.0	0	8
	1340156	1	0.0	0	84
	1340714	1	0.0	0	57
	1379544	1	0.0	0	52
	1381075	1	0.0	0	27
	1386786	0	123.55	2	68
	1394542	2	14.19	1	130
	1436794	1	0.0	0	0
	1452695	0	39.86	1	30
	1475117	0	0.0	0	18
	1529495	1	0.0	0	136
	2240900	0	25.5	1	0
	3579163	0	60.21	2	75
	3701863	1	0.0	0	11
	3856503	0	16.66	2	71
	3868642	0	12.69	1	26
	3871287	1	0.0	0	88
	3907298	1	0.0	0	14
	3984348	0	113.78	1	437

4034801	1	0.0	0	21
4055582	0	17.87	1	38
4079406	1	0.0	0	27
4086546	0	58.24	1	1

Tabela 5: Resultados obtidos pelo algoritmo LRN.

Algoritmo	Nº Cliente	Nº Total de Rescisões	Valor em Dívida (€)	Nº de Faturas Atrasadas	Quantidade de água consumida (m ³ /ano)
LRN	1316219	4	0.0	0	60

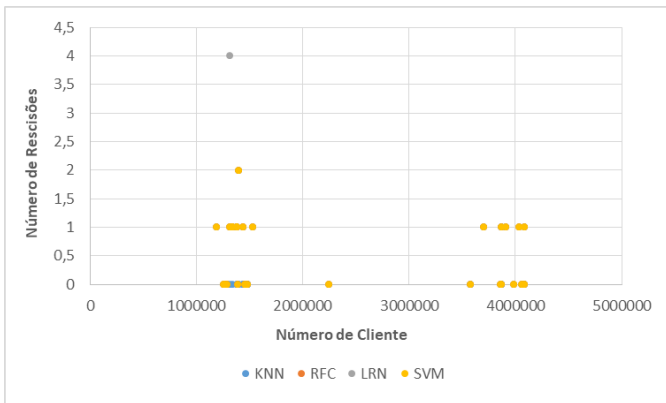
Tabela 6: Resultados obtidos pelo algoritmo SVM.

Algoritmo	Nº Cliente	Nº Total de Rescisões	Valor em Dívida (€)	Nº de Faturas Atrasadas	Quantidade de água consumida (m ³ /ano)
SVM	1190624	1	0.0	0	70
	1252908	0	163.37	2	223
	1258795	0	11.17	1	42
	1290024	0	35.85	1	110
	1312427	1	0.0	0	8
	1340156	1	0.0	0	84
	1340714	1	0.0	0	57
	1379544	1	0.0	0	52
	1381075	1	0.0	0	27
	1386786	0	123.55	2	68
	1394542	2	14.19	1	130
	1436794	1	0.0	0	0
	1452695	0	39.86	1	30
	1475117	0	0.0	0	18
	1529495	1	0.0	0	136
	2240900	0	25.5	1	0
	3579163	0	60.21	2	75
	3701863	1	0.0	0	11
	3856503	0	16.66	2	71
	3868642	0	12.69	1	26
3871287	1	0.0	0	88	
3907298	1	0.0	0	14	
3984348	0	113.78	1	437	
4034801	1	0.0	0	21	
4055582	0	17.87	1	38	

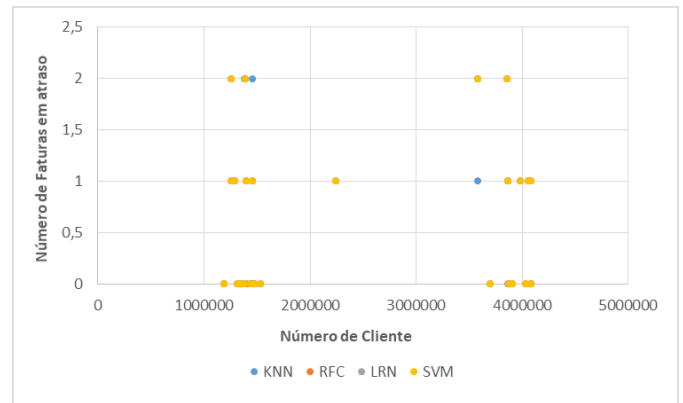
4079406	1	0.0	0	27
4086546	0	58.24	1	1

Na Figura 40 encontram-se as representações gráficas dos resultados previstos por cada algoritmo estudado. Na Figura 40 a) é possível analisar a variação do número de cliente em função do número de rescisões, chegando-se à conclusão que o mesmo número de cliente foi detetado por vários algoritmos.

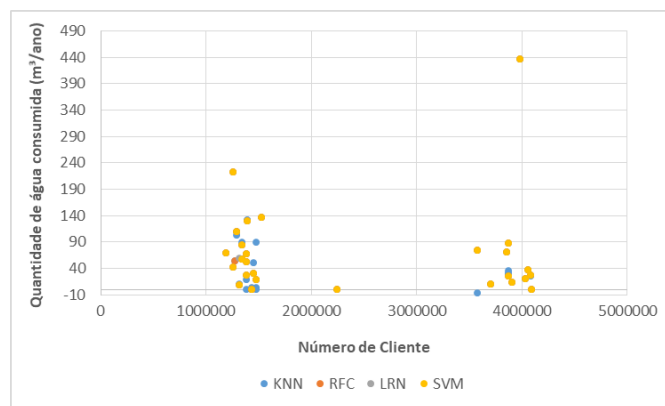
O mesmo é observado nas Figuras 40 b) e c) relativas, respectivamente, ao número de faturas em atraso e à quantidade de água consumida anualmente em que, também para estes dois parâmetros, diferentes algoritmos previram os mesmos clientes.



a)



b)



c)

Figura 40: Representação gráfica dos resultados obtidos pelos 4 algoritmos relativos ao a) número de rescisões b) número de faturas em atraso e c) quantidade de água consumida.

5.3 Análise Final

A seguir é apresentada a Tabela 7 que nos mostra os resultados obtidos pelos quatro algoritmos desenvolvidos ao longo deste trabalho.

Tabela 7: Resultados obtidos pelos 4 algoritmos.

Algoritmos	Precisão	Número de casos previstos	
KNN	0,99972	48	20 Ilícitos desconhecidos 28 Ilícitos conhecidos
RFC	0,99987	29	1 Ilícitos desconhecidos 28 Ilícitos conhecidos
LRN	0,99965	1	1 Ilícitos desconhecidos 0 Ilícitos conhecidos
SVM	0,99987	27	0 Ilícitos desconhecidos 27 Ilícitos conhecidos

Pela análise da Tabela 7 é possível observar que o algoritmo KNN previu 48 casos, onde 28 correspondem a ilícitos já conhecidos, enquanto que os 20 correspondem a ilícitos desconhecidos. Já os algoritmos RFC e SVM apresentam valores muito semelhantes: o primeiro apresenta 29 casos detetados, sendo que destes 28 são ilícitos conhecidos e 1 é ilícito desconhecido; já o SVM detetou 27 casos, sendo todos correspondentes a casos ilícitos conhecidos. Por último, o algoritmo LRN apresentou apenas 1 caso ilícito desconhecido.

De todos os algoritmos, o KNN para além de apresentar o maior número de casos previstos é o algoritmo que melhor se adequa ao caso de estudo, uma vez que foi desenvolvido para realizar uma análise mais pormenorizada quando os valores que se quer prever são desconhecidos ou de difícil determinação.

Depois de recolhidos os resultados finais, procedeu-se ao deslocamento ao local de forma a verificar caso a caso se estavam a cometer alguma ilicitude. Contudo, em nenhum destes casos foi detetado.

Os resultados obtidos por esta metodologia não foram totalmente os expectáveis uma vez que, apesar do algoritmo KNN ter sido o que previu mais casos, 48 casos no total, onde 28 corresponde a casos de ilícitos conhecidos e 20 a casos de ilícitos

desconhecidos, foi possível verificar-se presencialmente que cada um destes casos não apresentavam ilicitudes, apenas alguns dos casos apresentavam baixos consumos. O que não invalida o funcionamento do modelo criado, uma vez que um dos primeiros passos para perceber se um modelo está bem definido passa por detetar alguns dos ilícitos conhecidos fornecidos ao modelo o que foi verificado com este caso uma vez que este detetou 28 casos de ilícitos conhecidos o que significa que consegue prever os clientes que estejam a praticar ilicitudes.

Uma das razões que justificam estes resultados está relacionada com o facto de que os dados fornecidos correspondem a ilícitos aos quais foi instruído um processo jurídico. Ou seja, anualmente o número de ilícitos detetados é bastante superior ao que foi fornecido ao programa, porém a empresa apenas apresenta a informação discriminada para os ilícitos que sofreram um processo jurídico. Isto irá levar a que a amostra de casos ilícitos seja bastante pequena relativamente aos casos ativos que existe na empresa Águas de Gaia EM, S.A.

Outra das razões que está relacionada com estes resultados obtidos é baseada no funcionamento do algoritmo KNN que consiste na classificação de valores padrões, ou seja, irá funcionar com base nos valores mais próximos do valor a estimar. Assim, uma das principais desvantagens deste algoritmo pretende-se com encontrar os dados mais próximos para cada amostra. Por esta razão, se o número de valores fornecidos não for suficiente haverá uma maior dificuldade em detetar os valores mais próximos e, desta forma, os valores previstos.

Relativamente às precisões estas apresentam todos valores elevados, apenas variam na quarta casa decimal, isto acontece porque como a amostra é muito pequena (casos ilícitos 43 casos) relativamente aos casos ativos (124973 casos), os algoritmos irão prever mais os casos com valores mais baixos.

Uma forma de melhorar esta metodologia passa por adicionar mais variáveis ao ficheiro de *Excel*. Tal como foi supramencionado uma variável que poderia ser adicionada é a estimativa do consumo de água associada a cada cliente. Este parâmetro será bastante útil na deteção de algumas irregularidades, para isso basta relacionar esta informação com a quantidade de água contabilizada pela empresa.

Apesar dos 20 ilícitos desconhecidos detetados pelo algoritmo KNN não apresentarem ilicitudes, é necessário para futuros trabalhos nesta área ter atenção aos clientes que realizaram contraordenações e que são submetidos inicialmente ao modelo (input) para que este consiga detetar bons resultados.

6 Conclusões

O presente trabalho teve como objetivo a minimização das perdas aparentes dos sistemas de abastecimento de água utilizando para isso algoritmos de ML. A diminuição deste tipo de perdas irá levar conseqüentemente a uma diminuição da quantidade de água não faturada, parâmetro bastante importante para a avaliação da eficiência de uma EG.

Este projeto baseou-se na recolha e análise de dados fornecidos pela empresa referentes aos clientes ativos e aos casos ilícitos. Depois de analisados e organizados todos esses dados foi desenvolvido um modelo em linguagem *Python* que tem por objetivo prever casos ilícitos que até ao momento eram desconhecidos.

Este modelo baseou-se inicialmente na importação do ficheiro *Excel* dos dados fornecidos pela empresa (*input*), onde depois com o auxílio de 4 algoritmos diferentes (KNN, LRN, RFC, SVM) foi possível treinar estes dados de forma a converter o *input* em *output*. Por último, depois de o modelo ter sido treinado é possível prever os casos ilícitos desconhecidos, isto é, com os valores de *output* e com o auxílio dos 4 algoritmos é possível obter os valores *predicted* (casos ilícitos previstos).

É necessário ter em atenção os casos ilícitos que são fornecidos ao programa, pois, estes devem ser referentes a todos os casos em que foi detetada alguma ilicitude pela empresa e não apenas aos que foi instruído um processo jurídico, daí os resultados não terem sido como o esperado. Assim, a amostra estudada foi bastante pequena quando comparada com o número de casos conhecidos que são detetados anualmente.

Assim, é possível concluir que os objetivos estipulados para este projeto foram concretizados, uma vez que foi possível desenvolver uma metodologia capaz de minimizar as perdas aparentes com recurso a algoritmos de ML. Este projeto elaborado na empresa Águas de Gaia, EM,S.A. poderá ser bastante relevante para a empresa uma vez que levará a uma diminuição das perdas aparentes e, conseqüentemente, a uma diminuição do consumo de água não faturada levando a um aumento da eficiência desta EG.

7 Avaliação do Trabalho Realizado

7.1 Objetivos Realizados

O objetivo deste projeto consistiu no desenvolvimento de um modelo com recurso a algoritmos de ML de forma a minimizar as perdas aparentes dos sistemas de abastecimento de água. Atualmente, não existem muitos métodos para a minimização das perdas aparentes, desta forma, com a evolução da IA é possível desenvolver modelos de forma a diminuir este tipo de perdas de água sem ser necessários muitos recursos.

7.2 Recomendações Futuras

Depois de realizado este projeto existem algumas recomendações de melhoria que não foi possível abordar neste projeto, mas que deverão ser tidas em conta para futuros trabalhos nesta área.

Os valores de casos ilícitos utilizados dizem respeito a ilícitos ao qual foi instruído um processo jurídico, o que isto significa que a amostra de casos ilícitos que vai ser estudada é bastante pequena relativamente à realidade, o que irá comprometer a eficiência dos algoritmos usados.

A segunda recomendação está relacionada com o aumento da eficiência desta metodologia. Se todos os clientes fossem categorizados pela tipologia que apresentam e a essa categorização fosse associado um consumo expectável, seria mais uma variável bastante importante para a deteção de casos ilícitos e, desta forma, aumentar-se-ia a eficiência deste método.

7.3 Apreciação Final

Este projeto mostra ser bastante promissor pois será aplicada a IA de forma a diminuir as perdas aparentes dos sistemas de abastecimento de água, tornando a sua detecção mais prática e com menos utilização de recursos, trazendo bastantes benefícios para a EG.

Esta dissertação permitiu-me explorar novas áreas, desenvolver novos conhecimentos e, ainda, ter tido a oportunidade de desenvolver este projeto em ambiente empresarial permitiu que esta dissertação fosse um grande desafio a nível pessoal e profissional.

8 Referências

- AdP - Águas de Portugal. (2016). *Águas de Portugal*. www.adp.pt
- AdP - Águas de Portugal. (2019). *Relatório de Contas Intercalar do Ano 2019 das Águas de Portugal*.
- Águas de Gaia, EM, S. . (2015). *Relatório e Contas 2015*.
- Águas de Gaia, EM, S. . (2018). *Relatório e Contas 2018*. https://portal-chsj.min-saude.pt/uploads/document/file/678/R_C_2018.pdf. Accessed 2020, August
- Águas de Gaia, EM, S. . (2020a). *Distribuição de Água*. <http://www.aguasgaia.pt/empresas/agua>
- Águas de Gaia, EM, S. . (2020b). *Relatório de Gestão e Contas 2019*.
- Águas do Douro e Paiva, S. .-Á. de P. (2017). *ETA de Lever*. <https://www.addp.pt/dados.php?ref=etadelever>
- Alegre, H., Coelho, S., Almeida, C., & Vieira, P. (2005). Controlo de Perdas de Água em Sistemas Públicos de Adução e Distribuição. In *Dspace.Ist.Utl.Pt*. <https://doi.org/235614/05>
- Anaconda: *spyder*. (2020). <https://anaconda.org/anaconda/spyder>
- Anaconda Documentation: *spyder*. (2020). <https://docs.anaconda.com/anaconda/user-guide/tasks/integration/spyder/>
- Cardoso, A., Oliveira, F., Lacerda, F., & Martins, P. (2013). *Redução de Perdas Reais em Sistemas de Abastecimento de Água: Otimização das Pressões numa Grande Rede de Distribuição de Água - Aplicação a um Subsistema da Cidade do Porto*.
- Carvalho, H. (2014). *Redução de Perdas Reais de Água em Sistemas de Abastecimento de Água - Definição de critérios para delimitação de zonas de medição e controlo*.
- Decreto-Lei nº 23/95. (1995). Regulamento Geral de Sistemas Públicos e Prediais de Distribuição de Água e de Drenagem de Águas Residuais. *Diário Da República - 1 Série-B, N.º 194*.
- Entidade Reguladora dos Serviços de Águas e Resíduos. (2019a). *Relatório Anual dos Serviços de Águas e Resíduos em Portugal - Controlo da Qualidade da Água para Consumo Humano (Vol. 2)*.

- Entidade Reguladora dos Serviços de Águas e Resíduos. (2019b). *Relatório Anual dos Serviços de Águas e Resíduos em Portugal 2019* (Vol. 1).
- Época Negócios. (2018). *Falta de água afetará 5 biliões de pessoas até 2050, diz ONU*. <https://epocanegocios.globo.com/Mundo/noticia/2018/03/falta-de-agua-afetara-5-bilhoes-de-pessoas-ate-2050-diz-onu.html>
- ERSAR - Entidade Reguladora dos Serviços de Águas e Resíduos. (2015). *Abastecimento de Água*. <http://www.ersar.pt/pt/setor/caracterizacao/abastecimento-de-agua>
- Farley, M., Wyeth, G., Ghazali, Z. B. M., Istandar, A., & Sigh, S. (2008). The Manager's Non-Revenue Water Handbook. In *A Guide to Understanding Water Losses, Ranhill Utilities Bernhad and USAID, Malaysia*.
- Frade, J. V., Lacasta, N., Mendes, P., Cardoso, P., Trindade, I., Newton, F., Franco, P., Serra, A., Póvoa, C., Narciso, F., & Brôco, N. (2015). PENSAR 2020. In *Pensar 2020* (Vol. 2). <https://doi.org/10.1017/CBO9781107415324.004>
- Lédo, P. (1999). Combate ao Roubo de Água - Uma Experiência no Sistema de Guanambi. *20º Congresso Brasileiro de Engenharia Sanitária e Ambiental, Rio de Janeiro*.
- Malheiro, R., Poças-Martins, J., & Lacerda, F. (2011). Controlo de Perdas Aparentes em Sistemas de Abastecimento de Água com Utilização de Telecontagem. *6as Jornadas de Hidráulica, Recursos Hídricos e Ambiente*.
- Observador. (2019). *DECO denuncia desperdício de 90 milhões de euros por ano com perdas de água*. <https://observador.pt/2019/06/25/deco-denuncia-desperdicio-de-90-milhoes-de-euros-por-ano-com-perdas-de-agua/>
- Oliveira, F. (2013). *Redução das Perdas Reais em Sistemas de Abastecimento de Água - Otimização das pressões numa grande rede de distribuição de água (aplicação ao caso do Porto)*.
- Pedroso, V. M. R. (2007). *Manual dos Sistemas Prediais de Distribuição e Drenagem de Águas* (L. N. de E. C. (LNEC) (Ed.)).
- Pires, N. (2020). Introdução a Machine Learning com Python. *Formação OERN*.
- República, A. da. (2005). Lei n.º 58/2005. *Diário Da República – I Série - A N.º 249 - 29 de Dezembro de 2005*.
- scikit-learn: accuracy score.* (2020). https://scikit-learn.org/stable/modules/generated/sklearn.metrics.accuracy_score.html#sklearn.metrics.accuracy_score

scikit-learn: confusion matrix. (2020). https://scikit-learn.org/stable/modules/generated/sklearn.metrics.confusion_matrix.html

scikit-learn: KNeighbors Classifier. (2020). <https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html>

scikit-learn: Logistic Regression. (2020). https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

scikit-learn: Random Forest Classifier. (2020). <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>

scikit-learn: Support Vector Classification. (2020). <https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html>

Supervised learning: predicting an output variable from high-dimensional observations. (2020). https://scikit-learn.org/stable/tutorial/statistical_inference/supervised_learning.html

Swamynathan, M. (2019). *Mastering Machine Learning with Python in Six Steps.* In *Mastering Machine Learning with Python in Six Steps.* <https://doi.org/10.1007/978-1-4842-4947-5>

Teixeira, J., Oliveira, M. J., Sousa, S., & Poças Martins, J. (2014). *Redução das Perdas Aparentes em Sistemas de Abastecimento de Água Através da Detecção e Controlo de Consumos Ilícitos - Reduction of Apparent Losses in Water Supply Systems Trough the Detection and Control of Illicit Consumption. i.*

VANGENTE, Argentino; GASPAR, R. (2018). *A Gestão Integrada De Águas Subterrâneas e Águas Superficiais. 6º SILUSBA-Simpósio de Hidráulica e Recursos Hídricos Dos Países de Língua Oficial Portuguesa.* http://www.aprh.pt/6_silusba/vol3_APRH_LF_643_696.pdf

Apêndices

Apêndice A - Dados fornecidos pela empresa

A.1 Clientes domésticos ativos

A tabela A.1.1 é referente aos clientes domésticos ativos dos meses de Janeiro a Maio do ano 2019.

Tabela A.1. 1: Dados referentes aos clientes domésticos ativos de Janeiro a Maio.

Data Faturação	Ano e Mês Contabilístico	Código fatura	Número Conta	Número Cliente	Instalação	Tipo de Cliente Contrato	Código Tarifa	Conceito	Saldo Qt (m ³)
20190218	201902	36190201004434	1186575002	1186575	1	5 - DOMESTICOS	510	A1A - Consumo	5
20190322	201903	36190301003539	1186575002	1186575	1	5 - DOMESTICOS	510	A1A - Consumo	6
20190416	201904	36190400004471	1186575002	1186575	1	5 - DOMESTICOS	510	A1A - Consumo	64
20190416	201905	36190500001500	1186575002	1186575	1	5 - DOMESTICOS	510	A1A - Consumo	-64
20190416	201905	36190500001501	1186575002	1186575	1	5 - DOMESTICOS	510	A1A - Consumo	4
20190125	201901	36190501003202	1186575002	1186575	1	5 - DOMESTICOS	510	A1A - Consumo	9
20190523	201905	36190501007387	1186575002	1186575	1	5 - DOMESTICOS	510	A1A - Consumo	7
20190218	201902	36190201004435	3049456001	3049456	2	5 - DOMESTICOS	510	A1A - Consumo	6
20190322	201903	36190301003538	3049456001	3049456	2	5 - DOMESTICOS	510	A1A - Consumo	10
20190416	201904	36190401004594	3049456001	3049456	2	5 - DOMESTICOS	510	A1A - Consumo	7
20190125	201901	36190501003201	3049456001	3049456	2	5 - DOMESTICOS	510	A1A - Consumo	11
20190523	201905	36190501007386	3049456001	3049456	2	5 - DOMESTICOS	510	A1A - Consumo	9
20181218	201901	36181201011166	1286895001	1286895	19	5 - DOMESTICOS	510	A1A - Consumo	3
20190218	201902	36190201004400	1286895001	1286895	19	5 - DOMESTICOS	510	A1A - Consumo	3
20190322	201903	36190301003507	1286895001	1286895	19	5 - DOMESTICOS	510	A1A - Consumo	3
20190416	201904	36190401004559	1286895001	1286895	19	5 - DOMESTICOS	510	A1A - Consumo	3
20190125	201901	36190501003598	1286895001	1286895	19	5 - DOMESTICOS	510	A1A - Consumo	6
20190523	201905	36190501007380	1286895001	1286895	19	5 - DOMESTICOS	510	A1A - Consumo	4
...									
20190524	201905	36190517004535	4085662001	4085662	10014982	5 - DOMESTICOS	510	A1A - Consumo	0
20190430	201904	36190416017526	1470599003	1470599	10014984	5 - DOMESTICOS	519	A1A - Consumo	2
20190531	201905	36190512049240	1470599003	1470599	10014984	5 - DOMESTICOS	519	A1A - Consumo	3

20190524	201905	36190517004568	3052256004	3052256	10014986	5 - DOMESTICOS	510	A1A - Consumo	0
20190528	201905	36190552015530	3793518002	3793518	10014987	5 - DOMESTICOS	510	A1A - Consumo	6
20190528	201905	36190552015532	4085743001	4085743	10014991	5 - DOMESTICOS	510	A1A - Consumo	11

A Tabela A.1.2 é a continuação da tabela supramencionada mas agora referente aos meses de Junho a Setembro do ano 2019.

Tabela A.1. 2: Valores referentes aos clientes domésticos ativos de Junho a Setembro.

Data Faturação	Ano e Mês Contabilístico	Código fatura	Número Conta	Número Cliente	Instalação	Tipo de Cliente Contrato	Código Tarifa	Conceito	Saldo Qt (m ³)
20190619	201906	361906010 03857	118657500 2	1186575	1	5 - DOMESTIC OS	510	A1A - Consumo	5
20190719	201907	361907010 03536	118657500 2	1186575	1	5 - DOMESTIC OS	510	A1A - Consumo	9
20190820	201908	361908010 05749	118657500 2	1186575	1	5 - DOMESTIC OS	510	A1A - Consumo	5
20190923	201909	361909010 04917	118657500 2	1186575	1	5 - DOMESTIC OS	510	A1A - Consumo	8
20190619	201907	361906000 05339	304945600 1	3049456	2	5 - DOMESTIC OS	510	A1A - Consumo	6
20190719	201907	361907010 03535	304945600 1	3049456	2	5 - DOMESTIC OS	510	A1A - Consumo	10
20190820	201908	361908010 05750	304945600 1	3049456	2	5 - DOMESTIC OS	510	A1A - Consumo	9
20190923	201909	361909010 04916	304945600 1	3049456	2	5 - DOMESTIC OS	510	A1A - Consumo	7
20190619	201906	361906010 03833	128689500 1	1286895	19	5 - DOMESTIC OS	510	A1A - Consumo	3
20190719	201907	361907010 03576	128689500 1	1286895	19	5 - DOMESTIC OS	510	A1A - Consumo	3
20190820	201908	361908010 05713	128689500 1	1286895	19	5 - DOMESTIC OS	510	A1A - Consumo	4
20190923	201909	361909010 05411	128689500 1	1286895	19	5 - DOMESTIC OS	510	A1A - Consumo	7
...									
20190524	201906	361906000 04618	408566200 1	4085662	10014982	5 - DOMESTIC OS	510	A1A - Consumo	0
20190524	201906	361906000 04619	408566200 1	4085662	10014982	5 - DOMESTIC OS	510	A1A - Consumo	40
20190628	201907	361906190 00099	408566200 1	4085662	10014982	5 - DOMESTIC OS	510	A1A - Consumo	10

20190723	201907	361907520 12507	408566200 1	4085662	10014982	5 - DOMESTIC OS	510	A1A - Consumo	15
20190827	201908	361908520 15570	408566200 1	4085662	10014982	5 - DOMESTIC OS	510	A1A - Consumo	14
20190924	201909	361909520 14235	408566200 1	4085662	10014982	5 - DOMESTIC OS	510	A1A - Consumo	15
20190628	201907	361906160 17619	147059900 3	1470599	10014984	5 - DOMESTIC OS	519	A1A - Consumo	19
20190731	201907	361907090 26071	147059900 3	1470599	10014984	5 - DOMESTIC OS	519	A1A - Consumo	10
20190830	201909	361908210 00004	147059900 3	1470599	10014984	5 - DOMESTIC OS	519	A1A - Consumo	-5
20190930	201909	361909120 20389	147059900 3	1470599	10014984	5 - DOMESTIC OS	519	A1A - Consumo	7
20190930	201909	361909120 20390	408848400 1	4088484	10014985	5 - DOMESTIC OS	510	A1A - Consumo	0
20190619	201906	361906160 14728	305225600 4	3052256	10014986	5 - DOMESTIC OS	510	A1A - Consumo	20
20190724	201907	361907160 09260	305225600 4	3052256	10014986	5 - DOMESTIC OS	510	A1A - Consumo	10
20190823	201908	361908160 06513	305225600 4	3052256	10014986	5 - DOMESTIC OS	510	A1A - Consumo	4
20190923	201909	361909120 16157	305225600 4	3052256	10014986	5 - DOMESTIC OS	510	A1A - Consumo	8
20190625	201906	361906520 13513	379351800 2	3793518	10014987	5 - DOMESTIC OS	510	A1A - Consumo	4
20190726	201907	361907520 15751	379351800 2	3793518	10014987	5 - DOMESTIC OS	510	A1A - Consumo	10
20190827	201908	361908520 15257	379351800 2	3793518	10014987	5 - DOMESTIC OS	510	A1A - Consumo	6
20190926	201909	361909520 17600	379351800 2	3793518	10014987	5 - DOMESTIC OS	510	A1A - Consumo	7
20190726	201907	361907160 10426	127705200 3	1277052	10014988	5 - DOMESTIC OS	510	A1A - Consumo	0
20190827	201908	361908120 19627	127705200 3	1277052	10014988	5 - DOMESTIC OS	510	A1A - Consumo	0
20190926	201909	361909160 04503	127705200 3	1277052	10014988	5 - DOMESTIC OS	510	A1A - Consumo	5
20190827	201908	361908520 15253	408731200 1	4087312	10014989	5 - DOMESTIC OS	510	A1A - Consumo	0
20190926	201909	361909520 17604	408731200 1	4087312	10014989	5 - DOMESTIC OS	510	A1A - Consumo	14

20190625	201906	361906520 13510	408617400 1	4086174	10014990	5 - DOMESTIC OS	510	A1A - Consumo	3
20190726	201907	361907520 15753	408617400 1	4086174	10014990	5 - DOMESTIC OS	510	A1A - Consumo	5
20190827	201908	361908520 15254	408617400 1	4086174	10014990	5 - DOMESTIC OS	510	A1A - Consumo	4
20190926	201909	361909520 17602	408617400 1	4086174	10014990	5 - DOMESTIC OS	510	A1A - Consumo	1
20190625	201906	361906520 13511	408574300 1	4085743	10014991	5 - DOMESTIC OS	510	A1A - Consumo	7
20190726	201907	361907520 15754	408574300 1	4085743	10014991	5 - DOMESTIC OS	510	A1A - Consumo	2
20190827	201908	361908520 15255	408574300 1	4085743	10014991	5 - DOMESTIC OS	510	A1A - Consumo	6
20190926	201909	361909520 17603	408574300 1	4085743	10014991	5 - DOMESTIC OS	510	A1A - Consumo	4

A.2 Clientes que rescindiram contrato

A Tabela A.2.1 apresenta os clientes domésticos que rescindiram contrato em 2019.

Tabela A.2. 1: Clientes domésticos que rescindiram contrato

Nº Cliente Contrato	Nº Total de Baixas	Tipo Cliente	Instalação	Morada Instalação	Tarifa	Estado Contrato	Data Valor	Data de Criação	Data Proc. Rescisão	Motivo da Baixa
1400840	1	5 - DOMESTIC COS	99978	RUA PINHAIS BASTOS 151 3DT	510 DOMESTICO	21 - Rescisão Voluntária	200705 23	200705 23	201903 08	21 - Rescisão Voluntária
1467035	1	5 - DOMESTIC COS	99962	LRGO ALEMA 132 N/A	510 DOMESTICO	21 - Rescisão Voluntária	199206 08	199206 08	201910 01	21 - Rescisão Voluntária
1289944	1	5 - DOMESTIC COS	99897	RUA ALEXANDRE O NEIL 22 RCDT	510 DOMESTICO	21 - Rescisão Voluntária	199905 31	199905 31	201906 11	21 - Rescisão Voluntária
1488986	2	5 - DOMESTIC COS	99777	RUA ESCOLA LABORIM 324 3DF	510 DOMESTICO	21 - Rescisão Voluntária	201906 03	201905 31	201909 27	21 - Rescisão Voluntária
1466934	1	5 - DOMESTIC COS	9976	RUA FERNANDES DOS ANJOS 235 N/A	510 DOMESTICO	21 - Rescisão Voluntária	196005 05	196005 05	201909 05	21 - Rescisão Voluntária
1466881	1	5 - DOMESTIC COS	99642	RUA VISCONDE DAS DEVESAS 414 2	510 DOMESTICO	21 - Rescisão Voluntária	199202 18	199202 18	201903 06	21 - Rescisão Voluntária

1187734	1	5 - DOMESTICOS	9960	RUA RUI DE PINA 71 RC	510 DOMESTICO	21 - Rescisão Voluntária	201811 23	201811 23	201910 09	21 - Rescisão Voluntária
4083055	1	5 - DOMESTICOS	99579	PRCT HEROIS DO ULTRAMAR 144 2C3	510 DOMESTICO	21 - Rescisão Voluntária	201811 26	201811 13	201910 07	21 - Rescisão Voluntária
1385196	1	5 - DOMESTICOS	99527	RUA AQUILINO RIBEIRO 251 RCEQ	510 DOMESTICO	21 - Rescisão Voluntária	200512 19	200512 19	201907 15	21 - Rescisão Voluntária
4045733	1	5 - DOMESTICOS	99460	RUA 14 DE OUTUBRO 36 N/A	510 DOMESTICO	21 - Rescisão Voluntária	201707 31	201707 31	201902 08	21 - Rescisão Voluntária
1511436	1	5 - DOMESTICOS	99422	RUA AREIAS 266 N/A	510 DOMESTICO	21 - Rescisão Voluntária	201703 07	201703 01	201911 04	21 - Rescisão Voluntária
2113861	1	5 - DOMESTICOS	99395	RUA JOAQUIM LOPES PINTOR 118 3DT	510 DOMESTICO	21 - Rescisão Voluntária	201506 09	201506 05	201908 01	21 - Rescisão Voluntária
4019328	2	5 - DOMESTICOS	99300	RUA DR ANTONIO GRANJO 475 RCDF	510 DOMESTICO	21 - Rescisão Voluntária	201710 17	201710 17	201908 14	21 - Rescisão Voluntária
4042872	1	5 - DOMESTICOS	99286	PRCT TIMOR 48 3CT	510 DOMESTICO	21 - Rescisão Voluntária	201707 07	201707 06	201912 12	21 - Rescisão Voluntária
3001587	1	5 - DOMESTICOS	99274	RUA D PEDRO V 149 RCDF	510 DOMESTICO	21 - Rescisão Voluntária	201010 25	201010 25	201902 04	21 - Rescisão Voluntária
1466673	1	5 - DOMESTICOS	99201	RUA CASTANHEIRAS 35 N/A	510 DOMESTICO	21 - Rescisão Voluntária	199201 31	199201 31	201901 08	21 - Rescisão Voluntária
3915723	1	5 - DOMESTICOS	99197	RUA JOSE MARIANI 29 1	510 DOMESTICO	21 - Rescisão Voluntária	201703 30	201703 30	201903 18	21 - Rescisão Voluntária
1466653	1	5 - DOMESTICOS	99169	AVEN MAJOR BOTELHO MONIZ 306 1DF	510 DOMESTICO	21 - Rescisão Voluntária	199202 03	199202 03	201904 08	21 - Rescisão Voluntária
3549492	1	5 - DOMESTICOS	99116	RUA CRUZ DE CRISTO 37 3	510 DOMESTICO	21 - Rescisão Voluntária	201307 16	201307 16	201909 02	21 - Rescisão Voluntária
4083102	2	5 - DOMESTICOS	98958	RUA MARCIANO AZUAGA 92 1 E	510 DOMESTICO	21 - Rescisão Voluntária	201811 15	201811 15	201909 20	21 - Rescisão Voluntária
1525695	1	5 - DOMESTICOS	98955	RUA D ANTONIO FERREIRA GOMES 320 4DT	510 DOMESTICO	21 - Rescisão Voluntária	199201 21	199201 21	201910 01	21 - Rescisão Voluntária
4036857	1	5 - DOMESTICOS	98912	AVEN REPUBLICA 1273 5 DT	510 DOMESTICO	21 - Rescisão Voluntária	201907 01	201907 01	201911 27	21 - Rescisão Voluntária

3916024	1	5 - DOMESTI COS	98912	AVEN REPUBLICA 1273 5 DT	510 DOMESTICO	21 - Rescisão Voluntária	201602 25	201602 25	201907 01	21 - Rescisão Voluntária
3051731	1	5 - DOMESTI COS	98901	RUA GRACIOSA 148 N/A	510 DOMESTICO	21 - Rescisão Voluntária	201712 05	201712 05	201908 06	21 - Rescisão Voluntária
1466230	1	5 - DOMESTI COS	98900	TRAV EIRAS 24 2	510 DOMESTICO	21 - Rescisão Voluntária	199201 07	199201 07	201912 05	21 - Rescisão Voluntária
1466194	1	5 - DOMESTI COS	98897	AVEN ESCULTORES 42 9 C	510 DOMESTICO	21 - Rescisão Voluntária	199112 13	199112 13	201904 25	21 - Rescisão Voluntária
1312479	1	5 - DOMESTI COS	9877	RUA FERNAO MENDES PINTO 81 N/A	510 DOMESTICO	21 - Rescisão Voluntária	201507 20	201507 16	201905 03	21 - Rescisão Voluntária
1466482	1	5 - DOMESTI COS	98729	PRCT COOPERATIVA O TELEFONE 113 7EQ	510 DOMESTICO	21 - Rescisão Voluntária	199201 07	199201 07	201904 27	21 - Rescisão Voluntária
4051118	1	5 - DOMESTI COS	9872	RUA SOARES DOS REIS 113 1	510 DOMESTICO	21 - Rescisão Voluntária	201709 08	201709 08	201908 05	21 - Rescisão Voluntária
4083895	1	5 - DOMESTI COS	98680	RUA BELGICA 2240 2 CT	510 DOMESTICO	21 - Rescisão Voluntária	201812 28	201812 28	201911 12	21 - Rescisão Voluntária

A.3 Clientes com dívida

A Tabela A.3.1 representa os clientes domésticos com dívidas no ano 2019.

Tabela A.3. 1: Valores referentes aos clientes com dívida.

Data Contabilística	Data Faturação	Data Processamento	Nº Instalação	Nº Conta	Nº Fatura	Nº Cliente	Cod. Tipo Cliente	Tipo Cliente	Estado Contrato	Morada Cliente Fatura	Valor	Valor Total	Qualidade da Dívida
30/dez/19	27/dez/19	30/dez/19	100009	118659 7001	3,619 12E+13	118659 7	5	5 - DOMESTI COS	12 - Contrato em Vigor	RUA POUSA DA 104 99999	29,6 4	29,6 4	Corrente
10/dez/19	09/dez/19	23/dez/19	100013	118660 6001	3,619 12E+13	118660 6	5	5 - DOMESTI COS	12 - Contrato em Vigor	RUA ENTRE CAMPOS 252 99999	53,2	53,2	Corrente
13/dez/19	12/dez/19	27/dez/19	100020	118662 9001	3,619 12E+13	118662 9	5	5 - DOMESTI COS	12 - Contrato em Vigor	RUA QUINTA DO ESTEVE S 70 99999	32,8 8	32,8 8	Corrente
27/dez/19	23/dez/19	27/dez/19	100033	118665 5001	3,619 12E+13	118665 5	5	5 - DOMESTI COS	12 - Contrato em Vigor	RUA LUIS DE CAMOES 231 99999	41,9 5	41,9 5	Corrente

29/out/19	25/out/19	29/out/19	100052	1186665001	3,6191E+13	1186665	5	5 - DOMESTI COS	12 - Contrato em Vigor	RUA MIGUEL BOMBARDA 68 99999	64,23	64,23	Corrente
27/nov/19	27/nov/19	27/nov/19	100052	1186665001	3,61911E+13	1186665	5	5 - DOMESTI COS	12 - Contrato em Vigor	RUA MIGUEL BOMBARDA 68 99999	91,22	91,22	Corrente
27/dez/19	26/dez/19	27/dez/19	100052	1186665001	3,61912E+13	1186665	5	5 - DOMESTI COS	12 - Contrato em Vigor	RUA MIGUEL BOMBARDA 68 99999	109,16	109,16	Corrente
20/nov/19	18/nov/19	20/nov/19	100055	1186672001	3,61911E+13	1186672	5	5 - DOMESTI COS	12 - Contrato em Vigor	PRCT PADRE FLORO 43 5 C 99999	43,6	43,6	Corrente
11/dez/19	11/dez/19	11/dez/19	100055	1186672001	3,61912E+13	1186672	5	5 - DOMESTI COS	12 - Contrato em Vigor	PRCT PADRE FLORO 43 5 C 99999	35,82	35,82	Corrente
17/dez/19	17/dez/19	17/dez/19	100080	1186705001	3,61912E+13	1186705	5	5 - DOMESTI COS	12 - Contrato em Vigor	RUA ADRIANO C OLIVEIRA 177 RC 99999	46,52	46,52	Corrente
18/dez/19	17/dez/19	18/dez/19	100081	1186706001	3,61912E+13	1186706	5	5 - DOMESTI COS	12 - Contrato em Vigor	RUA ADRIANO C OLIVEIRA 177 1 99999	1,23	1,23	Corrente
11/dez/19	11/dez/19	23/dez/19	100082	1186710001	3,61912E+13	1186710	5	5 - DOMESTI COS	12 - Contrato em Vigor	AVEN REPUBLICA 855 5 FR 99999	12,01	12,01	Corrente
09/dez/19	09/dez/19	23/dez/19	100083	1186714001	3,61912E+13	1186714	5	5 - DOMESTI COS	12 - Contrato em Vigor	RUA BARAO DO CORVO 508 2 21 99999	24,34	24,34	Corrente
10/dez/19	09/dez/19	23/dez/19	100100	1186757001	3,61912E+13	1186757	5	5 - DOMESTI COS	12 - Contrato em Vigor	RUA CANELAS DE BAIXO 142 99999	25,83	25,83	Corrente
07/out/19	07/out/19	07/out/19	100125	1186812001	3,6191E+13	1186812	5	5 - DOMESTI COS	12 - Contrato em Vigor	RUA PEDREIRA 9 RC 99999	60,34	60,34	Corrente
08/nov/19	08/nov/19	08/nov/19	100125	1186812001	3,61911E+13	1186812	5	5 - DOMESTI COS	12 - Contrato em Vigor	RUA PEDREIRA 9 RC 99999	9,11	9,11	Corrente

04/dez/19	04/dez/19	04/dez/19	100125	1186812001	3,619 12E+1 3	1186812	5	5 - DOMESTI COS	12 - Contrato em Vigor	RUA PEDREIRA 9 RC 99999	39,75	39,75	Corrente
06/dez/19	05/dez/19	06/dez/19	100147	1186845001	3,619 12E+1 3	1186845	5	5 - DOMESTI COS	12 - Contrato em Vigor	RUA QUELHA 93 99999	29,06	29,06	Corrente
13/dez/19	13/dez/19	13/dez/19	100150	1186851001	3,619 12E+1 3	1186851	5	5 - DOMESTI COS	12 - Contrato em Vigor	TRAV MONTE 83 99999	26,95	26,95	Corrente
16/dez/19	16/dez/19	16/dez/19	100153	1186857001	3,619 12E+1 3	1186857	5	5 - DOMESTI COS	12 - Contrato em Vigor	BAIR CAVAC O II S/N B1 RC DT 99999	29,36	29,36	Corrente
13/dez/19	13/dez/19	13/dez/19	100129	1186865001	3,619 12E+1 3	1186865	5	5 - DOMESTI COS	12 - Contrato em Vigor	TRAV MONTE 67 99999	36,08	36,08	Corrente
27/dez/19	23/dez/19	27/dez/19	100164	1186882001	3,619 12E+1 3	1186882	5	5 - DOMESTI COS	12 - Contrato em Vigor	PRCT HEROIS DO ULTRAMAR 98 2EQ	19,88	19,88	Corrente
16/dez/19	16/dez/19	16/dez/19	100176	1186899001	3,619 12E+1 3	1186899	5	5 - DOMESTI COS	12 - Contrato em Vigor	BAIR CAVAC O II S/N 99999 1 DT	17,87	17,87	Corrente

A.4 Casos Ilícitos por Contraordenação da Água

A Tabela A.4.1 apresenta os casos ilícitos de clientes domésticos aos quais foi instruído um processo jurídico no ano 2019 por terem cometido contraordenações ao nível da água.

Tabela A.4. 1: Casos ilícitos por contraordenação ao nível da água.

Participação	Nº Instalação	Nº Cliente
20/05/2019	115751	3907298
15/07/2019	20926	3579163
28/10/2019	238351	1340156

11/12/2019

158627

1529495

A.5 Casos Ilícitos por Contraordenação dos Contadores

A Tabela A.5.1 representa os casos ilícitos de clientes domésticos aos quais foi instruído um processo jurídico no ano 2019 por terem praticado contraordenações ao nível do contador.

Tabela A.5. 1: Casos ilícitos por contraordenação dos contadores.

Data Participação	Nº Cliente
24/01/2019	3984348
16/01/2019	1307176
16/01/2019	1340714
04/02/2019	1276681
05/02/2019	1297081
22/02/2019	4034801
25/02/2019	1258795
25/02/2019	4081816
25/02/2019	1290024
27/02/2019	1397799
11/03/2019	4055582
14/03/2019	3701863
25/03/2019	3868642
18/04/2019	4072066
03/05/2019	1190624
22/05/2019	3229408
30/05/2019	4086546
05/06/2019	1486128

12/06/2019	1386786
16/07/2019	1381075
23/07/2019	1312427
25/07/2019	1436794
10/08/2019	1394542
06/09/2019	3792599
12/09/2019	1495244
12/09/2019	1379544
13/09/2019	3871287
26/09/2019	3756407
26/09/2019	1458848
11/10/2019	1463972
16/10/2019	3940782
24/10/2019	4053902
24/10/2019	1475117
24/10/2019	2240900
28/10/2019	1452695
28/10/2019	1252908
11/11/2019	1379804
21/11/2019	4079406
25/11/2019	3856503

Apêndice B - Gráficos do *Machine Learning*

As Figuras B.1, B.2, B.3, B.4, B.5 referem-se à representação gráfica dos dados fornecidos ao ML dos parâmetros estudados, para os clientes ativos (pontos azuis) e para os casos ilícitos (pontos laranja). Estes gráficos relacionam-se entre si pelos parâmetros estudados.

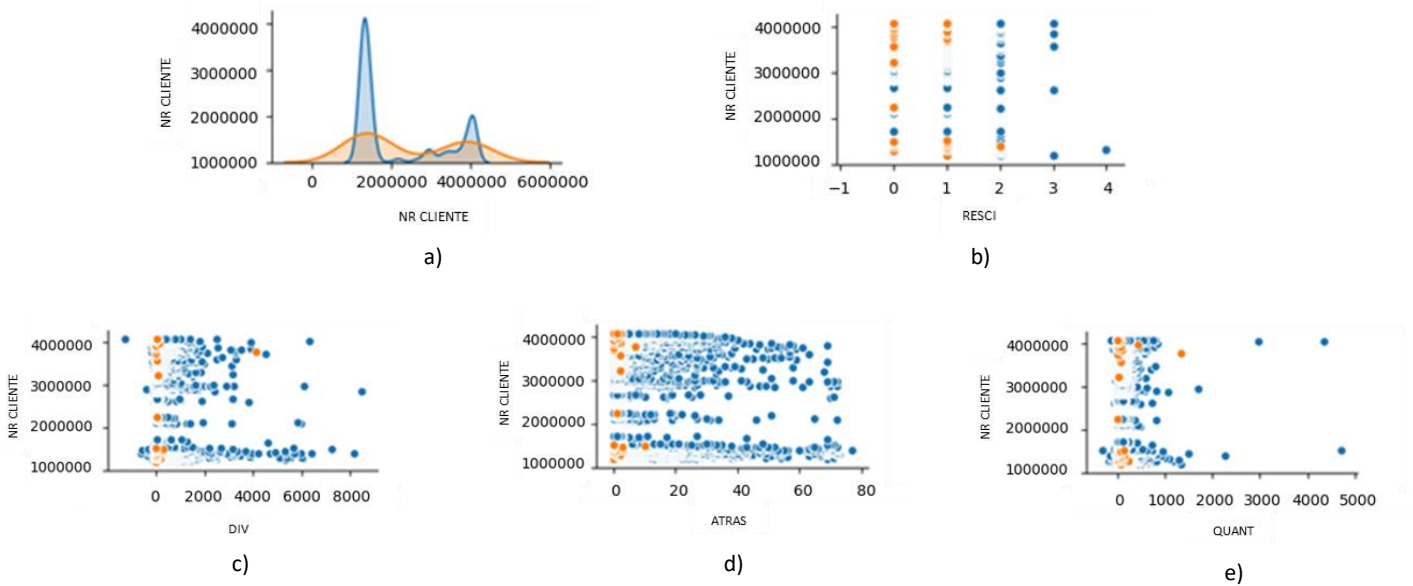


Figura B. 1: Representação gráfica do número de cliente em função a) número de cliente, b) número de rescisões, c) valor em dívida, d) número de faturas em atraso, e) quantidade de água consumida.

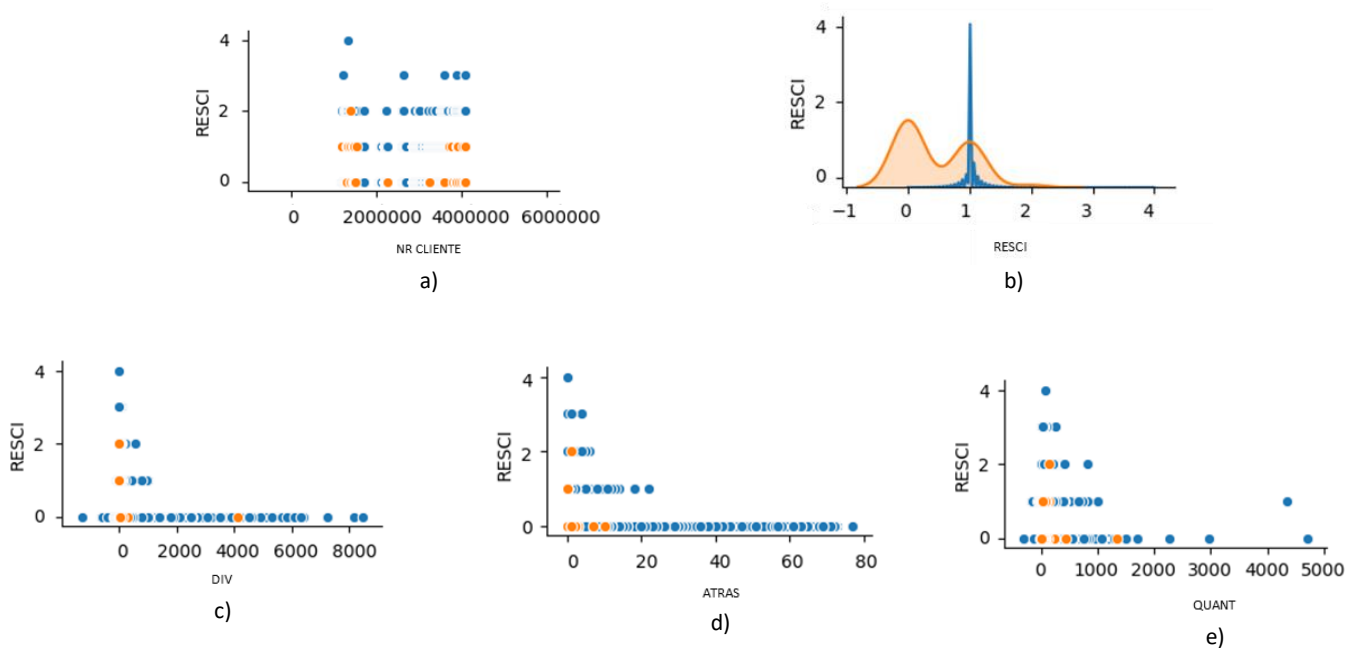


Figura B. 2: Representação gráfica do número de rescisões em função a) número de cliente, b) número de rescisões, c) valor em dívida, d) número de faturas em atraso, e) quantidade de água consumida.

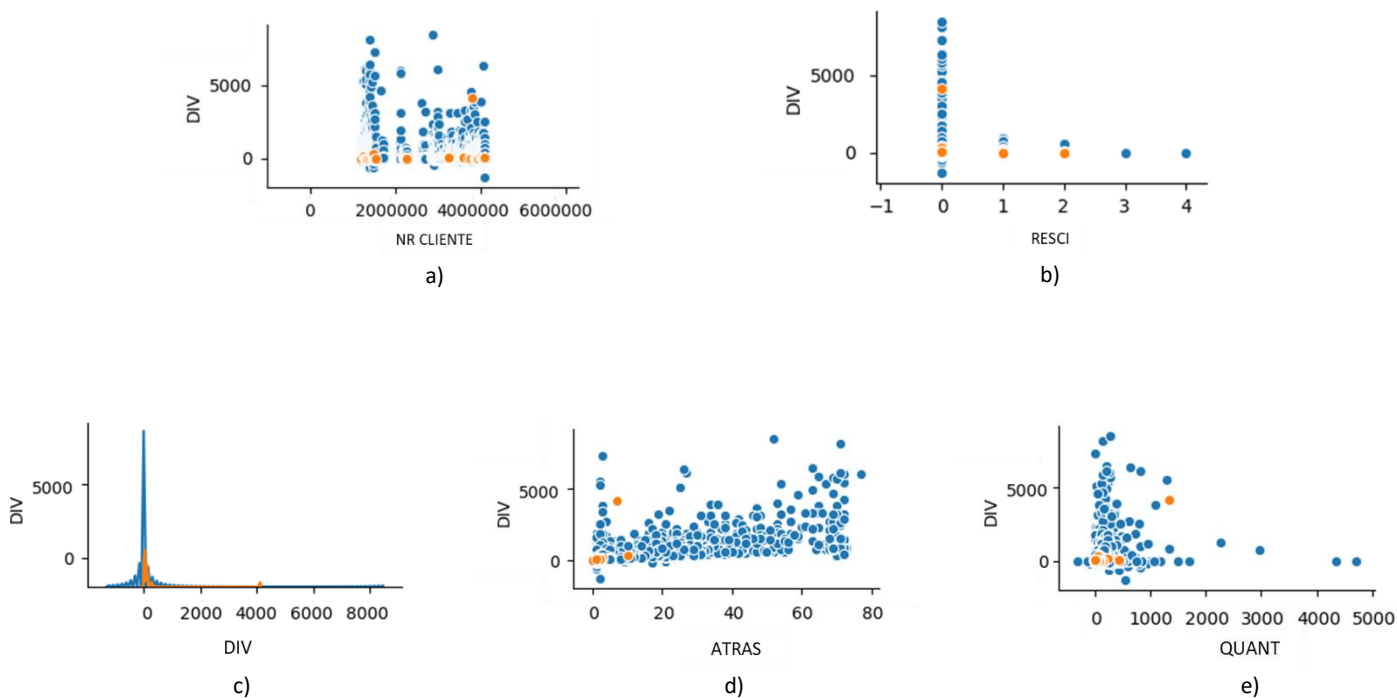


Figura B. 3: Representação gráfica do valor em dívida em função a) número de cliente, b) número de rescisões, c) valor em dívida, d) número de faturas em atraso, e) quantidade de água consumida.

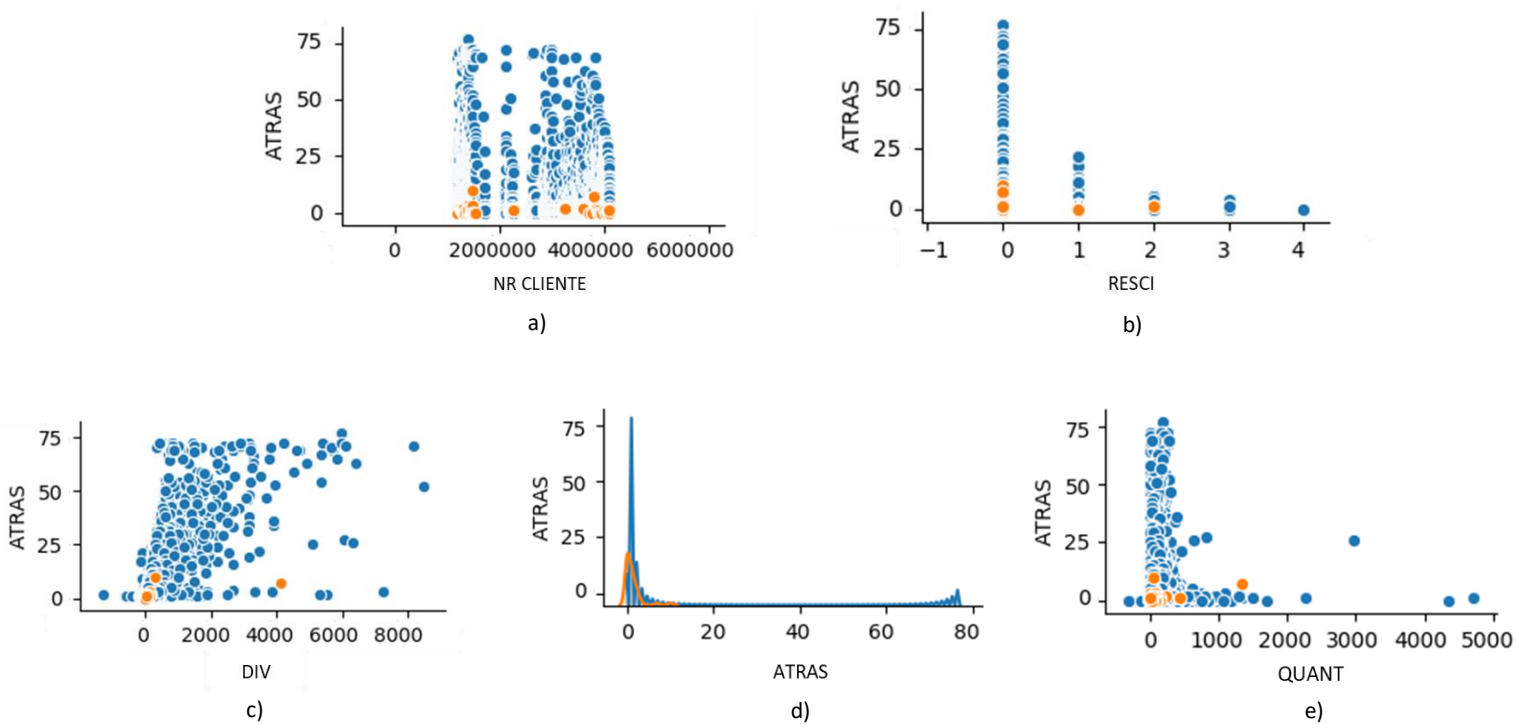


Figura B. 4: Representação gráfica do número de faturas em atraso em função a) número de cliente, b) número de rescisões, c) valor em dívida, d) número de faturas em atraso, e) quantidade de água consumida.

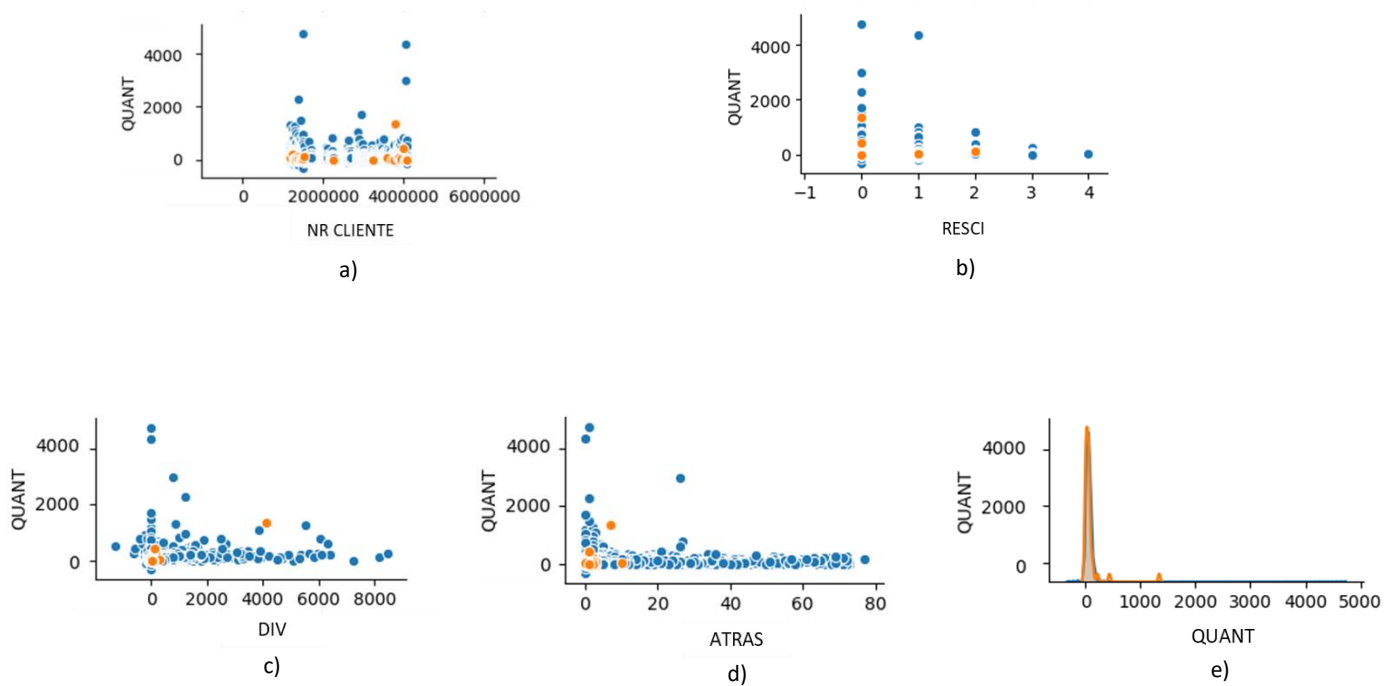


Figura B. 5: Representação gráfica da quantidade de água consumida em função a) número de cliente, b) número de rescisões, c) valor em dívida, d) número de faturas em atraso, e) quantidade de água consumida.