

GESTÃO DE SOLICITAÇÕES DE CLIENTES

Uma Abordagem com Recurso à Inteligência
Artificial

LEANDRA ELISA DA ROCHA

Dissertação submetida para satisfação parcial dos requisitos do grau de
MESTRE EM ENGENHARIA DO AMBIENTE

Orientador: Professor Doutor Joaquim Manuel Veloso Poças Martins
(Professor Associado com Agregação do Departamento de Engenharia Civil da Faculdade de
Engenharia da Universidade do Porto)

Coorientador: Eng^a Maria João Leite de Castro Marques de Oliveira
(Coordenadora da Unidade Orgânica de Gestão de Clientes da Águas
do Porto, EM)

SETEMBRO DE 2019

À minha Mãe

Tudo começa com o cliente
Luís XIV de França, o “Rei Sol”

AGRADECIMENTOS

Uma dissertação final de mestrado traduz-se numa tarefa árdua apenas concretizada quando se busca pelo saber que, de uma maneira ou outra, se transmite de pessoas para pessoas. O seu contributo foi decisivo em todas as fases do trabalho, tornando-se inevitável expressar a minha gratidão, em especial:

Ao meu orientador académico, o Professor Doutor Joaquim Poças Martins, por me conceder este primeiro contacto com o mundo empresarial, no qual pude aprender com os vários colaboradores. Retribuo com agradecimento a sua orientação, a sua sabedoria contagiante, a constante disponibilidade para elucidar todo o tipo de questões e por despertar sua visão vanguardista nas gerações mais novas.

À minha orientadora na empresa, a Engenheira Maria João Oliveira, pela amabilidade com que me acolheu, por me guiar nesta etapa e facultar o acesso ilimitado a fontes essenciais à elaboração da tese. Sem dúvida a pessoa mais inspiradora para mim em todos os momentos.

À Doutora Susana Sousa, que vestiu a pele de uma segunda mentora, pela receção calorosa, por exigir sempre o melhor acima de tudo, perseverança, jocosidade e estimulação do pensamento crítico.

Ao Gil Rocha e Luís Barbosa, do Laboratório de Inteligência Artificial e Ciência de Computadores (DEI-FEUP), pela ajuda e disponibilidade na compreensão e validação do algoritmo construído em *Machine Learning*, utilizando a linguagem de programação *Python*.

A todos os funcionários da UO de Gestão da Dívida, Maria João Freitas, Dr. Énio Guimarães, Carla Mourão, Magda Chichorro, às equipas de cortes, e particularmente Manuel Ribeiro, Rui Martins, Alexandre Neves e Beatriz Pimenta, por me proporcionarem uma melhor integração através das muitas conversas decorridas neste período, pela preocupação, júbilo e coadjuvação quando mais precisei. A par, dirijo igualmente a mensagem à Paula Malheiro do Planeamento e Controlo.

A todos os colaboradores da UO de Gestão de Clientes, que me apresentaram ao serviço de atendimento e a todos fatores de índole contratual, faturação e leitura fundamentais a esta dissertação, a contar da equipa que agrega o atendimento ao público ao *Backoffice*. Quero gratular principalmente, ao Pedro Moura e ao Dr. Manuel Monteiro pelo obséquio de satisfazerem a minha curiosidade.

A todos os elementos da UO de Informática, com principal enfoque para o João Lopes, João Vaz, Engenheiro Filipe Costa, Valdemar Freitas, Luís Fernandes e Daniel Madureira, pelo suporte frequente relativo a *hardware/software* basilar à evolução do projeto e ambiente descontraído. Agradeço ainda à Liliana Castro da Inovação pelo apoio e dedicação na componente de SIG.

À Madalena Fardilha e Cláudia Severino, pela simpatia e disponibilidade no esclarecimento de eventuais assuntos relacionados com o registo de e-mails e gestão de reclamações, respetivamente.

Ao grupo de almoço multidisciplinar, Nuno Neves, Marisa Fernandes, Moisés Neves, Joana Araújo e aos demais integrantes por demonstrarem boa disposição, sugestões e partilha de experiências.

Aos meus amigos e colegas que caminharam comigo ao longo de todo o percurso académico, por todos os momentos memoráveis e paciência astronómica. À Cristiana Ribeiro, por acompanhar de perto esta jornada, e à Sara Cunha, pela motivação sobretudo na empresa.

Aos meus pais, irmão, avós e restante família, figuras onnipresentes apesar da distância, a quem devo tudo por me catapultarem a procura por um futuro próspero e promissor durante toda a vida, ternura, ânimo, confiança e estabilidade. Ficarei eternamente grata pela esperança que depositaram em mim.

Ao Hugo, o meu braço direito carismático, por me escutar carinhosamente desde os primórdios da época universitária, pelo conforto, proatividade, sentido de humor incomparável e incessantes palavras encorajadoras que me mantiveram erguida, mesmo quando duvidei das minhas capacidades.

RESUMO

Esta dissertação apresenta como finalidade a gestão de solicitações de clientes, recorrendo a uma linguagem de programação no domínio da Inteligência Artificial aplicada a casos reais da cidade do Porto.

O estudo teve por base os pedidos recebidos por via eletrónica em duas contas pertencentes à empresa, pelo que se analisou em primeiro lugar o trajeto das emails, desde o envio por parte do cliente até à receção por um colaborador, de modo a fazer um enquadramento da situação atual. Numa segunda abordagem, definiu-se uma metodologia de classificação com o intuito de associar os inúmeros pedidos analisados às categorias correspondentes ao longo de um determinado período de tempo. Com base neste procedimento, efetuou-se inicialmente a georreferenciação das interações dos consumidores por correio eletrónico distribuídos pelas diversas categorias, a partir do qual foi possível constatar a predominância em torno das tipologias de Contratação e Faturação e Leitura.

O universo de classes foi adaptado à construção do algoritmo em *Python*, de modo a otimizar os resultados, e procedeu-se à conceção de uma base de dados com vários exemplos de solicitações recolhidas durante a categorização, sendo depois adicionada ao código a desenvolver em que cada categoria teve um único *template* genérico. Após a importação do *dataset* e realizada uma primeira análise estatística à amostragem, transformou-se o texto para assegurar que não fossem introduzidas quaisquer incoerências no modelo e testou-se os seguintes classificadores no ambiente de *Machine Learning*, mais concretamente em Aprendizagem Supervisionada (*Supervised Learning*): *Linear SVC*, *Bagging*, *RandomForest*, *Multi-layer Perceptron*, *Stochastic Gradient Descent*, *DecisionTree*, *ExtraTree*, *KNeighbors*, *BernoulliNB*, *Gaussian NB*, *MultinomialNB* e *ComplementNB*. Deste estudo, concluiu-se que o melhor classificador foi o *Linear SVC* que demonstrou uma exatidão de treino referente a 100% e uma exatidão de teste com 74%. Através da representação da matriz de confusão, verificou-se a atribuição incorreta de exemplos servidos como suporte. Na etapa final desta técnica, configurou-se a resposta automática e o envio foi testado utilizando um endereço institucional por meio da linha de comandos, despoletando assim a instrução com sucesso.

De seguida, enveredou-se pelo mesmo código, mas com alterações nas categorias que passaram a designar-se em função das várias respostas tipo produzidas. Testou-se de novo todos os classificadores usados no tópico anterior, onde o *Linear SVC* voltou a estar em destaque com um valor medíocre de 53% na exatidão de teste, contudo os resultados obtidos nesta fase revelaram-se inferiores comparado com o método transato.

Numa última fase, projetou-se um código para geração de texto automático utilizando redes neuronais no âmbito da aprendizagem profunda (*Deep Learning*). Optou-se pelas bases de dados onde constam as sete categorias, uma vez que se ajustaria melhor à situação, e selecionou-se uma rede neuronal com duas camadas em que cada camada subsistem 256 neurónios e um comprimento máximo de 100 caracteres que se treinou ao longo de 50 épocas. Apesar de não reconhecer os caracteres findo a sequência, o resultado foi favorável para os parâmetros fixados.

PALAVRAS-CHAVE: categorias, georreferenciação, *Machine Learning*, resposta automática, gestão de solicitações de clientes.

ABSTRACT

This dissertation is about the management of customer requests, using real cases in Porto city with the aid of a programming language in the field of Artificial Intelligence.

The study was based on the requests received electronically in the two accounts belonging to the company. First, the path of the emails sent from the customers that are received by a collaborator was analysed in order to determine the framework of the current situation. In the second approach, a classification methodology was defined to associate the numerous requests comparing with the corresponding categories over a given period of time. Based on this procedure, it was initially geo-referenced the interactions of consumers via e-mail distributed among the various categories, from which it was possible to verify the predominance around the types such as **hiring, billing and reading of water meter**.

The various classes were adapted to the Python algorithm construction in order to optimize the results. A database with numerous examples of requests collected during the categorization was designed, and then added to the code to be developed. Each category had a single generic template. After importing the dataset and performing the first statistical analysis on the sample, the text was transformed to ensure that no inconsistencies were introduced into the model and the following classifiers were tested in the Machine Learning environment, namely in Supervised Learning: Linear SVC, Bagging, RandomForest, Multi-layer Perceptron, Stochastic Gradient Descent, DecisionTree, ExtraTree, KNeighbors, BernoulliNB, Gaussian NB, MultinomialNB and ComplementNB.

From this study, it was concluded that the best classifier was the Linear SVC, which demonstrated 100% accuracy for the training and 74% for the testing. Through the representation of the confusion matrix, the incorrect attribution of examples that served as support was verified. In the final step of this technique, the automatic response was configured, and the submission was tested using an institutional address through the command line, thus successfully triggering the instruction.

Then it went through the same code, but with changes in the categories that were renamed due to the various types of responses produced. All the classifiers used in the previous topic were retested, where Linear SVC was again highlighted with a mediocre value of 53% for the test accuracy, however the results obtained at this stage were lower compared to the transact method.

Ultimately, a code for automatic text generation using neural networks in the context of Deep Learning was designed. It chose the database with the seven categories, as it would be better adjusted to the situation, and selected a two-layer neuronal network where each layer has 256 neurons and a maximum length of 100 characters that were trained over 50 epochs. Despite not recognizing the characters after the sequence, the result was auspicious for the set parameters.

KEYWORDS: automatic response, categories, georeferencing, Machine Learning, management of customer requests.

ÍNDICE GERAL

AGRADECIMENTOS	i
RESUMO	iii
ABSTRACT	v
 1. INTRODUÇÃO	 1
1.1. APRESENTAÇÃO DO PROBLEMA	1
1.2. ORGANIZAÇÃO DA DISSERTAÇÃO	2
 2. ESTADO DA ARTE	 5
2.1. INTRODUÇÃO AO TEMA.....	5
2.2. O SETOR DAS ÁGUAS A NÍVEL NACIONAL	6
2.2.1. SUSTENTABILIDADE DO SETOR.....	7
2.2.2. ENQUADRAMENTO ESTRATÉGICO DO SETOR: PENSAAR 2020	8
2.2.3. MODELOS DE GESTÃO	10
2.2.4. SISTEMA DE ABASTECIMENTO DE ÁGUA	12
2.2.4.1. Sistemas De Abastecimento De Adução	13
2.2.4.2. Sistemas De Abastecimento De Distribuição.....	15
2.2.5. SISTEMA DE SANEAMENTO DE ÁGUAS RESIDUAIS URBANAS.....	17
2.3. DEPARTAMENTO COMERCIAL	21
2.3.1. PROCESSO DE GESTÃO COMERCIAL	21
2.3.2. NOVOS REGULAMENTOS: IMPACTO NAS ENTIDADES GESTORAS.....	22
2.3.2.1. Regulamento de Relações Comerciais dos Serviços de Águas e Resíduos	22
2.3.2.2. Projeto de Regulamento Tarifário dos Serviços de Águas	23
2.4. SATISFAÇÃO DO CLIENTE	25
2.4.1. ESTUDO DE SATISFAÇÃO DOS UTILIZADORES DOS SERVIÇOS DE ÁGUAS E RESÍDUOS (ESUAR 2017)	25
2.4.2. GESTÃO DE RECLAMAÇÕES ERSAR.....	29
2.5. SISTEMAS DE INFORMAÇÃO GEOGRÁFICA	31
2.6. APLICABILIDADE DA INTELIGÊNCIA ARTIFICIAL NA COMUNICAÇÃO COM O CLIENTE.....	33
2.6.1. APRENDIZAGEM DE MÁQUINA	34
2.6.1.1. Aprendizagem Supervisionada	35

2.6.1.2. Aprendizagem Não Supervisionada	35
2.6.1.3. Aprendizagem Aprofundada	36
2.6.2. PROGRAMAÇÃO EM PYTHON	37
2.7. CONCLUSÃO DO ESTADO DA ARTE	39

3. ÂMBITO E OBJETIVOS	41
3.1. ÂMBITO	41
3.2. OBJETIVOS	41

4. CMPEA – EMPRESA DE ÁGUAS DO MUNICÍPIO DO PORTO, EM	43
4.1. APRESENTAÇÃO DA EMPRESA	43
4.2. UNIDADE ORGÂNICA DE GESTÃO DE CLIENTES	44
4.2.1. ENQUADRAMENTO	44
4.2.2. EVENTOS RELEVANTES EM 2018	45
4.2.2.1. Atendimento Comercial.....	46
4.2.2.2. Interação Digital com os Utilizadores Finais	47
4.2.2.3. Atendimento Geral e Técnico.....	48
4.2.2.4. Gestão de Reclamações.....	48

5. METODOLOGIA PARA GESTÃO DE SOLICITAÇÕES DE CLIENTES	49
5.1. PROCEDIMENTO ADOTADO PELA EMPRESA	49
5.2. PROPOSTA DE CLASSIFICAÇÃO PARA SOLICITAÇÕES	52
5.2.1. TIPOLOGIAS	52
5.2.1.1. Atendimento	52
5.2.1.2. Contratação.....	53
5.2.1.3. Faturação, Leitura e Tarifário.....	54
5.2.1.4. Ligação e Disponibilidade	56
5.2.1.5. Prestação de Serviço Técnico.....	57
5.2.1.6. Restantes Categorias	58
5.3. GEORREFERENCIAÇÃO DE SOLICITAÇÕES E RECLAMAÇÕES PROVENIENTES DE CORREIO ELETRÔNICO NA CIDADE DO PORTO	60

5.3.1. REPRESENTAÇÃO DAS SOLICITAÇÕES DA CATEGORIA ATENDIMENTO	61
5.3.2. REPRESENTAÇÃO DAS SOLICITAÇÕES DA CATEGORIA CONTRATAÇÃO	62
5.3.3. REPRESENTAÇÃO DAS SOLICITAÇÕES DA CATEGORIA FATURAÇÃO E LEITURA	63
5.3.4. REPRESENTAÇÃO DAS SOLICITAÇÕES DA CATEGORIA LIGAÇÃO E DISPONIBILIDADE	64
5.3.5. REPRESENTAÇÃO DAS SOLICITAÇÕES DA CATEGORIA PRESTAÇÃO DE SERVIÇO TÉCNICO	65
5.3.6. REPRESENTAÇÃO DAS SOLICITAÇÕES DAS RESTANTES CATEGORIAS.....	66
5.3.7. REPRESENTAÇÃO DAS RECLAMAÇÕES DAS VÁRIAS CATEGORIAS.....	67
5.3.8. SOLICITAÇÕES NÃO COBERTAS PELO SIG	67
5.4. CÓDIGO DESENVOLVIDO EM APRENDIZAGEM SUPERVISIONADA PARA RESPONDER A SOLICITAÇÕES VIA E-MAIL.....	70
5.4.1. APLICAÇÃO DOS VÁRIOS CLASSIFICADORES COM ATRIBUIÇÃO DE RESPOSTAS TIPO GENÉRICAS.....	70
5.4.1.1. Conceção da base de dados	70
5.4.1.2. Análise dos dados	71
5.4.1.3. Tratamento do texto.....	73
5.4.1.4. Classificadores	76
5.4.1.5. Resposta automática.....	85
5.4.2. APLICAÇÃO DE VÁRIOS CLASSIFICADORES COM ATRIBUIÇÃO DE RESPOSTAS TIPO PERSONALIZADAS.....	88
5.4.2.1. Conceção da base de dados	88
5.4.2.2. Análise dos dados	89
5.4.2.3. Classificadores	89
5.4.2.4. Resposta automática.....	91
5.5. GERAÇÃO DE TEXTO AUTOMÁTICO IMPLEMENTANDO UMA ARQUITETURA DE REDE NEURONAL RECORRENTE ARTIFICIAL (LSTM).....	93
5.5.1. IMPORTAÇÃO DOS MÓDULOS E CONVERSÃO DE TEXTO	93
5.5.2. MODELAÇÃO DA REDE NEURONAL	94
5.5.3. USO DO MODELO DE TREINO PARA GERAÇÃO DE TEXTO	96
5.5.4. PREVISÃO DA SAÍDA DE TEXTO	97
5.6. ANÁLISE CONTABILÍSTICA DE SOLICITAÇÕES	100
5.6.1. UO GESTÃO DE CLIENTES	100
5.6.2. UO SALA DE COMANDO.....	100
6. CONCLUSÕES E RECOMENDAÇÕES.....	101
6.1. CONCLUSÕES	101

6.2. RECOMENDAÇÕES PARA TRABALHOS FUTUROS	102
--	------------

REFERÊNCIAS BIBLIOGRÁFICAS	103
---	------------

ÍNDICE DE FIGURAS

Fig. 2.1 - Cobertura da água potável e águas residuais e tratamento nos países da Europa Ocidental	5
Fig. 2.2 - Evolução da cobertura de abastecimento de água e saneamento de águas residuais.....	7
Fig. 2.3 - Níveis de sustentabilidade do setor	7
Fig. 2.4 - Visão, eixos e objetivos operacionais determinados no PENSAAR 2020	10
Fig. 2.5 - Etapas do serviço de abastecimento de água.....	12
Fig. 2.6 - Distribuição geográfica das EG incumbidas da operação de adução.....	14
Fig. 2.7 - Disposição territorial das EG responsáveis pelos serviços de distribuição de água	16
Fig. 2.8 - Constituição típica de um SAR	18
Fig. 2.9 - Distribuição geográfica das entidades gestoras de serviços de saneamento de águas residuais urbanas em alta	18
Fig. 2.10 - Distribuição geográfica das entidades gestoras de serviços de saneamento de águas residuais urbanas em baixa.....	20
Fig. 2.11 - Fases do processo de gestão comercial	22
Fig. 2.12 - Distribuição relativa das respostas facultadas acerca da satisfação global dos utilizadores finais dos serviços de águas e resíduos	25
Fig. 2.13 - Percentagem de inquiridos que contactaram as suas entidades gestoras de serviços de águas e resíduos no último ano.....	26
Fig. 2.14 - Principais motivos pelos quais os inquiridos contactaram a sua entidade gestora de serviços de águas e resíduos.....	27
Fig. 2.15 - Médias de respostas dadas no âmbito da avaliação do serviço de apoio ao cliente	28
Fig. 2.16 - Distribuição relativa das respostas dadas no âmbito de como as reclamações ou sugestões foram resolvidas por parte da respetiva entidade gestora	28
Fig. 2.17 - Evolução das reclamações rececionadas na entidade reguladora desde 2007.....	29
Fig. 2.18 - Progresso das reclamações para a Águas do Porto.....	30
Fig. 2.19 - Representação de diferentes tipos de informação geográfica nos SIG	31
Fig. 2.20 - Características gerais dos dados matriciais	32
Fig. 2.21 - <i>Raster</i> empregue como mapa base	32
Fig. 2.22 - Exemplificação de uma rede neuronal	36
Fig. 4.1 - Objeto social	43
Fig. 4.2 - Organograma da empresa.....	44
Fig. 4.3 - Principais linhas de atuação da UO	45
Fig. 4.4 - Meios de contacto utilizados pelos clientes em 2018	46
Fig. 5.1 - Fluxograma das solicitações enviadas para o endereço cliente@aguasdoporto.pt.....	50

Fig. 5.2 - Fluxograma das solicitações remetidas para o endereço geral@aguasdoporto.pt.....	51
Fig. 5.3 - Georreferenciação das solicitações do tipo atendimento.....	61
Fig. 5.4 - Georreferenciação das solicitações do tipo contratação.....	62
Fig. 5.5 - Georreferenciação das solicitações do tipo faturação, leitura e tarifário.....	63
Fig. 5.6 - Georreferenciação das solicitações do tipo ligação e disponibilidade	64
Fig. 5.7 - Georreferenciação das solicitações do tipo prestação de serviço técnico.....	65
Fig. 5.8 - Georreferenciação das solicitações do tipo restantes categorias.....	64
Fig. 5.9 - Georreferenciação de reclamações de várias tipologias	66
Fig. 5.10 - Representação do número de emails afetos a cada categoria	71
Fig. 5.11 - Percentagem de emails repartidos pelas categorias	72
Fig. 5.12 - Distribuição dos tamanhos dos emails presentes na base de dados	73
Fig. 5.13 - Matriz de confusão obtida para o classificador <i>Linear SVC</i>	78
Fig. 5.14 - Execução do comando para gerar resposta automática.....	87
Fig. 5.15 - Receção do email no destinatário indicado	88
Fig. 5.16 - Produção da resposta automática através da linha de comando <i>Anaconda</i>	92
Fig. 5.17 - Exemplo de duas respostas acionadas.....	92
Fig. 5.18 - Demonstração do processo de geração de texto de uma porção de email aleatório.....	98
Fig. 5.19 - Exemplificação da criação de texto envolvendo uma resposta tipo genérica e parte de um email aleatório	99

ÍNDICE DE TABELAS

Tabela 2.1 - Quadro legal da gestão e da exploração dos sistemas multimunicipais e municipais em Portugal	11
Tabela 2.2 - Vista geral das EG que compõem os serviços de abastecimento público de água	13
Tabela 2.3 - Perspetiva dos serviços de abastecimento de água em adução.....	15
Tabela 2.4 - Panorama dos serviços de abastecimento de água em baixa	17
Tabela 2.5 - Vista dos serviços de saneamento de águas residuais urbanas em alta.....	19
Tabela 2.6 - Visão dos serviços de saneamento de águas residuais urbanas em baixa	21
Tabela 4.1 - Evolução do atendimento aos clientes	46
Tabela 4.2 - Evolução do TME e do TMA no <i>Call Center</i>	47
Tabela 4.3 - Crescimento do TME e TMA nos balcões de atendimento	47
Tabela 5.1 - Solicitações do tipo atendimento.....	53
Tabela 5.2 - Solicitações do tipo contratação.....	54
Tabela 5.3 - Solicitações do tipo faturação, leitura e tarifário	55
Tabela 5.4 - Solicitações do tipo ligação e disponibilidade	56
Tabela 5.5 - Solicitações do tipo prestação de serviço técnico.....	57
Tabela 5.6 - Solicitações das categorias excedentes.....	58
Tabela 5.7 - Evolução do número de clientes por tipologia	60
Tabela 5.8 - Número de solicitações das várias tipologias rejeitadas por dados insuficientes.....	68
Tabela 5.9 - Número de solicitações sem categoria atribuída	69
Tabela 5.10 - Agrupamento das categorias e respetiva designação.....	70
Tabela 5.11 - Performance do classificador usado	77
Tabela 5.12 - Perspetiva geral do desempenho dos vários classificadores	85
Tabela 5.13 - Resultados obtidos para os diversos classificadores	89
Tabela 5.14 - Vista geral do desempenho dos vários classificadores atendendo às inúmeras categorias	91

ACRÓNIMOS E ABREVIATURAS

AA – Abastecimento Público de Água

APDA – Associação Portuguesa de Distribuição e Drenagem de Águas

AMP – Área Metropolitana do Porto

AdDP - Águas do Douro e Paiva, S.A.

AdPorto – Águas do Porto, EM

AEP – Associação Empresarial de Portugal

AR – Saneamento de Águas Residuais Urbanas

BAV – Balcão de Atendimento Virtual da Câmara Municipal do Porto

CIL – Código de Identificação Local

CMP – Câmara Municipal do Porto

CNB - *ComplementNaiveBayes*

CPU – *Central Process Unit*

DECO – Associação Portuguesa para a Defesa do Consumidor

EE – Estação Elevatória

EG – Entidade Gestora

EPAL – Empresa Portuguesa das Águas Livres, S.A.

ERSAR – Entidade Reguladora dos Serviços de Águas e Resíduos

ESUAR – Estudo de Satisfação dos Utilizadores dos Serviços de Águas e Resíduos

ETA – Estação de Tratamento de Água

ETAR - Estação de Tratamento de Águas Residuais

FEUP – Faculdade de Engenharia da Universidade do Porto

GNR – Guarda Nacional Republicana

IA – Inteligência Artificial

LSTM – *Long Short-Term Memory*

ML – *Machine Learning*

MLP – *Multi-Layer Perceptron*

NB – *Naive Bayes*

NLP – Processamento de Linguagem Natural

NLTK – *Natural Language Toolkit*

PENSAAR 2020 – Plano Estratégico de Abastecimento de Água e Saneamento de Águas Residuais 2020

PJ – Polícia Judiciária

PSP – Polícia de Segurança Pública

PT – Planta Topográfica

RASARP – Relatório Anual dos Serviços de Águas e Resíduos em Portugal

RNN – *Artificial Recurrent Neural Network*

SAA – Sistema de Abastecimento de Água

SAR – Sistema de Saneamento de Águas Residuais

SIG – Sistemas de Informação Geográfica

SGD – *Stochastic Gradient Descent*

SVC – *Support Vector Classification*

SVM – *Support Vector Machine*

TF-IDF – *Term Frequency - Inverse Document Frequency*

UE – União Europeia

UO – Unidade Orgânica

1

INTRODUÇÃO

1.1. APRESENTAÇÃO DO PROBLEMA

Uma entidade que se dedica ao serviço de águas intervém em vários setores que estão interligados por algo em comum, nomeadamente, o cliente. Este é a fonte de todo o “motor” das entidades gestoras, o que conduz à área comercial, onde ocorre o contacto direto, telefónico ou digital.

O relacionamento do cliente começou a ganhar alguma importância para as entidades gestoras de serviços públicos, no entanto a variedade dos modelos de gestão, a sua dimensão e o regime de monopólio na prestação do serviço, constituem fatores determinantes na forma como se encara a relação com o cliente. Muitas são as entidades prestadoras que desvalorizam esta vertente da gestão nesta conjuntura, sobretudo com recursos técnicos e humanos, dado o enfoque especial no investimento em infraestruturas e o progresso das taxas de cobertura. A ausência de concorrência não conduz à quebra da responsabilidade que as entidades gestoras têm para com o usuário, visto que o princípio de “utilizador-pagador” se rege essencialmente pela prestação do serviço que obriga ao cliente a pagar, porém a outra parte possui as suas obrigações independentemente das condições legítimas e regulatórias. Superado o problema da conceção de plataformas que possibilitem a atuação nos vários campos de aplicação, surge a necessidade de colmatar essa lacuna com vista à melhoria da qualidade de serviço e satisfação do cliente.

Os serviços comerciais adotam uma posição importante na promoção e desenvolvimento da proximidade com os clientes, devendo por isso canalizar toda a informação e instituir o elo de ligação entre a entidade gestora e os seus clientes, disponibilizando-se a compreender as necessidades e expectativas dos mesmos, o que contribui para uma melhoria na qualidade do serviço prestado.

Os meios atuais de que o cliente dispõe, por vezes, tornam-se ineficientes dado o tempo de espera e em casos excecionais, por erro humano ou técnico.

A gestão de solicitações é marcante na interação da entidade gestora com os consumidores, na medida em que permite otimizar todo o processo desde o contacto até à resposta aos diversos assuntos requisitados e assim proporcionar um serviço eficiente e consequente redução de custos. A criação de um modelo de classificação e a produção de resposta automática são soluções ao atual procedimento que devem ser tidas em consideração para que se possa redarguir em tempo útil e evitar constrangimentos com o utilizador final.

1.2. ORGANIZAÇÃO DA DISSERTAÇÃO

Com esta dissertação pretende-se demonstrar a valor da Inteligência Artificial na relação entre cliente e EG com o desígnio de minorar o tempo de espera por uma resposta. Para tal, começou-se por consultar a bibliografia existente de modo a contextualizar o assunto, concluindo-se a mesma neste seguimento. Posteriormente, são fixados os objetivos gerais deste ensaio, advindo depois o estudo de caso onde é patenteado um projeto realizado no âmbito do tema, e respetiva reflexão acerca das metas atingidas no decorrer do estudo.

Os capítulos que compõem esta dissertação encontram-se dispostos conforme a edificação que se exhibe de imediato.

- **Capítulo 1 – Introdução**

O capítulo inicial relata genericamente o enquadramento da temática e o caso prático desenvolvido culminando na importância para entidade gestora como uma melhoria do serviço prestado ao cliente. Na segunda parte, define-se a estruturação desta dissertação.

- **Capítulo 2 – Estado da Arte**

Dedica-se basicamente a toda a pesquisa bibliográfica pertinente ao objeto em que se debruça a análise.

Primeiramente, refere-se a importância da água na sociedade atual e uma descrição do setor de abastecimento e saneamento de águas.

Em seguida, são exploradas todas as temáticas consideradas de enorme relevância para a área comercial, em concreto, o processo de gestão comercial que foca as várias etapas que envolvem o contacto direto com o cliente e suas características, bem como os regulamentos mais recentes publicados neste domínio: o Regulamento das Relações Comerciais dos Serviços de Águas e Resíduos, que clarifica vários aspetos não contemplados na legislação, e o Projeto de Regulamento Tarifário dos Serviços de Águas, cujo tópico de interesse são as alterações à atual estrutura tarifária.

Na terceira parte do documento, aborda-se a satisfação do cliente onde se insere o estudo de satisfação dos utilizadores dos serviços de águas e resíduos e a gestão de reclamações perpetrada pela ERSAR.

Depois expõe-se uma componente tecnológica, apresentando-se a aplicabilidade dos Sistemas de Informação Geográfica e a sua importância no contexto do problema. Por outro lado, tem-se a Inteligência Artificial que se estende a vários campos de interesse, cujo potenciador passa pela linguagem de programação *Python*.

No final, faz-se uma sinopse conclusiva da pesquisa bibliográfica, reconhecendo possíveis falhas a complementar numa dissertação académica.

- **Capítulo 3 – Âmbito e Objetivos**

Neste capítulo estabelece-se com transparência o âmbito e os objetivos específicos deste trabalho após toda a revisão bibliográfica e conhecimentos aprofundados. Destaca-se como alvo essencial a gestão de solicitações por via da Inteligência Artificial, provenientes de clientes do município do Porto como caso de estudo.

- **Capítulo 4 – Apresentação da Empresa**

Aqui é caracterizado o âmbito da Águas do Porto, EM, a entidade gestora selecionada para a conceção da presente exposição escrita. Concentra-se nas suas principais atividades e resultados obtidos no ano transato respeitante aos indicadores de interesse para este estudo.

- **Capítulo 5 – Gestão de Solicitações de Clientes**

O projeto elaborado assenta principalmente em três patamares: numa primeira abordagem introduz-se um modelo aplicável à gestão de solicitações categorizando-as em várias tipologias, seguindo-se a georreferenciação de pedidos rececionados por correio eletrónico num determinado período de investigação com base nessa metodologia. Por fim, são expostos os códigos criados nos ambientes de aprendizagem de máquina e profunda para responder às solicitações de clientes em prazo útil, acionando a par os devidos recursos e o impacto económico que tem numa empresa de águas.

- **Capítulo 6 – Conclusões e Recomendações**

Surge como último capítulo, onde são descritas todas as considerações finais captadas do estudo executado, e apresentam-se sugestões que ficaram por concretizar devido ao tempo despendido no desenrolamento desta dissertação, recomendando-se assim para trabalhos exequíveis no futuro.

2

ESTADO DA ARTE

2.1. INTRODUÇÃO AO TEMA

A água constitui um bem de cariz vital para o ser humano, porém no decorrer do tempo o ciclo da água à mercê da ação antropogénica tem-se sujeitado a modificações que resultaram na escassez de água desprovida de impurezas em pleno século XXI. A degradação consequente do consumo irracional da água doce atingiu níveis alvoroçantes, apesar da quantidade disponível apenas representar 1% da água total no planeta, e ainda se consta que 24% da população mundial não possui acesso à água tratada. Isto traduz-se num panorama atual em que o setor da água se torna cada vez mais desafiante rogando por soluções inovadoras (AEP, 2014).

Em Portugal, o setor das águas que engloba os serviços de abastecimento público de água e saneamento de águas residuais urbanas é crucial na sociedade moderna. Anteriormente a 1993, a situação global de ambos os serviços prestados às populações apresentava sérias fragilidades em resposta aos novos procedimentos impostos pela UE, pelo que o Governo atuou com vista à reestruturação do setor (AEP, 2014).

De acordo com o Decreto-Lei n.º 194/2009, de 20 de agosto, o serviço de abastecimento público de água às populações, de saneamento de águas residuais urbanas e de gestão de resíduos urbanos representam serviços públicos de carácter estrutural, fundamentais ao bem-estar comum, à saúde pública e à segurança coletiva das populações, às atividades económicas e ao zelo do meio ambiente, regendo-se pelo acesso universal e contínuo da população, elevados níveis de qualidade de serviço, eficiência e equidade dos tarifários praticados. Na Figura 2.1. é observável a percentagem de população abrangida pelos serviços de águas, figurando Portugal entre os vários países da Europa Ocidental.

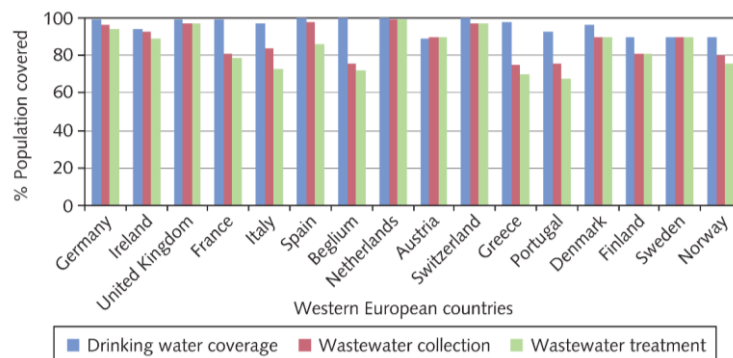


Fig. 2. 1 - Cobertura da água potável e águas residuais e tratamento nos países da Europa Ocidental (Water & Wastewater International, 2011).

Foram efetuados investimentos nas últimas décadas que permitiram um progresso considerável no abastecimento de água e saneamento. Como tal, deve-se perpetuar a sustentabilidade desses investimentos a longo prazo, por forma a asseverar a melhoria do ambiente, o incremento da eficiência e a evolução de um setor que pode contribuir para o crescimento da economia no país. Agora o setor depara-se com novos desafios, quer sejam a resiliência e segurança dos serviços básicos de água e saneamento aclamados como um direito humano, com enfoque para a gestão eficiente dos seus recursos, promovendo a qualidade do serviço conferido que se reflita na sustentabilidade social, económica e financeira. O conteúdo destes direitos é determinado pelos critérios de acessibilidade física, disponibilidade, qualidade, acessibilidade económica e aceitabilidade, no entanto o acesso à água e ao saneamento devem ser afiançados mediante os princípios de não discriminação, permitindo o acesso a minorias ou grupos desfavorecidos, a participação dos interessados e a responsabilização coletiva (ERSAR, 2012) (PENSAAR 2020, 2015).

Deseja-se, portanto, o acesso a um serviço público de água e saneamento de qualidade e adequado às suas necessidades com custos socialmente plausíveis junto das populações (PENSAAR 2020, 2015).

2.2. O SETOR DAS ÁGUAS A NÍVEL NACIONAL

Os últimos 20 anos destacaram uma evolução positiva e importantíssima na qualidade da água distribuída aos cidadãos portugueses e na qualidade do tratamento das águas residuais urbanas, expressa em indicadores de qualidade e de cobertura que colocam Portugal ao nível dos países mais desenvolvidos da UE e do mundo (Águas do Porto, 2018).

No início da década de 90, a cobertura do serviço de abastecimento de água era cerca de 80%, tendo vindo a aumentar desde então esta percentagem contínua e significativamente para atingir 95% em 2011, com 99% nas zonas urbanas e 90% nas rurais. Assim foi assegurado o cumprimento do objetivo definido no PEAASAR 2007-2013 em relação ao abastecimento de água com um valor de 95% (Águas do Porto, 2018).

Segundo o Relatório Anual dos Serviços de Águas e Resíduos em Portugal (RASARP) de 2018, publicado pela ERSAR com dados que remontam a 2017, este valor não se modificou até 2014, assinalando-se em 2015 uma melhoria na cobertura do serviço de abastecimento de água, cujo valor alcançou os 96 % e se manteve em 2017. De notar que, apesar de ter sido ultrapassada a meta definida para a cobertura do serviço de abastecimento de água (calculada pelo indicador “Acessibilidade física do serviço (AA01)”), averigua-se que o valor da adesão ao serviço apresenta ainda potencialidade de melhoria, com 87% em 2017, refletindo assim situações decorrentes da existência de habitações devolutas ou da utilização de origens alternativas de água (Águas do Porto, 2019b).

Em relação às atividades de drenagem e tratamento de águas residuais perduram ainda dificuldades diversas no cumprimento das obrigações legais nacionais e comunitárias, apesar de se terem vindo a registar melhorias importantes. O progresso da cobertura do serviço de saneamento tem vindo a aumentar na última década, ainda que a um ritmo mais vagaroso quando comparado com o serviço de AA. Atendendo ao RASARP 2018, a cobertura do serviço de drenagem de águas residuais e de tratamento de águas residuais era ainda de 84% e 83%, respetivamente, não tendo sido alcançado a meta fixada no PEAASAR II para 2013, de prover 90% da população com rede de drenagem e com tratamento de águas residuais. Quanto à adesão ao serviço de saneamento de águas residuais, este mostra ainda potencial de melhoria, de maneira análoga ao que se constata no abastecimento de água, registando 89,4% em 2017, o que reflete situações decursivas da utilização de soluções de tratamento individuais (Águas do Porto, 2019b).

A evolução da cobertura dos serviços de águas pode ser visualizada na Figura 2.2.

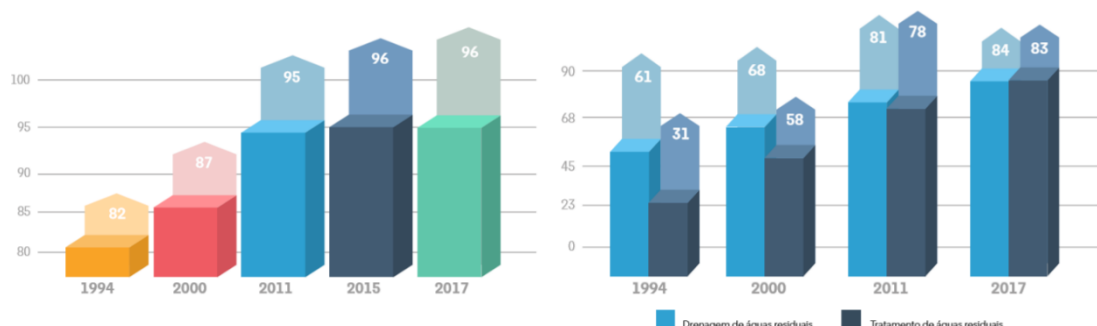


Fig. 2. 2 - Evolução da cobertura de abastecimento de água e saneamento de águas residuais (Adaptado de Águas do Porto, 2019b)

O setor de águas e resíduos contribui significativamente para o desenvolvimento económico e social do País, tanto pela capacidade de gerar atividade económica e de criar emprego e riqueza, como pela crescente melhoria que tem conferido às condições de vida da população, gerando externalidades económicas noutros setores (ERSAR, 2018b).

2.2.1. SUSTENTABILIDADE DO SETOR

A ERSAR atua a nível da regulação dos setores dos serviços de abastecimento público de água, de saneamento de águas residuais urbanas e de gestão de resíduos urbanos e é a autoridade à qual compete a coordenação e fiscalização do regime da qualidade da água para consumo humano. A ERSAR tem ainda como missão assegurar as condições de igualdade e de transparência no acesso e no exercício da atividade no setor, bem como consolidar um efetivo direito público à informação geral sobre os serviços e sobre cada uma das entidades gestoras (ERSAR, 2018b).

Intervém em três patamares, cuja incumbência passa por garantir a sustentabilidade global do setor.

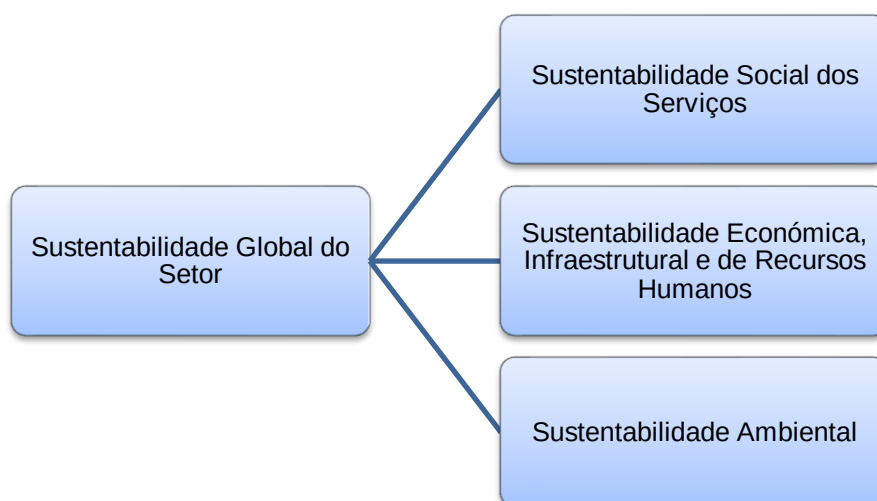


Fig. 2. 3 - Níveis de sustentabilidade do setor (ERSAR, 2018b)

No âmbito da sustentabilidade social dos serviços, a entidade atua com o propósito de assegurar a correta proteção dos interesses dos utilizadores finais dos serviços de águas e resíduos permitindo o acesso ao serviço, a sua adequada qualidade e da razoabilidade do preço e relativamente à sustentabilidade

ambiental no uso de recursos ambientais e na prevenção da poluição, a nível dos impactos dos serviços na água, no ar e no solo (ERSAR, 2018b).

2.2.2. ENQUADRAMENTO ESTRATÉGICO DO SETOR: PENSAAR 2020

O PENSAAR 2020 – Uma Nova Estratégia para o Setor de Abastecimento de Água e Saneamento de Águas Residuais comporta a ferramenta estratégica para o setor das águas em Portugal Continental, entre 2014 a 2020, sucedendo ao Plano Estratégico de Abastecimento de Água e de Saneamento de Águas Residuais 2007-2013 (PEAASAR II).

O balanço do PEAASAR II assentou na avaliação da evolução atingida para cada um dos 9 objetivos operacionais e das 8 medidas definidas no Plano, a partir da qual se reconheceram melhorias a colocar em prática no futuro, bem como objetivos e medidas importantes, justificando a sua agregação na estratégia sucessora. Dessa avaliação sobressaem os aspetos atinentes aos objetivos operacionais, resultados alcançados e relevância para o PENSAAR 2020 (PENSAAR 2020, 2015):

- Êxito no incremento da acessibilidade do serviço de abastecimento de água à população, que cumpriu o objetivo fixado de uma taxa de cobertura de 95%;
- Execução dos objetivos consequentes do normativo nacional e comunitário, apesar da ocorrência de alguns casos de incumprimento que têm sido solucionados;
- Obtenção de níveis adequados de qualidade do serviço comensuráveis através dos indicadores definidos pela ERSAR;
- Prestação de serviço a 90% da população com sistemas públicos de saneamento de águas residuais não foi sucedida, podendo aumentar indiretamente através do cumprimento do normativo e a melhoria da qualidade das massas de água;
- Repercussões positivas na qualidade das águas balneares, resultante do investimento no tratamento de águas residuais com benefícios para o setor do turismo e na qualidade de vida da população em geral. O impacto positivo na qualidade das massas de água também foi notório, embora de uma forma menos evidente, como atestam os valores apresentados, devido aos efeitos contrários da poluição não-doméstica originada noutros sectores, aliado ao aumento da informação sobre a qualidade das massas de água, que veio evidenciar problemas não quantificados no passado;
- Dinamização do setor empresarial privado através de concessões nos sistemas municipais e de subcontratação nos sistemas multimunicipais, a otimização da gestão operacional e a anulação de custos de ineficiência, a recuperação completa dos gastos dos serviços e a abordagem integrada na prevenção e controlo da poluição estão entre os objetivos operacionais que ficaram mais aquém do previsto.

Os pressupostos para o novo instrumento consistem em (APA, s.d.):

- Apoiar a nova estratégia para o setor nos pilares em que assentaram os anteriores planos estratégicos para o setor, nomeadamente o PEAASAR I (Plano Estratégico de Abastecimento de Água e de Saneamento de Águas Residuais 2000-2006) e o PEAASAR II para o período de 2007-2013;
- Reconhecer e aclarar consistentemente os problemas que lesam o setor;
- Estabelecer a estratégia com base em metas de sustentabilidade em todas as suas atividades – técnica, ambiental, económica, financeira e social – gerando um contexto de aceitação global a médio, para o período de regência da nova estratégia de 2014-2020, e a longo prazo, que se estende para além de 2020;

- Agrupar essa estratégia de sustentabilidade a médio e longo prazo a uma parceria vantajosa para que todas as vertentes setoriais possam coligar-se e obter ganhos em conjunto, possibilitando um salto qualitativo do setor à semelhança do passado quando foi possível reunir esse consenso e compromisso alargados;
- Produzir um estratagema dinâmico cuja implementação possa ser assegurada através de um Grupo de Apoio à Gestão (GAG), que assevere o suporte à boa governança do setor de uma forma contínua, incluindo a monitorização e atualização anual do PENSAAR 2020 a partir de uma plataforma de informação setorial a nível nacional que incorpore os dados das entidades encarregues pelo planeamento e regulação do setor, partilhada por todos os parceiros setoriais e acessível aos utilizadores e cidadãos;
- Colaborar para que o setor alcance a excelência com desempenho elevado num cenário que exige também solidariedade e equidade, concentrando as forças num setor que processa um bem económico e social.

Para o diagnóstico da situação atual, divulgam-se as repressões mais relevantes que devem ser resolvidas através da estratégia setorial para o período 2014-2020, sobretudo, o desconhecimento das infraestruturas de AA e de SAR por elevado número de EG, quanto à distribuição de água e drenagem de águas residuais; o grau de utilização de algumas infraestruturas; a aderência dos utilizadores ao serviço em baixa e a ligação dos sistemas municipais aos multimunicipais, com valores abaixo do que se esperava; o ritmo de reabilitação dos ativos, nitidamente insuficiente e estimado em cerca de 50% do recomendado pelas boas práticas, com riscos para a qualidade dos ativos que se reflete nas perdas físicas de água distribuída e nas afluências indevidas às redes de coletores de saneamento através de infiltrações; a presença de volumes de água não faturada ainda bastante elevados; o desconhecimento dos gastos na prestação dos serviços de AA e de AR de um grande número de EG, principalmente nos sistemas em baixa; a recuperação de gastos longe de níveis que garantam a sustentabilidade económica e financeira das EG e do setor em geral; as tarifas atualmente em vigor são razoavelmente baixas para permitir uma acessibilidade económica que se estende a quase todos os utilizadores, porém com consequências negativas na sustentabilidade económico-financeira de algumas EG averiguando-se grandes disparidades e ainda os problemas atuais de disponibilidade de recursos financeiros, nomeadamente, o acesso ao endividamento das EG. Os resultados do diagnóstico da situação atual conduzem a um novo paradigma que orientou as fases seguintes do PENSAAR 2020, sustentado numa estratégia menos centrada na realização de infraestruturas para aumento da cobertura e focalizando-se mais na gestão dos ativos, no seu funcionamento e na qualidade dos serviços prestados (PENSAAR 2020, 2015).

Ponderando o balanço do PEAASAR II e a avaliação da situação atual, o PENSAAR 2020 apresenta cinco objetivos estratégicos, com a designação de eixos, e dezanove objetivos operacionais (OP) que se podem visualizar no diagrama da Figura 2.4.

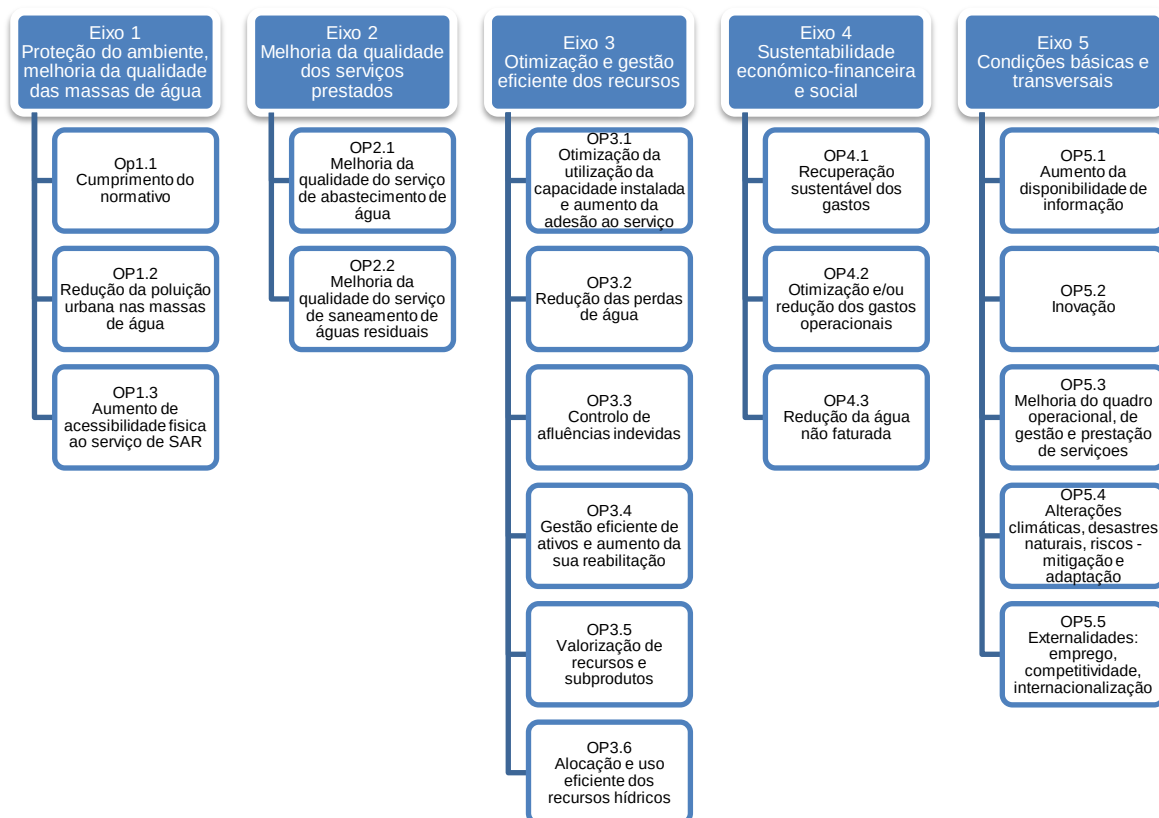


Fig. 2. 4 - Eixos e objetivos operacionais determinados no PENSAAR 2020 (Adaptado de PENSAAR 2020, 2015)

O PENSAAR 2020 propõe um conjunto importante de indicadores que permitirão avaliar o progresso do Plano para cada um dos objetivos operacionais. Como tal, foram definidos valores de base e as respetivas metas que se pretende atingir em 2020 para cada indicador. Alguns destes indicadores são apenas utilizados para aferir a evolução do indicador sem estabelecer uma meta. A maioria dos objetivos operacionais contempla vários indicadores, de forma a avaliar a evolução de forma minuciosa (ERSAR, 2018b).

Com as medidas e as ações tenciona-se concretizar e detalhar cada um dos objetivos operacionais definidos no Quadro Estratégico, constituindo assim parte de um plano de ação da estratégia que ajudará nas decisões a tomar pelos parceiros sectoriais, particularmente os promotores responsáveis pela realização dos projetos e as entidades responsáveis pela sua aprovação, financiamento e monitorização (ERSAR, 2018b).

2.2.3. MODELOS DE GESTÃO

A gestão dos serviços de abastecimento de água é da responsabilidade do Estado e autarquias, sobretudo os sistemas municipais e multimunicipais. São considerados multimunicipais, os sistemas de titularidade estatal que sirvam pelo menos dois municípios com a intervenção do Estado por razões de interesse nacional, e sistemas municipais, nos quais é da competência dos municípios, separadamente ou em conjunto, através de associações de municípios ou em parceria com o Estado, definir o modelo de organização e gestão (ERSAR, 2018b).

A Lei n.º 88-A/97, de 25 de julho, alterada pela Lei n.º 35/2013, de 11 de junho (denominada Lei de Delimitação de Setores), estabelece as circunstâncias em que se admite o ingresso de entidades privadas na gestão destes serviços. Nos sistemas de abastecimento de água e de saneamento de águas residuais, os operadores privados apenas podem assumir uma posição minoritária no capital das empresas concessionárias no caso dos sistemas multimunicipais (ERSAR, 2018b).

A Tabela 2.1 apresenta as várias tipologias de entidades gestoras que advêm dos diferentes modelos de gestão adotados pelos municípios e pelo Estado para a prestação dos serviços de águas e resíduos.

Tabela 2. 1 - Quadro legal da gestão e da exploração dos sistemas multimunicipais e municipais em Portugal (Adaptado de ERSAR, 2012)

Titularidade	Modelo de gestão	Entidade gestora
Estatual	Direta	Estado (não aplicável atualmente)
	Delegada	Empresa pública (no caso da EPAL)
	Concessionada	Empresa de capitais maioritariamente públicos do Estado ou dos municípios
Municipal ou Intermunicipal		Serviços municipais (não aplicável)
	Direta	Serviços municipalizados ou intermunicipalizados
		Associação de municípios
	Delegada	Empresa de capitais estatais e municipais em parceria com os municípios e o Estado
		Empresa do setor empresarial local sem participação do Estado
	Concessionada	Entidade concessionária municipal

Nos termos do Decreto-Lei n.º 194/2009, de 20 de agosto, o modelo de gestão delegada compreende a operação, a manutenção e conservação do sistema e pode incluir ainda a construção, renovação e substituição das infraestruturas, instalações e equipamentos na totalidade ou em parte do território da entidade delegante. A delegação é estabelecida através da celebração de contrato de gestão delegada entre o município, a associação de municípios ou a área metropolitana e a empresa municipal delegatária, no qual deve constar o âmbito da delegação especificando os serviços, a tipologia de utilizadores, o espaço territorial abrangido, assim como a data a partir da qual a empresa municipal delegatária assume a responsabilidade pela prestação dos serviços. O prazo mínimo de vigência é de 10 anos, onde devem ser definidos os objetivos, a identificação das principais iniciativas de carácter estratégico a implementar, o plano de investimentos e o tarifário, pelo qual a empresa municipal delegatária se deve reger.

Quanto às especificações definidas para a concessão, esta apresenta as mesmas atividades descritas na situação de delegação. De acordo com o Decreto-Lei n.º 194/2009, de 20 de agosto, o prazo da concessão fixado obedece ao disposto no n.º 1 do artigo 410.º do Código dos Contratos Públicos, não podendo exceder, incluindo a duração de qualquer prorrogação, 30 ou 15 anos conforme haja ou não investimento significativo de expansão, modernização ou reabilitação a cargo do concessionário. No entanto, a decisão de conceder deve ser precedida de estudo que demonstre a viabilidade financeira da

concessão e a racionalidade económica e financeira acrescida decorrente do desenvolvimento da atividade através deste modelo de gestão, designadamente em função de expectáveis ganhos de eficiência e de transferência para o concessionário de riscos passíveis por este serem melhor geridos. Para tal, do contrato de concessão constam, obrigatoriamente, o tarifário a seguir, o plano de investimentos da concessão, o caso base do modelo financeiro da concessão e os proveitos mínimos anuais. No caderno de encargos devem estar explicitados os objetivos e condições a atingir no exercício de funções, o modelo de partilha de riscos, eventuais investimentos da responsabilidade do concedente, eventuais obras da responsabilidade de terceiros, a identificação dos serviços passíveis de faturação, o regime de multas contratuais, entre outros, sendo possível abranger exigências do concedente e eventuais restrições ao modelo de financiamento e em anexo devem encontrar-se, por exemplo, o inventário dos bens e relações jurídicas e avaliação das suas condições de conservação e funcionamento.

2.2.4. SISTEMA DE ABASTECIMENTO DE ÁGUA

O serviço de abastecimento público de água abrange várias etapas, das quais é fulcral referir a captação, o tratamento, a elevação da água até aos pontos de cota mais acentuada, o transporte, o armazenamento, a distribuição pelos ramais de água na rede pública e a utilização da água pelos consumidores a partir de uma combinação de equipamentos e instalações.

Nesta atividade, os sistemas subdividem-se em sistemas em alta e sistemas em baixa. Os primeiros são constituídos por um conjunto de componentes a montante da rede de distribuição que unem o meio hídrico ao sistema em baixa, enquanto este último caracteriza-se por um conjunto de elementos que permitem a prestação do serviço de abastecimento de água ao utilizador final. Ambos os sistemas podem ser incorporados sempre que haja vínculo entre o meio hídrico e o consumidor (Olaia, A., 2012).

Na figura que se segue é revelada a cadeia de valor das atividades do setor de abastecimento de água.



Fig. 2. 5 - Etapas do serviço de abastecimento de água (ERSAR, 2019a)

- Captação – Recolha de água bruta do meio hídrico superficial ou subterrâneo;
- Tratamento na ETA – Instalações onde é conferido à água as características físicas, químicas e microbiológicas para se tornar potável;
- Estação Elevatória (EE) – Elevação de água até aos pontos de cota mais elevada por meio de bombas hidráulicas com o intuito desta circular sob pressão;
- Adução – Transporte de água tratada da zona de produção através de tubagens até à rede de abastecimento, conectando os diversos equipamentos e instalações;
- Reservatório – Armazenamento de água tratada na instalação e regularização dos consumos, preservando a continuidade no abastecimento;
- Rede de distribuição – Compreende a rede de entrega aos consumidores, constituída por condutas instaladas na rede pública que conduzem a água até aos ramais e ligam os prédios

domiciliário em quantidade e pressão necessárias, sendo assegurado o fornecimento do serviço aos alojamentos (Águas do Porto, 2014) (ERSAR, 2018b).

Os SAA normalmente são classificados em adução e distribuição, mediante as atividades realizadas pelas EG donde se destacam os consumos domésticos, comerciais, industriais, públicos e de incêndio. Assim, surgiram os sistemas multimunicipais, maioritariamente encargos pela adução, e os sistemas municipais responsáveis pela distribuição.

Na Tabela 2.2 expõe-se o total de entidades gestoras que constam da vista geral do domínio regulado pela ERSAR.

Tabela 2. 2 - Vista geral das EG que compõem os serviços de abastecimento público de água (Adaptado de ERSAR, 2018b)

Modelo de Gestão	Submodelo de Gestão	Abastecimento de Água		
		Alta	Baixa	Total
Gestão Concessionada	Concessões Multimunicipais	6	1	6
	Concessões Municipais	1	28	29
Gestão Delegada	Delegações Estatais	1	1	1
	Parcerias Estado/Municípios	1	2	3
	Empresas Municipais ou Intermunicipais	1	23	24
	Juntas de Freguesia	0	61	61
Gestão Direta	Associações de Municípios	0	0	0
	Serviços Municipalizados ou Intermunicipalizados	1	20	20
	Serviços Municipais	1	183	183
Total		12	319	327

2.2.4.1. Sistemas De Abastecimento De Adução

Tal como se explicitou no ponto anterior, o sistema de adução inclui os componentes a montante da rede de abastecimento, isto é, desde a extração do meio hídrico até ao armazenamento onde se observam diversas ligações entre condutas adutoras e reservatórios. A finalidade das entidades gestoras que atuam neste ramo do serviço é captar, tratar e vender água tratada às EG designadas para o fornecimento de água ao utilizador final.

Na Figura 2.6 encontra-se patente a distribuição geográfica das EG do serviço de abastecimento público de água em alta a operar até à data de 31 de dezembro de 2017.

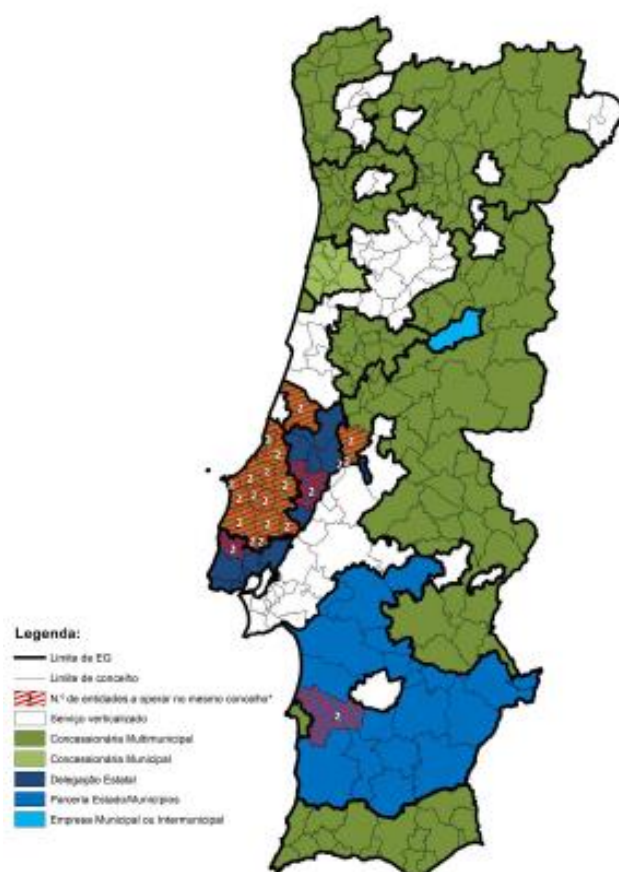


Fig. 2. 6 - Distribuição geográfica das EG incumbidas da operação de adução (ERSAR, 2018b)

Atendendo à figura prévia constata-se um elevado número de municípios em Portugal Continental geridos por entidades concessionárias, mais precisamente concessões multimunicipais que se estendem a 72% da população, depois seguem-se as parcerias estado/municípios, as concessões municipais e, por fim, as empresas municipais ou intermunicipais. Ainda é possível denotar a presença de municípios cujo serviço prestado é verticalizado, ou seja, as entidades que efetuam todas as operações do sistema que se inicia pela captação a culminar posteriormente no fornecimento do serviço, o que em suma corresponde a uma cadeia de valor. Existem 120 municípios sob a alçada deste submodelo com um total de 3 milhões de habitantes, aglomerando-se especialmente no Centro e Norte do país (ERSAR, 2018b).

Na Tabela 2.3 mostra-se a informação detalhada para os submodelos de gestão.

Tabela 2. 3 - Perspetiva dos serviços de abastecimento de água em adução (Adaptado de ERSAR, 2018b)

Submodelo de Gestão	EG	Concelhos abrangidos	Área abrangida (km ²)	População abrangida (milhares de hab.)	Densidade populacional (hab./km ²)
Concessões Multimunicipais	6	174	51 863	5 105	98
Concessões Municipais	1	8	899	190	211
Delegações Estatais	1	25	3 885	1 779	458
Parcerias Estado/Municípios	1	20	15 553	245	16
Empresas Municipais ou Intermunicipais	1	1	435	49	112
Serviços Municipalizados ou Intermunicipalizados	1	1	22	0,27	12,2
Serviços Municipais	1	1	72	0,02	0,3

Com base na tabela anterior, depreende-se que as concessões multimunicipais dominam o setor em alta com 6 EG num universo de 174 municípios, onde se averigua o maior número de habitantes como seria expectável. As delegações estatais mostram-se como o segundo submodelo de gestão com maior peso, alusivas a 25 municípios que ascendem aos cerca de 1,8 milhões de habitantes apesar de constituir apenas 1 EG responsável. Os restantes submodelos de parcerias estado/municípios, concessões municipais, empresas municipais e outros submodelos encontram-se menos representados, no entanto é de salientar que as parcerias estado/municípios intervêm em aproximadamente 21% do território continental, numa área que cobre grande porção do Alentejo, mormente a Águas Públicas do Alentejo (ERSAR, 2018b).

De referir ainda que em 2017 foi marcado pela desagregação do sistema administrado pela Águas do Norte, passando a Águas do Douro e Paiva a gerir o sistema resultante deste processo fornecendo o serviço de abastecimento público de água de adução em 20 municípios localizados maioritariamente na AMP, enquanto as Águas do Norte ficou encarregue do serviço em alta dos outros 53 municípios que permaneceram no seu sistema (ERSAR, 2018b).

2.2.3.2. Sistemas De Abastecimento De Distribuição

Os sistemas de abastecimento em baixa estão associados ao transporte de água a partir dos reservatórios em direção aos utilizadores finais, cujos componentes permitem o seu fornecimento em pressões e qualidade adequadas.

A figura subsequente ostenta a distribuição geográfica das EG que compõem o setor de abastecimento de água em baixa em função dos submodelos de gestão a 31 de dezembro de 2017.

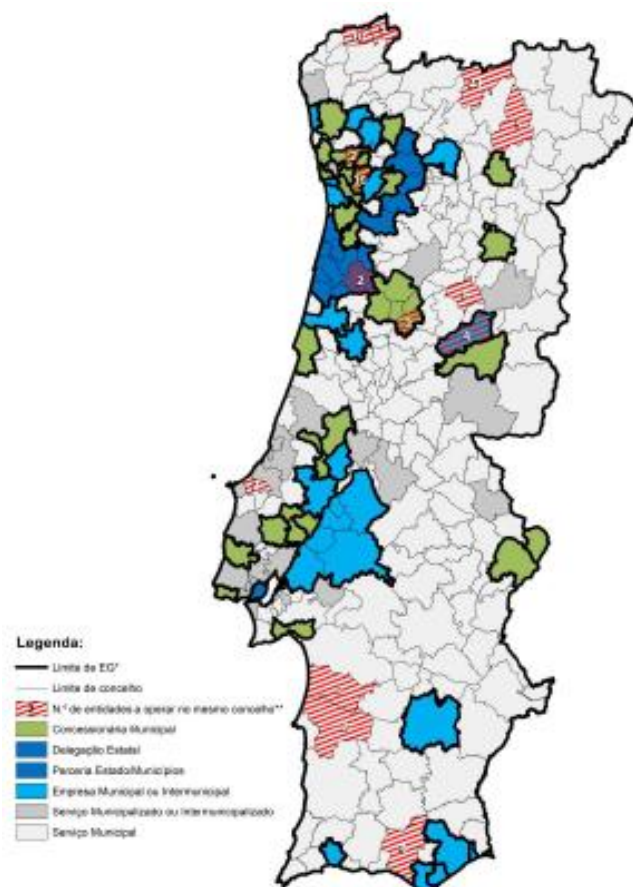


Fig. 2. 7 - Disposição territorial das EG responsáveis pelos serviços de distribuição de água (ERSAR, 2018b)

Considerando a figura antecedente, é exequível inferir que o provimento desta vertente do serviço é assegurado predominantemente pelos modelos de gestão direta (serviços municipais) que ronda 70% do total de municípios e perto de 52% da população de Portugal Continental, sendo evidente também alguma representatividade dos submodelos de concessão municipal, as empresas municipais ou intermunicipais e os serviços municipalizados ou intermunicipalizados que geralmente intercedem no litoral ou áreas densamente urbanas.

Na Tabela 2.4 estão representados pormenorizadamente estes serviços.

Tabela 2. 4 - Panorama dos serviços de abastecimento de água em baixa (Adaptado de ERSAR, 2018b)

Submodelo de Gestão	EG	Concelhos abrangidos	Área abrangida (km ²)	População abrangida (milhares de hab.)	Densidade populacional (hab./km ²)
Concessões Multimunicipais	1	1	75	11	142
Concessões Municipais	28	33	7 609	1 971	259
Delegações Estatais	1	1	100	533	5 524
Parcerias Estado/Municípios	2	15	2 702	479	177
Empresas Municipais ou Intermunicipais	23	28	8 888	1 793	202
Juntas de Freguesia	61	9	1 276	41	32
Serviços Municipalizados ou Intermunicipalizados	20	22	7 688	2 316	301
Serviços Municipais	183	183	61 939	2 922	47

Atentando à tabela precedente, apura-se um número total de 319 entidades gestoras em que os serviços municipais incidem sobre o maior número de habitantes com 2,9 milhões de habitantes respeitantes a 183 municípios, contudo imperam em zonas rurais sobretudo no interior onde se regista menor densidade populacional, o que corresponde a 47 hab./km². Os restantes submodelos de gestão apesar de abarcarem menos municípios, acabam por servir ainda um elevado número de habitantes dada a sua intervenção mais propensa a áreas urbanas (ERSAR, 2018b).

A segmentação do setor que originou o elevado número de EG deve-se à prestação do serviço pelas autarquias e juntas de freguesia (micro entidades), facto que motivou a concretização da relação contratual estabelecida entre ambas as entidades, abrangendo 41 mil habitantes. Para além disso existem igualmente situações em que um concelho detém mais do que uma EG a operar (ERSAR, 2018b).

Nos últimos 20 anos, assistiu-se ao aumento gradual de entidades empresariais no setor, pelo que na atualidade 48% do setor é formado por modelos de gestão concessionada e delegada. Confirma-se o mesmo número de EG à semelhança do apresentado no RASARP 2017 (ERSAR, 2018b).

2.2.5. SISTEMA DE SANEAMENTO DE ÁGUAS RESIDUAIS URBANAS

O ciclo urbano da água aglomera todas as etapas das atividades de abastecimento de água e saneamento de águas residuais, desde a captação da água até à rejeição final da água residual na natureza (ERSAR, s.d.). A Figura 2.8 representa todas as fases de um sistema de saneamento de águas residuais.



Fig. 2. 8 - Constituição típica de um SAR (ERSAR, 2019)

A atividade de saneamento de águas residuais urbanas engloba os estágios de descarga, drenagem, elevação, transporte e tratamento das águas residuais de origem urbana, bem como a sua rejeição no meio hídrico. Esta atividade é elementar para assegurar a qualidade das massas de água, sendo determinante no condicionamento dos outros usos do domínio hídrico, particularmente a captação de água para consumo humano. O sistema de saneamento de águas residuais em baixa atesta a drenagem de águas residuais urbanas junto ao produtor, rejeitando-as num sistema em alta, ou, em destino final adequado, caso se trate de um sistema integrado. Um sistema em alta é constituído por um conjunto de componentes que permite a ligação do sistema em baixa ao ponto de rejeição (ERSAR, s.d.).

A Figura 2.9 apresenta a distribuição geográfica das entidades gestoras do serviço de saneamento de águas residuais em alta, por submodelo de gestão, a 31 de dezembro de 2017.

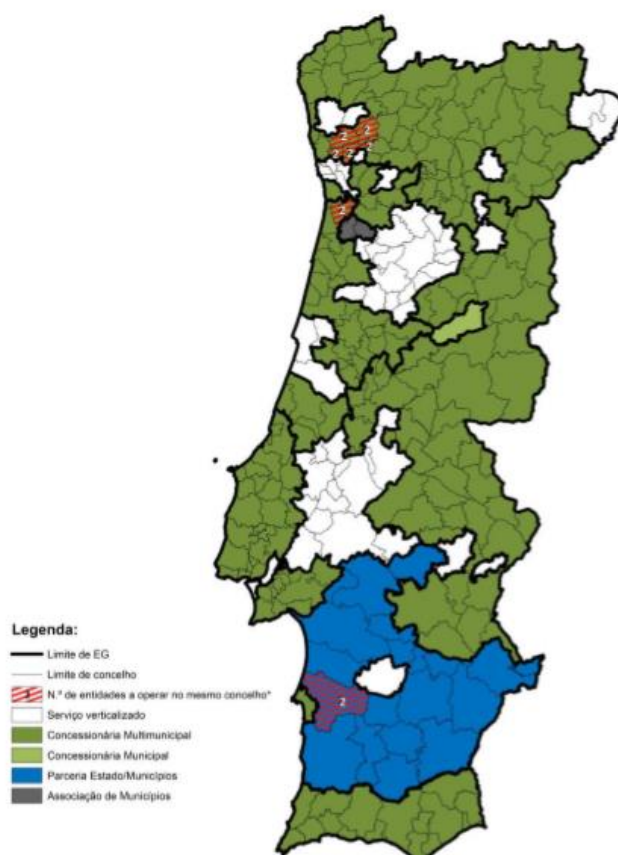


Fig. 2. 9 - Distribuição geográfica das entidades gestoras de serviços de saneamento de águas residuais urbanas em alta (ERSAR, 2018b)

A maior quota deste serviço continua a ser coberto por entidades concessionárias, englobando perto de 96% da população e 90% do número de municípios abrangidos por entidades gestoras que prestam este processo produtivo. Relativamente à incorporação das operações desde a recolha e drenagem do efluente até ao seu tratamento e destino final por parte de uma entidade gestora, esta é inferior ao que se assinala no abastecimento. De facto, abrange 150 municípios e sensivelmente 2,7 milhões de habitantes (ERSAR, 2018b).

Tabela 2. 5 - Vista dos serviços de saneamento de águas residuais urbanas em alta (Adaptado de ERSAR, 2018b)

Submodelo de gestão	Entidades gestoras	Concelhos abrangidos	Área abrangida (km ²)	População abrangida (milhares de hab.)	Densidade populacional (hab./km ²)
Concessões multimunicipais	8	202	47 245	6 957	147
Concessões municipais	2	6	1 031	402	390
Parcerias Estado/municípios	1	20	15 553	245	16
Associações de municípios	1	4	176	93	530

O submodelo de gestão preponderante no SAR em alta é o das concessões multimunicipais, que contém aproximadamente 7,0 milhões de habitantes e um total de 202 municípios, 90% e 87% do universo atinente ao serviço de saneamento em alta, respetivamente. O submodelo das parcerias Estado/municípios detém, ainda assim alguma representatividade em termos da área abrangida, com 24% do total, à semelhança do registado no abastecimento de água em alta no caso da Águas Públicas do Alentejo, cuja prestação de serviço alcança grande parte do Alentejo especialmente em áreas maioritariamente rurais (ERSAR, 2018b).

Na Fig. 2.10 pode ver-se a distribuição geográfica das EG do serviço de abastecimento público de água de distribuição, por submodelo de gestão, a 31 de dezembro de 2017.

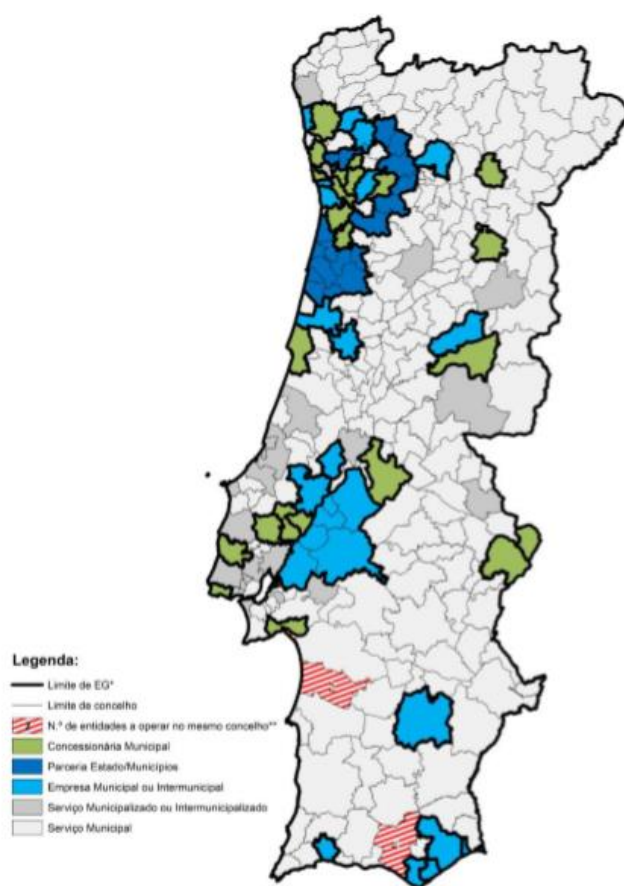


Fig. 2. 10 - Distribuição geográfica das entidades gestoras de serviços de saneamento de águas residuais urbanas em baixa (ERSAR, 2018b)

O setor das águas residuais em baixa é identicamente subjugado pela existência de um número elevado de entidades gestoras, 257, na sua generalidade de pequena dimensão, sendo o seu âmbito de atuação o próprio município em que se situam. De forma idêntica ao que sucede no abastecimento de água, o modelo com maior preeminência na prestação deste serviço é o da gestão direta, com 75% dos municípios e cerca de 59% da população de Portugal continental a ser compreendida por esta categoria (ligeiramente mais do que se verifica no abastecimento de água). Encontra-se maior predominância dos modelos de gestão delegada ou concessão no litoral ou nos grandes centros urbanos (ERSAR, 2018b).

Na Tabela 2.6 apresenta-se o panorama do setor de saneamento de águas residuais em baixa.

Tabela 2. 6 - Visão dos serviços de saneamento de águas residuais urbanas em baixa (Adaptado de ERSAR, 2018b)

Submodelo de gestão	Entidades gestoras	Concelhos abrangidos	Área abrangida (km ²)	População abrangida (milhares de hab.)	Densidade populacional (hab./km ²)
Concessões municipais	23	23	6 262	1 714	274
Parcerias Estado/municípios	2	18	3 129	642	205
Empresas municipais ou intermunicipais	23	28	8 961	1 796	200
Serviços municipalizados ou intermunicipalizados	19	21	6 973	2 277	327
Serviços municipais	190	190	64 022	3 635	57

Averigua-se uma maior representatividade a nível dos serviços municipais na prestação do serviço de saneamento de águas residuais em baixa, abrangendo aproximadamente de 3,6 milhões de habitantes e 190 municípios. Por regra, esta tipologia de entidades atua em meios com pouca densidade populacional (57 hab./km²), com características rurais, o que vai de encontro ao observado no mapa da Figura 2.10, onde se pode apurar que a distribuição dos serviços municipais se encontra principalmente no interior do País em zonas menos povoadas. Em relação aos restantes submodelos verifica-se que, apesar de não terem a mesma predominância que os serviços municipais, as concessões municipais, as empresas municipais ou intermunicipais e os serviços municipalizados ou intermunicipalizados estendem-se aos 1,7 milhões de habitantes, 1,8 milhões de habitantes e 2,3 milhões de habitantes, respetivamente (ERSAR, 2018b).

2.3. DEPARTAMENTO COMERCIAL

2.3.1. PROCESSO DE GESTÃO COMERCIAL

Geralmente o primeiro contacto do cliente com a empresa passa pela celebração da contratação do serviço que possui um contador instalado, sujeito a leitura periódica. Considerando a leitura registada, estima-se o consumo de água e a fatura é emitida. Ao cliente é cobrado um valor faturado a pagar dentro dos prazos previstos, ciclo de leitura, faturação e cobrança repete-se até ao momento em que termine o contrato de abastecimento (Sousa, S. [et al.], 2017).

No entanto, existem vários fatores e circunstâncias que condicionam a concretização do processo conforme pretendido. Na gestão contratual com os clientes, averigua-se um certo nível de complexidade e elevado número de interfaces, quer com os clientes quer com outras áreas da empresa, havendo uma grande margem para erros ou falhas operacionais que, direta ou indiretamente, podem provocar perdas de receitas e/ou insatisfação dos clientes (Sousa, S. [et al.], 2017).

A seguinte esquematização sintetiza o processo de gestão comercial.



Fig. 2. 11 - Fases do processo de gestão comercial (Sousa, S. [et al.], 2017)

O processo de gestão comercial engloba todas as atividades que admitem a gestão contratual entre a EG e os seus clientes, incluindo os procedimentos operacionais dos quais contam a faturação dos consumos e serviços, a assistência comercial e técnica, a gestão de reclamações e a gestão de toda a informação acerca do cliente, local de consumo e contador. Numa perspetiva mais operacional, abrange as seguintes atividades (Sousa, S. [et al.], 2017):

- Leitura, faturação e cobranças;
- Gestão de contratos e novas ligações;
- Instalação e substituição de contadores.

Na gestão do relacionamento com o cliente, tem-se o atendimento, gestão de contactos e resposta a solicitações e reclamações. Quanto à gestão do cadastro das infraestruturas de clientes, incluem-se os ramais, contadores e locais de consumo e na gestão de informação e indicadores de desempenho, incide nos consumos, receitas de perdas aparentes, dívida, etc. (Sousa, S. [et al.], 2017).

É fulcral monitorizar o processo dada a diversidade e complexidade das diferentes tarefas, pois torna-se vulnerável a situações de conflituosidade com os clientes, insatisfação ou perdas de receitas (Sousa, S. [et al.], 2017).

2.3.2. NOVOS REGULAMENTOS: IMPACTO NAS ENTIDADES GESTORAS

2.3.2.1. Regulamento de Relações Comerciais dos Serviços de Águas e Resíduos

Segundo o descrito na alínea c) do artigo 11º dos Estatutos da Entidade Reguladora dos Serviços de Águas e Resíduos, compete à mesma proceder à elaboração e à aprovação de um regulamento em que vigore explicitamente as regras de relacionamento entre as entidades gestoras em alta e em baixa e entre as últimas e os respetivos utilizadores, no que concerne às condições de acesso e contratação do serviço, medição, faturação, pagamento, cobrança, prestação de informação e resolução de litígios, regulamentando os respetivos regimes jurídicos e a proteção dos utilizadores de serviços públicos essenciais.

Como tal, este regulamento de relações comerciais visa clarificar um conjunto de situações que, não obstante previstas na lei, suscitavam dúvidas interpretativas quanto ao seu âmbito de abrangência e aplicação, bem como definir critérios, conceitos, prazos e estabelecer deveres de informação e outras práticas consideradas como fundamentais para garantir uma maior clareza, equidade e uniformidade de procedimentos no âmbito das relações comerciais, com claro benefício para o setor em geral, e para todos os intervenientes na prestação do serviço público de abastecimento de água.

Após o parecer favorável, assim como os comentários apresentados pelo Conselho Consultivo e críticas e contributos recebidos em sede de consulta pública, o Conselho de Administração da ERSAR deliberou a aprovação do presente regulamento, sucedendo-se a respetiva publicação em Diário da República a 4 de setembro de 2018 e vigência de 90 dias depois da promulgação a 3 de dezembro de 2018. O intitulado Regulamento n.º 594/2018 aplicou-se aos contratos existentes à data da sua entrada em vigência, salvaguardando-se os efeitos já produzidos, no entanto os contratos de fornecimento e de recolha de AA e AR já celebrados com os utilizadores municipais devem ser objeto de aditamento no prazo de 1 ano, i.e, até 3 de dezembro de 2019.

Concretiza vários aspetos do relacionamento entre entidades prestadoras de serviços de águas e resíduos e os seus utilizadores. Este diploma apresenta as seguintes principais inovações:

- A EG deve fornecer uma alternativa de disponibilização de água potável, nas situações em que a interrupção do serviço, por razões de exploração, se mantenha por mais de 24 horas (artigo 53.º);
- Proibição de interrupções do serviço em datas que não permitam que o utilizador regularize a situação no dia seguinte, por motivo imputável à EG (art.º 54.º);
- Definição de um prazo máximo de 24h para o restabelecimento do serviço, após regularização da situação que originou a interrupção (art.º 55.º);
- Possibilidade da EG denunciar o contrato caso o utilizador não proceda ao pagamento da dívida no prazo de 2 meses (art.º 79.º);
- Regras sobre o pagamento de créditos a favor do utilizador em resultado de acertos (art.º 99.º);
- Obrigação de fracionamento do pagamento de acertos que resultem num débito de valor superior ao consumo médio mensal do local de consumo (valor a pagar não pode ultrapassar 25% do consumo médio mensal dos últimos 6 meses - art.º 99.º).

2.3.2.2. Projeto de Regulamento Tarifário dos Serviços de Águas

No final de 2018, a ERSAR veio apresentar um projeto de Regulamento Tarifário para os serviços de águas, conforme igualmente estabelecido nos seus Estatutos que esteve em consulta pública até 15 de março do ano corrente, e que a seguir se analisa.

O presente Regulamento Tarifário abrange dois serviços, abastecimento público de água e saneamento de águas residuais urbanas, podendo ainda ser incluídas neste último serviço as atividades de gestão de águas pluviais e de reutilização de águas residuais tratadas. Incorpora soluções que visam promover o aumento da eficiência produtiva e uma maior racionalidade nas decisões de investimento, de modo a que as tarifas reflitam apenas os custos efetivamente necessários à provisão dos serviços (ERSAR, 2018a).

O novo modelo regulatório para efeitos de determinação de tarifas e rendimento tarifários reflete a solução de afetação dos riscos à parte melhor colocada para o gerir, considerando-se, neste caso, não ser adequado que o risco de procura seja imputado à entidade gestora, porque esta tem uma reduzida capacidade de a influenciar, pelo que tenderia a exigir condições que a prazo se revelariam penalizantes para a tarifa e rendimentos tarifários. Todos os restantes riscos correm por conta da entidade gestora, designadamente os desvios de custos da operação não justificados por variações de volumes ou por outros fatores exógenos, os desvios de gastos de investimento e de financiamento (ERSAR, 2018a).

Os objetivos deste regulamento expressam-se na uniformização da formação das tarifas e dos rendimentos tarifários de cada um dos serviços, abastecimento público de água e saneamento de águas residuais urbanas, num cenário de eficiência produtiva, observando o cumprimento de princípios de

cariz económico, social e ambiental, e promover o equilíbrio entre a sustentabilidade da entidade gestora e a acessibilidade económica aos serviços por parte dos seus utilizadores, num quadro de estabilidade regulatória (ERSAR, 2018a).

Segundo o Artigo 13.º do presente regulamento, a estrutura tarifária encontra-se agregada em utilizadores domésticos e não-domésticos, sendo aplicável em cada sistema a tarifa de disponibilidade, devida em função do período de consumo objeto de faturação e expressa em euros por dia; a tarifa variável, devida em função do nível de utilização do serviço durante o período objeto de faturação e expressa em euros por unidade de medida e as tarifas de serviços auxiliares, devidas por cada serviço prestado, e em função da unidade correspondente.

Em conformidade com o Artigo 16.º, as diferenciações tarifárias devem reger-se pelas diretrizes que se seguem:

- Dentro de cada tipologia de utilizador, doméstico e não doméstico, não deve existir diferenciação das tarifas, à exceção das situações elencadas no presente artigo;
- As entidades gestoras podem diferenciar as tarifas em função do período do ano de modo a atender a flutuações elevadas da procura de ordem sazonal ou de escassez de recursos hídricos;
- As entidades gestoras podem diferenciar as tarifas em função do horário de consumo, desde que disponibilizem os instrumentos de medição adequados;
- Os tarifários podem ser diferenciados nas situações de tarifários sociais, nos termos previstos na lei desde que o respetivo financiamento seja suportado pela entidade titular;
- Deve existir um tarifário para famílias numerosas, nos termos do Artigo 17.º;
- No caso de agregação de sistemas, de titularidade estatal ou municipal, a respetiva entidade titular pode definir, com carácter excecional, a aplicação de um período para convergência dos tarifários dos municípios associados, devendo definir os montantes e respetivas regras de recuperação de custos;
- As exceções aos números anteriores, desde que devidamente fundamentadas, devem ser objeto de autorização da entidade reguladora.

Atualmente são vários os tipos de cliente que vigoram na Águas do Porto, tais como:

1. Doméstico
2. Social
3. Empresarial
4. Público
5. Autárquico
6. Próprios

Neste seguimento, o Regulamento Tarifário tem sido alvo de contestação por parte da Câmara do Porto que considera de "totalmente inaceitável" a proposta apresentada a discussão pública pela Entidade Reguladora dos Serviços de Águas e Resíduos, "por ter contida uma visão altamente centralista da gestão de recursos" e "além de representar um desincentivo ao investimento e à promoção da qualidade de vida nas cidades, o regulamento penalizaria as empresas locais mais eficientes, representando um retrocesso na promoção de hábitos de consumo sustentáveis e ecológicos" (Jornal de Notícias, 2019).

2.4. SATISFAÇÃO DO CLIENTE

2.4.1. ESTUDO DE SATISFAÇÃO DOS UTILIZADORES DOS SERVIÇOS DE ÁGUAS E RESÍDUOS (ESUAR 2017)

A ERSAR deliberou a elaboração do “Estudo de satisfação dos utilizadores dos serviços de águas e resíduos” (ESUAR 2017) cumprindo com o estipulado no número 5 do Artigo 10.º do Decreto-Lei n.º 194/2009, de 20 de agosto, que estabelece o regime jurídico dos serviços municipais de abastecimento público de água, de saneamento de águas residuais e de gestão de resíduos urbanos mencionando que “cabe à entidade reguladora medir, avaliar e divulgar os níveis de satisfação dos utilizadores dos serviços”.

O estudo em questão pretendeu aferir a satisfação dos utilizadores finais dos serviços de águas e resíduos em Portugal continental para o ano de 2017, cujos objetivos específicos para os setores foram os seguintes:

- Satisfação global;
- Satisfação das expetativas;
- Satisfação específica (qualidade da água; pressão da água e frequência de interrupções no abastecimento de água; rapidez na resolução de interrupções/avarias do abastecimento de água e frequência de situações em que se vê esgoto escorrer na via pública e rapidez de intervenção);
- Perceção do abastecimento de água (facilidade de obter informação da qualidade da água e aviso da entidade antes de proceder ao corte de água) e do saneamento de águas residuais;
- Preço a percebido;
- Valor a percebido;
- Imagem;
- Serviço de apoio ao cliente;
- Perceção do serviço de apoio ao cliente;
- Reclamações ou sugestões;
- Confiança.

No enquadramento do tema desta dissertação, apenas alguns tópicos são relevantes, nomeadamente, a satisfação global, a perceção do serviço de apoio ao cliente e as reclamações ou sugestões.

Numa primeira mão, observa-se a Figura 2.12 que demonstra a distribuição relativa das respostas facultadas acerca da satisfação global dos utilizadores finais.

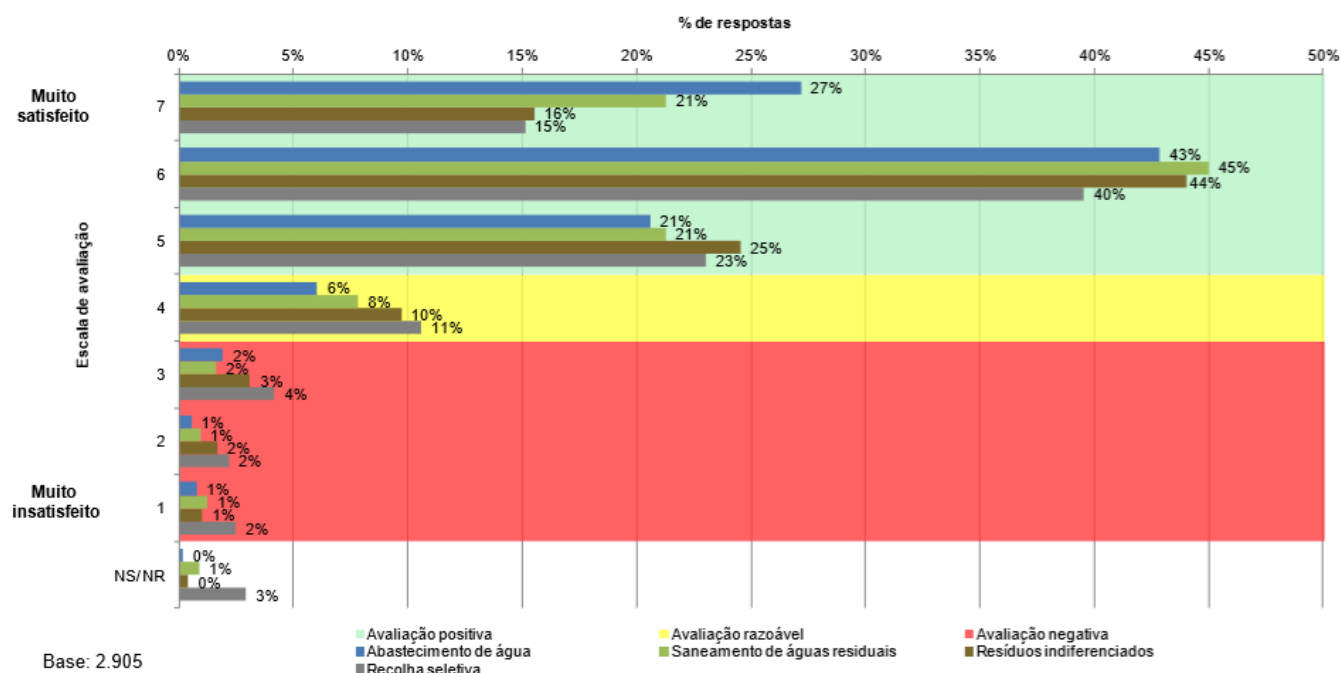


Fig. 2. 12 - Distribuição relativa das respostas facultadas acerca da satisfação global dos utilizadores finais dos serviços de águas e resíduos (ERSAR, 2017)

No cômputo geral, os utilizadores finais estão satisfeitos com o serviço que lhes é prestado no âmbito do abastecimento público de água, do saneamento de águas residuais, da gestão de resíduos urbanos indiferenciados e da recolha seletiva de resíduos.

Em todos os serviços em causa, as avaliações positivas (classificações iguais ou superiores a 5) afetam, pelo menos 78% dos utilizadores, variando entre um mínimo de 78%, para o serviço de recolha seletiva de resíduos, e um máximo de 91%, para o serviço de abastecimento de água. De notar ainda, os 27% de utilizadores que, para o serviço de abastecimento de água, e os 21% de utilizadores que, para o serviço de saneamento de águas residuais, escolhem o grau mais positivo (de valor 7 na escala de avaliação) da escala utilizada. Em termos médios, é sobre o serviço de abastecimento de água que se observa o maior nível de satisfação (ERSAR, 2017).

No que toca à interface com o consumidor, a metodologia aplicada nestes inquéritos permitiu constatar o seguinte:

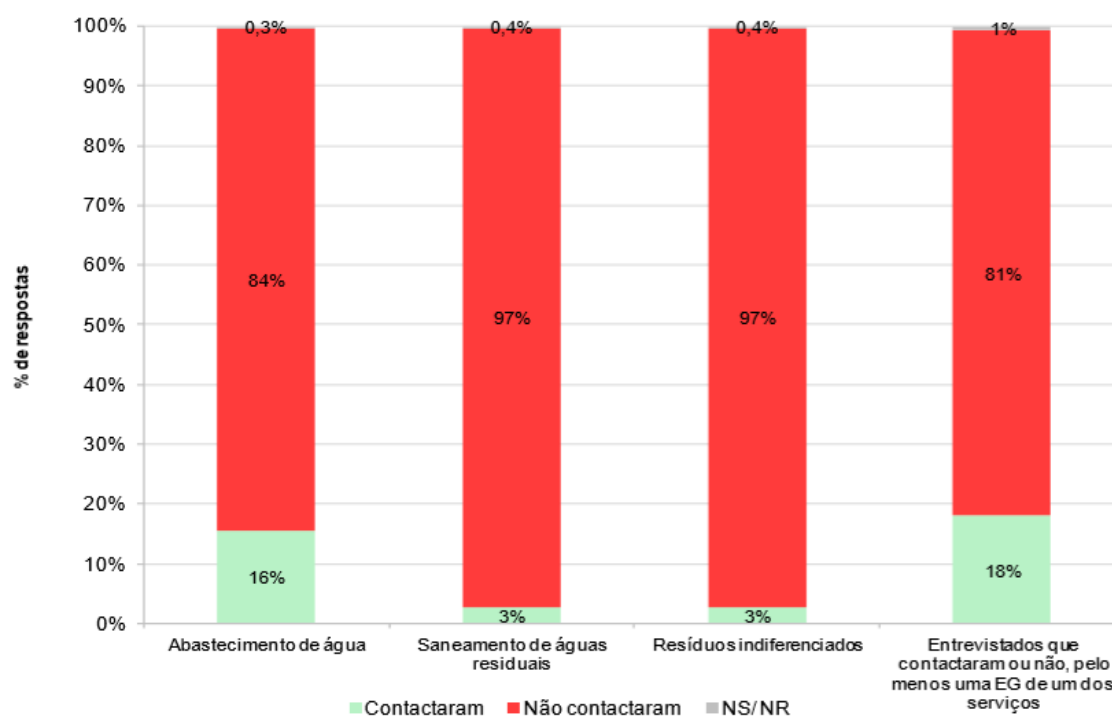


Fig. 2. 13 - Percentagem de inquiridos que contactaram as suas entidades gestoras de serviços de águas e resíduos no último ano (ERSAR, 2017)

No decurso do último ano, 18% dos utilizadores finais contactaram, pelo menos, uma das entidades gestoras dos serviços de águas e resíduos. Os serviços de abastecimento de água foram os mais contactados, com 16% do total de utilizadores a terem-no feito no último ano. Os serviços de saneamento de águas residuais foram alvo de um volume muito reduzido de contactos, envolvendo, em ambos os casos, 3% do total de inquiridos (ERSAR, 2017).

Na figura seguinte constam os motivos que conduziram os inquiridos a contactar a EG.

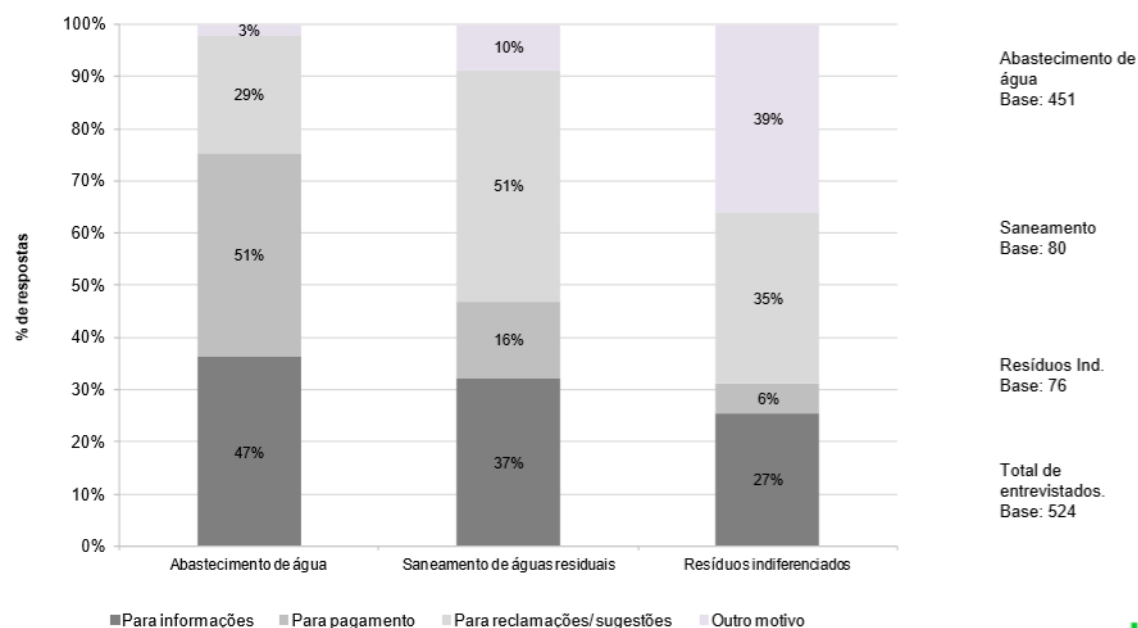


Fig. 2. 14 - Principais motivos pelos quais os inquiridos contactaram a sua entidade gestora de serviços de águas e resíduos (ERSAR, 2017)

No âmbito dos utilizadores que, no último ano, contactaram o serviço de abastecimento de água, as motivações mais referidas tiveram a ver ou com a necessidade de realização de pagamentos (51%) ou com a necessidade de obtenção de algum tipo de informação (47%). 29% do conjunto de utilizadores que contactaram o serviço fizeram-no para apresentação de uma reclamação/sugestão.

Quanto aos utilizadores que contactaram com os serviços de saneamento de águas residuais, a apresentação de uma reclamação/sugestão (51%) foi a motivação prevalecente. Num segundo plano de intensidade, situam-se os contactos para obtenção de informações (37%). O contacto para realização de pagamentos foi referido por 16% deste conjunto de utilizadores. Tal pode ser observado na Figura 2.15.

Relativamente ao serviço de apoio ao cliente, a distribuição da sua avaliação pode ser observada na Figura 2.15.

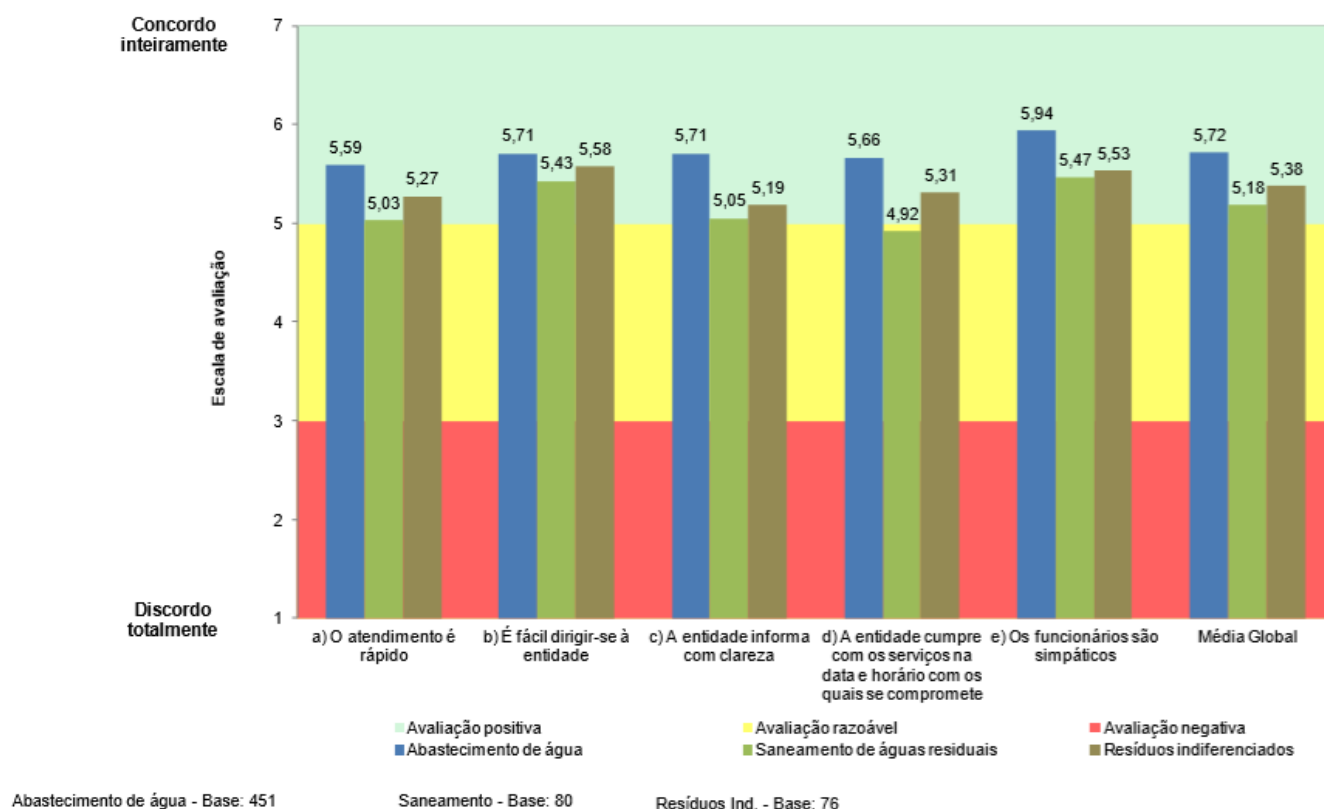


Fig. 2. 15 - Médias de respostas dadas no âmbito da avaliação do serviço de apoio ao cliente (ERSAR, 2017)

No contexto particular dos inquiridos habilitados a responder a este conjunto de questões para cada um dos serviços referenciados, há uma avaliação positiva de todos eles, com vantagem para o serviço de abastecimento de água: Serviço de abastecimento de água, com média global de 5,71 e o Serviço de saneamento de águas residuais, com média global de 5,18 (ERSAR, 2017).

O serviço de abastecimento de água obtém avaliações positivas em todas as dimensões consideradas, com particular destaque para a simpatia dos funcionários (média de 5,94), no caso do serviço de abastecimento de água e a facilidade em se dirigir à entidade (média de 5,58) (ERSAR, 2017).

Quanto ao modo de resolução das reclamações ou sugestões, a Figura 2.16 traduz o panorama das respostas facultadas pelos utilizadores dos serviços neste âmbito.

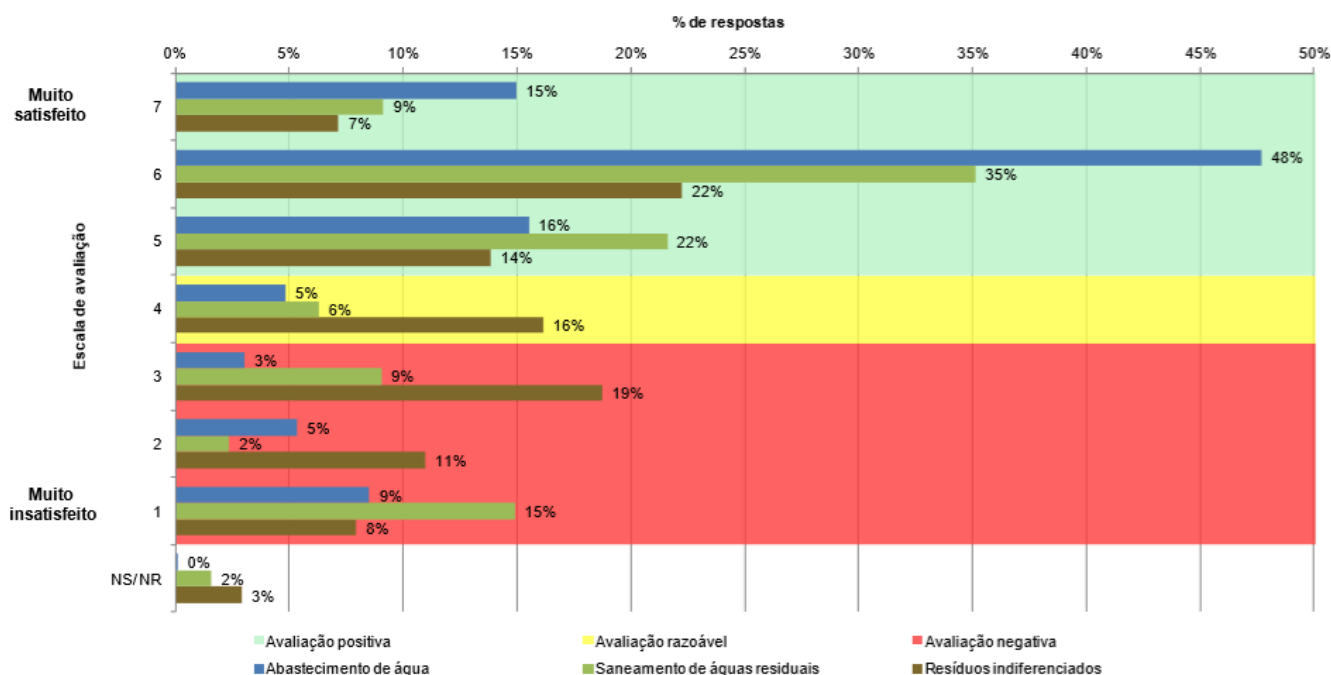


Fig. 2. 16 - Distribuição relativa das respostas dadas no âmbito de como as reclamações ou sugestões foram resolvidas por parte da respetiva entidade gestora (ERSAR, 2017)

Sobre o serviço de abastecimento de água, os inquiridos habilitados a pronunciar-se nesta matéria, avaliam positivamente a forma como a entidade resolveu a sua reclamação/sugestão: 78% de avaliações positivas; média de 5,17. No caso do outro serviço, a avaliação é perspectivada apenas como razoável com 66% de avaliações positivas versus 26% de avaliações negativas e média de 4,62 (ERSAR, 2017).

De uma forma geral, os utilizadores finais estão satisfeitos com os serviços de abastecimento de água e de saneamento de águas residuais. Uma satisfação tanto evidenciada em termos de avaliação global dos serviços, como em termos da contextualização dessa mesma avaliação para o plano das expectativas existentes. Do total dos utilizadores finais, 18% contactaram, no decurso do último ano, pelo menos uma das entidades gestoras dos serviços, mas com claro predomínio do contacto com o serviço de abastecimento de água (ERSAR, 2017).

2.4.2. GESTÃO DE RECLAMAÇÕES ERSAR

A ERSAR aprecia e supervisiona as reclamações apresentadas pelos utilizadores dos serviços de águas e resíduos que lhe sejam remetidas, com o objetivo de avaliar a forma como são respondidas e tratadas pelas entidades gestoras. A ERSAR pronuncia-se sobre as reclamações analisadas emitindo recomendações que não têm carácter vinculativo (ERSAR, 2019b).

Até 2005 a ERSAR, na altura ainda designada por Instituto Regulador das Águas e Resíduos (IRAR), recebia uma média de 45 reclamações por ano, situação que se alterou significativamente em 2006, atendendo a ter sido criada a obrigação de as empresas concessionárias e municipais, responsáveis pelo serviço de abastecimento de água, passarem a dispor de um livro de reclamações e de enviarem cópias das mesmas à ERSAR (Decreto-Lei n.º 156/2005, de 15 de setembro). A partir de 2008 todas as entidades gestoras dos serviços de águas e resíduos, independentemente do modelo de gestão, passaram a estar sujeitas à obrigação de dispor do livro de reclamações, o que provocou novo aumento

significativo das reclamações geridas pela ERSAR (Decreto-Lei n.º 371/2007, de 6 de novembro, que alterou o Decreto-Lei acima identificado) (ERSAR, 2019b).

No gráfico abaixo é possível constatar a evolução das reclamações recebidas na ERSAR, desde 2007, fragmentadas em função do respetivo tema.

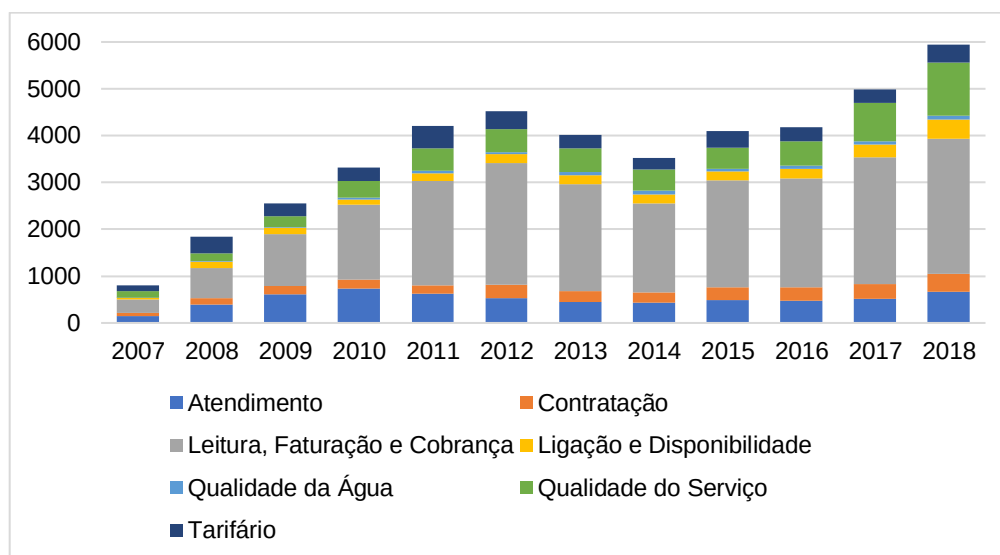


Fig. 2. 17 - Evolução das reclamações rececionadas na entidade reguladora desde 2007

Nota-se uma maior preponderância nos assuntos relacionados com a leitura, faturação e cobrança, seguindo-se a qualidade do serviço. Esta tendência manteve-se praticamente inalterada ao longo dos anos.

A nível local, a projeção das reclamações originárias da Águas do Porto desde 2017 foi a seguinte:

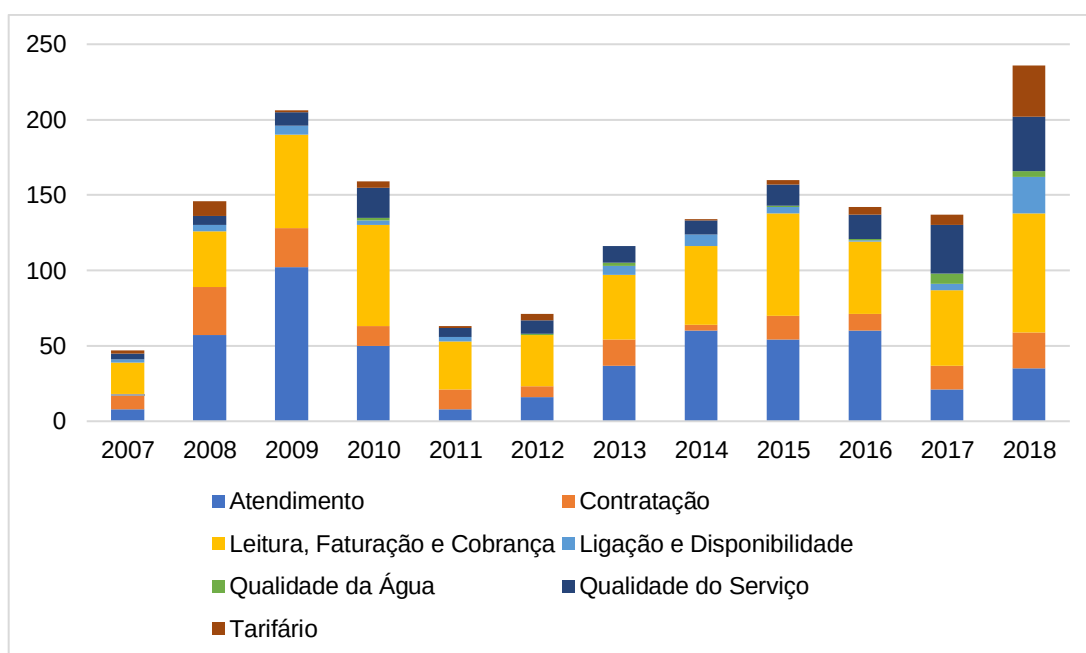


Fig. 2. 18 - Progresso das reclamações para a Águas do Porto

À semelhança da observação anterior, o cenário repete-se nesta EG, contudo a categoria de leitura, faturação e cobrança tem vindo a perder algum “terreno” para outros assuntos, como o atendimento, tarifário, qualidade de serviço e contratação que representam valores praticamente equiparados e a incrementar significativamente, com exceção do atendimento que vai oscilando durante os vários anos.

Na Águas do Porto, as reclamações apresentadas pelos clientes são encaradas como uma oportunidade de melhoria dos serviços prestados, de modo a prevenir as situações que lhes deram origem, e que traduzam as suas expectativas, preocupações e interesses. Nesta perspetiva, a análise de reclamações e a interligação com o cliente para a sua resolução e/ou elucidação é uma constante preocupação dos diferentes intervenientes internos (Águas do Porto, 2018).

2.5. SISTEMAS DE INFORMAÇÃO GEOGRÁFICA

Um SIG pode ser definido como uma plataforma de hardware e software que grande capacidade de armazenamento e compilação de informação. Apesar de ser utilizado com grande relevância a partir da década de 90, após haver uma evolução ao nível de hardware que permitiu uma adaptação as necessidades exigidas, o SIG é utilizado desde a década de 60 (Cunha, S., 2009).

Um SIG dá a possibilidade de visualizar, questionar, analisar e interpretar dados para compreender relações, padrões e tendências nos mais diversos setores de atividade independentemente da dimensão das organizações. Sendo assim, o SIG tem potencial de ajudar as empresas a reduzir custos através da eficiência, melhorar a tomada de decisão, ajudar na compreensão e relato de situações e ocorrências ou numa gestão geográfica mais eficiente. O funcionamento do SIG é um processo simples que se pode dividir em 5 passos: perguntar, adquirir, examinar, analisar e agir (ESRI, s.d.).

Assim sendo, as componentes essenciais para o SIG são os dados especiais, recursos humanos, *hardware*, *software*. Os dados especiais podem ser de variados tipos como por exemplo, fotografias aéreas, levantamentos topográficos entre outros sendo estes destinados a agregação e integração georreferenciada de várias fontes. Os recursos humanos são uma componente fundamental para o processo desde o seu início (fotografias aéreas, levantamentos topográficos, etc.) até ao seu final (tratamento dos dados, análises e apresentações de resultados). O *hardware* engloba a CPU, o armazenamento, dispositivos de entrada e de saída de informação. O *software* trata do processamento da informação especial, englobando a gestão da base geográfica e da base alfanumérica (Cunha, S., 2009).

A Figura 2.19 mostra a aplicação de diferentes tipos de informação geográfica: as linhas representam as condutas, os polígonos representam os edifícios, os pontos estão relacionados com as válvulas e a informação do objeto é dada pela imagem de satélite.

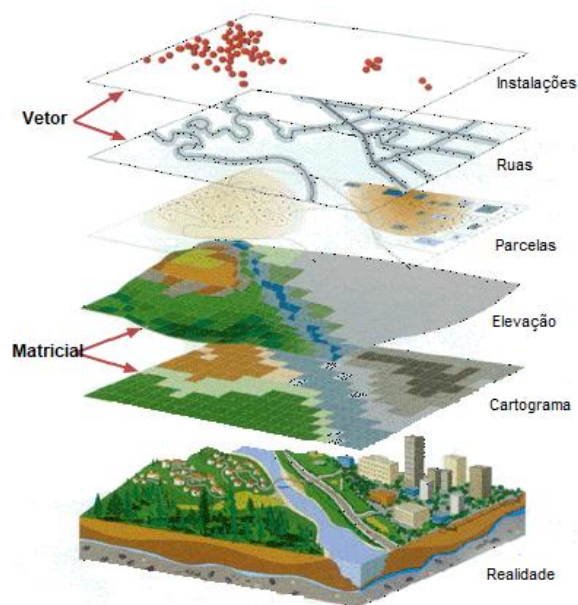


Fig. 2. 19 - Representação de diferentes tipos de informação geográfica nos SIG (Adaptado de (Creekside Center for Earth Observation, 2012))

A observação da figura anterior permite concluir que os SIG têm inúmeras potencialidades para os SAA. Os dados contidos num SIG podem ser vinculados a uma localização geográfica, como por exemplo os contadores de água de cada cliente e as condutas de distribuição de uma determinada zona (Teixeira, J., 2014).

Na sua forma mais simples, os dados matriciais ou *rasters* consistem num conjunto de células organizadas em linhas e colunas, em que cada célula contém um valor que representa a informação, por exemplo, a temperatura. São fotografias digitais aéreas, imagens de satélite ou até mesmo mapas digitalizados. Os dados matriciais adequam-se bem para representar dados que mudam continuamente ao longo de uma superfície. A elevação em relação à superfície da Terra é uma das aplicações mais frequentes dos dados matriciais, no entanto podem ser representadas outras características tais como a precipitação, temperatura, concentração e densidade populacional (ESRI, 2016b).

A figura que se segue ilustra as características gerais dos dados matriciais.

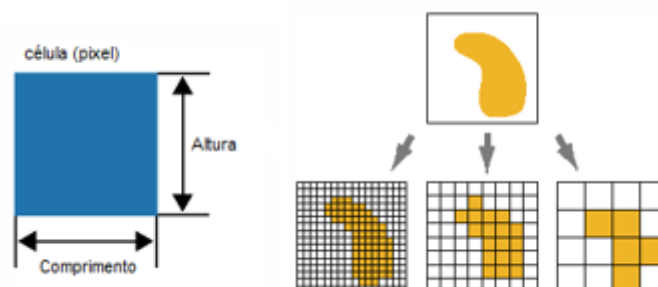


Fig. 2. 20 - Características gerais dos dados matriciais (Adaptado de ESRI, 2016b)

Embora a estrutura dos dados matriciais seja simples, é excepcionalmente útil para uma ampla variedade de aplicativos. Num GIS, o uso de dados *raster* pode aplicar-se principalmente a *rasters* como mapas

base, um objeto do estudo feito no capítulo 5. Um uso comum de dados *raster* em GIS é como uma exibição em segundo plano para outras camadas de recursos. Por exemplo, as ortofotografias ou ortofotomapas das quais foram removidas as distorções causadas pela inclinação da câmara e pelo relevo, encontram-se por baixo de outras camadas que fornecem ao usuário do mapa a confiança de que as camadas do mapa estão espacialmente alinhadas e representam objetos reais, além de informações adicionais. As três principais fontes de mapas de base *raster* são ortofotos de fotografias aéreas, imagens de satélite e mapas digitalizados, consoante a Figura 2.21 (ESRI, 2016b, Sinfic, s.d.).

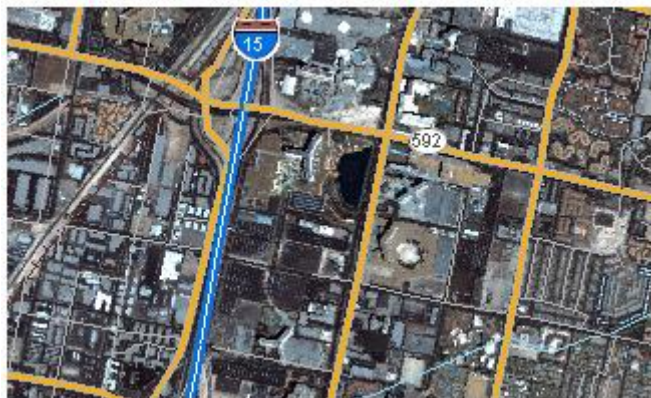


Fig. 2. 21 - Raster empregue como mapa base (ESRI, 2016b)

Cada célula tem a mesma largura e altura e é uma porção de toda a superfície representada pela matriz. Para uma melhor compreensão, considere-se o exemplo de uma matriz que represente uma área de 100 km², constituída por 100 células na qual cada uma destas teria, portanto, que ter 1 km² de área, com igual largura e altura (ou seja, 1 km × 1 km) (ESRI, 2016b).

O *software ArcGis* traduz-se na combinação de aplicações de geoprocessamento. O *ArcMap* representa informações geográficas como um conjunto de camadas e outros elementos num mapa. Os elementos comuns do mapa incluem, a tabela de dados que contém camadas do mapa para uma determinada extensão juntamente com uma barra de escala, seta a apontar para norte, título, texto descritivo, uma legenda de símbolo e assim em diante (ESRI, 2016a).

Das tarefas típicas executadas no *ArcMap*, a que mais se enquadra o estudo é a manipulação de mapas, onde se pode abrir e usar documentos para explorar informações, navegar pelos documentos do mapa, colocar e retirar camadas, consultar recursos para aceder aos dados de atributos avançados que estão por trás do mapa e para visualizar informações geográficas. Também se pode utilizar o geoprocessamento para automatizar o trabalho, efetuar análises e tem a capacidade de executar qualquer modelo ou script de geoprocessamento, além de visualizar e trabalhar com os resultados através da visualização de mapas. O geoprocessamento pode ser usado para análise e para automatizar muitas tarefas comuns, como a geração de livros de mapas, reparando links de dados corrompidos numa coleção de documentos de mapas e para executar o processamento de dados GIS (ESRI, 2016a).

2.6. APLICABILIDADE DA INTELIGÊNCIA ARTIFICIAL NA COMUNICAÇÃO COM O CLIENTE

A inteligência artificial, os algoritmos e a aprendizagem automática das máquinas estão a revolucionar a sociedade. O termo “Inteligência Artificial” encontra-se presente em diversas aplicações do dia a dia: da *Google Assistant* ao algoritmo do motor de busca da *Google*, dos carros autónomos aos serviços digitais de apoio ao cliente (os *chatbots*, cada vez mais exímios em processar a linguagem natural). Estes ainda não são IA no seu máximo potencial (capaz de replicar o raciocínio humano ou ir além disso),

mas estas aplicações na sua maioria baseadas no processamento de *big data* e *Machine Learning*, dão passos firmes para lá chegar um dia (Figueiredo, M., 2018).

Atualmente já existem empresas a recorrer à IA para descobrir padrões e identificar oportunidades, com benefício imediato no relacionamento com os clientes. A previsão é que até 2020, 85% das interações com clientes sejam efetuadas através de inteligência artificial e as organizações esperam contemplar um aumento de 39% na receita média em 2020 e uma redução de 37% nos custos (Salesforce, s.d.-b).

A IA promove uma nova experiência com os clientes através da oferta de recursos para prever o fecho de oportunidades e exceder as expectativas dos clientes, bem como a automatização de tarefas manuais do dia-a-dia que ocupam o tempo que a equipa poderia investir em estratégias de melhoramento da experiência dos clientes. No campo do atendimento, contribui para a ajuda na compreensão das necessidades dos clientes e torna a equipa de colaboradores ainda mais proativa ao resolver casos antes deles acontecerem. Quanto ao *marketing*, possibilita a criação de uma jornada preditiva com *insights* sobre os desejos e comportamento de cada cliente, tal como o acréscimo da inteligência à jornada de cada cliente para personalizar experiências como nunca, prevê o público ideal, conteúdo e tempo de envio para qualquer campanha e recomenda ainda especialistas, artigos e tópicos em comunidades. Nas plataformas, permite o treino com dados inseridos no CRM, potencializa o poder e a assertividade das recomendações com o passar do tempo, descobre *insights* mais profundos para impulsionar a inovação e criar aplicativos mais inteligentes, otimiza as regras de IoT (do inglês, *Internet of Things*), que abrange todos os aparelhos e objetos que se encontram habilitados a estarem permanentemente ligados à Internet, sendo capazes de se identificar na rede e de comunicar entre si, de forma automatizada (Salesforce, s.d.-a) (CNCS, s.d.).

A inteligência artificial é um conceito amplo que inclui a aprendizagem de máquina como um dos seus recursos que se passa a explicar no tópico seguinte.

2.6.1. APRENDIZAGEM DE MÁQUINA

A aprendizagem automática ou aprendizagem de máquina, traduzida do inglês *Machine Learning*, é um conceito que se encontra associado à Inteligência Artificial com várias definições possíveis. Trata-se de um sistema que pode modificar o seu comportamento autonomamente tendo como base a sua própria experiência, ou seja, a partir da realização do treino de um modelo onde a interferência humana é mínima. São estabelecidas regras lógicas que visam melhorar o desempenho de uma tarefa ou, dependendo da aplicação, tomar a decisão mais apropriada para o contexto do problema. Essas regras são geradas com base no reconhecimento de padrões dentro dos dados analisados (Alecrim, E., 2018).

Por exemplo, um motor de pesquisa precisa de analisar uma série de parâmetros com base numa quantidade significativa de dados para decidir se exhibe resultados equivalentes para o assunto que se pretende procurar, algo que as ferramentas de busca têm de sobra dados os milhões de acessos que recebem e que, consequentemente, servem de treino. Outro aspeto ilustrado é a constante entrada de dados que favorece a identificação de novos padrões que com ML, o mecanismo de pesquisa conseguirá identificar padrões que apontam para o novo significado do termo e, depois de algum tempo, estará apto a considerá-lo nos resultados das buscas (Alecrim, E., 2018).

São várias as abordagens para o *Machine Learning*, das quais se destacam a Aprendizagem Supervisionada (*Supervised Learning*), a Aprendizagem Não Supervisionada (*Unsupervised Learning*) e a Aprendizagem Aprofundada (*Deep Learning*).

2.6.1.1 Aprendizagem Supervisionada

A grande maioria da prática de *Machine Learning* recorre à aprendizagem supervisionada (em inglês, *Supervised Learning*). Aqui dispõe-se das variáveis de entrada x (*input*) e uma variável de saída Y (*output*), usadas no algoritmo para aprender a função de mapeamento da entrada para a saída, cuja expressão é (Brownlee, J., 2016):

$$Y = f(X)$$

O objetivo passa por aproximar a função de mapeamento tão bem que quando se tem novos dados de entrada (x) podem ser previstas as variáveis de saída (Y) para esses dados. É denominada de aprendizagem supervisionada, porque o processo de um algoritmo aprende com o conjunto de dados de treino pode ser idealizado como um “professor” a supervisionar o processo de aprendizagem. Sabe-se as respostas corretas, o algoritmo iterativamente faz previsões sobre os dados de treino e é corrigido pelo “professor”. A aprendizagem pára quando o algoritmo atinge um nível aceitável de desempenho (Brownlee, J., 2016).

Os problemas de aprendizagem supervisionada podem ser agrupados em problemas de regressão e classificação (Brownlee, J., 2016).

- Classificação: Um problema de classificação surge quando a variável de saída é uma categoria, por exemplo “vermelho” ou “azul” ou “doente” e “não doente”;
- Regressão: Aplica-se quando a variável de saída é um valor real, tal como “dólares” ou “peso”.

Alguns tipos comuns de problemas construídos no topo da classificação e regressão incluem a recomendação e previsão de séries temporais, respetivamente.

Alguns exemplos populares de algoritmos de aprendizagem de máquina supervisionados são (Brownlee, J., 2016):

- Regressão linear para problemas de regressão;
- “Floresta aleatória” (*Random Forest*) para problemas de classificação e regressão;
- Máquinas de vetores de suporte (Support Vector Machines, SVM) para problemas de classificação.

2.6.1.2 Aprendizagem Não Supervisionada

Algoritmos de aprendizagem de máquina não supervisionados inferem padrões de um conjunto de dados sem referência a resultados conhecidos ou rotulados. Diferentemente da aprendizagem de máquina supervisionado, os métodos não supervisionados de aprendizagem de máquina não podem ser aplicados diretamente a uma regressão ou a um problema de classificação, porque desconhece-se quais podem ser os valores para os dados de saída, impossibilitando o treino do algoritmo da maneira usual, e pode ser empregado para descobrir a estrutura subjacente dos dados (DataRobot, 2019)

O objetivo é descobrir padrões de dados anteriormente desconhecidos, mas na maioria das vezes esses padrões são aproximações fracas do que realmente pode alcançar. Além disso, não se sabe quais devem ser os resultados, não há como determinar a precisão deles, tornando-o mais aplicável a problemas do mundo real. O melhor momento para usar o modo não supervisionado é quando não se possui dados sobre os resultados desejados (DataRobot, 2019)

Algumas aplicações de técnicas não supervisionadas incluem (DataRobot, 2019):

- **Clustering:** Permite dividir automaticamente o conjunto de dados em grupos, de acordo com a similaridade. Muitas vezes, no entanto, a análise de *cluster* subestima a semelhança entre grupos

e não trata os pontos de dados como indivíduos. Por esse motivo, a análise de *cluster* é uma má opção para aplicativos como segmentação e segmentação de clientes;

- **Deteção de anomalias:** Pode descobrir automaticamente pontos de dados incomuns no conjunto de dados. Isso é útil para identificar transações fraudulentas, descobrir peças defeituosas de hardware ou identificar um erro externo causado por um erro humano durante a entrada de dados;
- **Mineração por associação:** Identifica conjuntos de itens que ocorrem frequentemente juntos no seu conjunto de dados. No comércio, costuma ser usado para análise de compras, pois permite que os analistas descubram mercadorias frequentemente adquiridas ao mesmo tempo e desenvolvem estratégias mais eficazes de marketing.
- **Modelos de variáveis latentes:** São habitualmente usados em pré-processamento de dados, como reduzir o número de recursos num conjunto de dados (redução de dimensionalidade) ou decompor o conjunto de dados em vários componentes

Os padrões patenteados com métodos de aprendizado de máquina não supervisionados também podem ser úteis ao implementar métodos de aprendizagem de máquina supervisionada posteriormente. Por exemplo, é possível aplicar uma técnica não supervisionada para executar a análise de *cluster* nos dados e, em seguida, usar o *cluster* ao qual cada linha pertence como um recurso extra no modelo supervisionado. Outro exemplo é um modelo de deteção de fraude que utiliza pontuações de deteção de anomalias como um recurso extra (DataRobot, 2019)

2.6.1.2 Aprendizagem Aprofundada

Também designada por *Deep Learning*, onde as grandes quantidades de dados são tratadas a partir de várias camadas de redes neurais artificiais (algoritmos inspirados na estrutura de neurónios do cérebro) que resolvem problemas muito complexos, conforme ilustrado na Figura 2.22 (Alecrim, E., 2018).

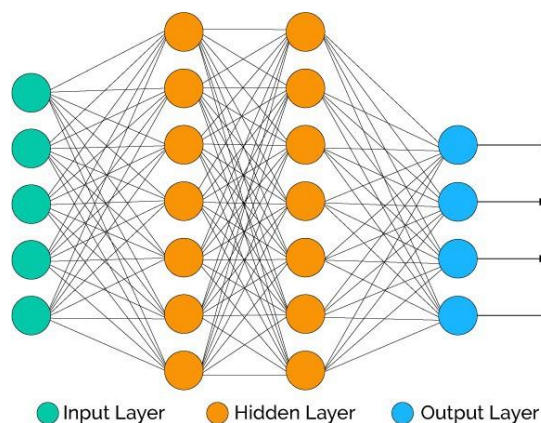


Fig. 2. 22 - Exemplificação de uma rede neuronal (Bhatia, R., 2018)

É preciso muito poder computacional para resolver problemas de aprendizagem aprofundada devido à natureza iterativa de seus algoritmos, a sua complexidade que cresce conforme o número de camadas aumenta e os grandes volumes de dados necessários para treinar as redes. A natureza dinâmica dos métodos, nomeadamente, a sua capacidade de continuamente melhorar e de se adaptar a mudanças no padrão de informações subjacentes, apresenta uma grande oportunidade de introduzir mais comportamentos dinâmicos aquando da análise de dados (SAS, 2018).

Esta componente deixa de transmitir ao computador como resolver o problema e passa a treinar o computador para resolver o problema sozinho. Uma abordagem tradicional para a análise de dados consiste na utilização de dados para a criação de novas variáveis, depois seleciona-se um modelo analítico e, finalmente, estimam-se os parâmetros (ou as incógnitas) desse modelo. Essas técnicas podem produzir sistemas preditivos que não difundem bem, uma vez que a integridade e a exatidão dependem da qualidade do modelo e de suas características. Por exemplo, se se desenvolver um modelo para prevenir fraudes, deve-se começar por um conjunto de variáveis que, provavelmente, derivarão um modelo usando transformações de dados em que o modelo pode chegar a depender de 30.000 variáveis, por isso deve ser moldado, descobrir quais variáveis são significativas e quais não são, e assim por diante. Porém, quando se adicionam mais dados exige é necessário efetuar tudo de novo (SAS, 2018).

A abordagem com *Deep Learning* compreende a substituição da formulação e a especificação do modelo por caracterizações hierárquicas (ou camadas) que aprendem a reconhecer as características latentes dos dados nas regularidades em camadas. A promessa deste tipo de aprendizagem passa por concretizar sistemas preditivos que se difundem e se adaptam bem, melhoram continuamente à medida que novos dados vão sendo adicionados e são mais dinâmicos do que sistemas preditivos baseados em regras de negócios. Aqui não se adapta um modelo, mas sim treiná-lo (SAS, 2018)

Para quem olha de fora, pode parecer um projeto em fase experimental enquanto investigadores de ciência da computação e cientistas de dados continuam a testar suas capacidades. Entretanto, possui muitas aplicações práticas que já são postas em práticas pelas empresas, e outras muitas que serão usadas com o desenvolvimento de novas abordagens. Exemplos de aplicações que se observam nos dias de hoje (SAS, 2018):

- **Reconhecimento de fala** nas plataformas *Xbox*, *Skype*, *Google Now* e a *Siri*®, que empregam tecnologias de *Deep Learning* nos seus sistemas para reconhecer padrões de fala e voz;
- **Reconhecimento de imagem**, que pode ser crucial em investigações criminais para identificar atividades transgressoras no meio de milhares de fotos enviadas por quem estiver presente em locais onde tenha ocorrido um crime. Os carros autônomos também serão beneficiados com reconhecimento de imagem através do uso de câmaras com tecnologia a 360°;
- **Processamento de linguagem natural** através de redes neurais, um componente central da aprendizagem aprofundada que tem sido usado para analisar e processar textos escritos durante anos. Com a mineração de texto (*text mining*) é exequível a descoberta de padrões em reclamações de clientes, relatórios médicos ou boletins informativos, entre outros;
- **Sistemas de recomendação** como a *Amazon* e a *Netflix*, que tornaram popular o conceito de sistemas de recomendação baseando-se em comportamentos anteriores. Também pode ser utilizado para aprimorar as recomendações em ambientes complexos, como músicas ou preferências de roupas em múltiplas plataformas.

2.6.2. PROGRAMAÇÃO EM PYTHON

Python consiste num código aberto, interpretado e fornece uma ótima abordagem para programação orientada a objetos. É uma linguagem popular de alto nível frequentemente utilizada em projetos de ciência de dados, destacando-se nas áreas de matemática, estatística e função científica com vastas bibliotecas de dados. A versatilidade e a sintaxe simples justificam a ampla adesão pelas comunidades científica e de pesquisa, sendo apto para pessoas sem conhecimentos de engenharia e para célere prototipagem (Codecademy, 2019).

Nas áreas de aplicação, os cientistas de ML mostram-se mais recetíveis ao *Python*, onde o NLP e a análise de sentimentos se destacam pela sua robustez dada a grande coleção de bibliotecas que

contribuem para a resolução de problemas complexos de negócios, construção de sistemas maciços e aplicativos de dados (Python Software Foundation, 2001).

Características desta linguagem (Singh, M., 2016):

- Sintaxe elegante, o que se reflete na clara leitura do programa;
- Acessibilidade;
- Uma imensa biblioteca padrão e comunidades de suporte;
- Modo interativo que torna simples o teste de códigos;
- Extensão de códigos através da junção de novos módulos praticável noutra linguagem compilada como C++ ou C;
- Linguagem expressiva possível de incorporar em aplicações, de modo a obter uma interface programável;
- Execução do código em qualquer sistema, inclusive Windows, Mac Os X, etc.;
- Software livre em algumas categorias, sem custos adicionais para usar ou descarregar.

As bibliotecas frequentemente usadas no âmbito da ciência de dados são (GeeksforGeeks, s.d.):

- **Numpy:** *Numpy*, ou *Python* Numérico, é uma biblioteca *Python* que fornece funções matemáticas para lidar com uma matriz de grandes dimensões. Ele fornece vários métodos/funções para matriz, métricas e álgebra linear. Apresenta muitos recursos úteis para operações em n-matrizes e matrizes em *Python*, bem como vetorização de operações matemáticas do tipo de matriz *NumPy*, que aprimoram o desempenho e aceleram a execução, sendo muito fácil trabalhar com grandes matrizes e matrizes multidimensionais usando o *NumPy*.
- **Pandas:** *Pandas* é uma das bibliotecas *Python* mais populares para manipulação e análise de dados. O *Pandas* fornece funções úteis para manipular grande quantidade de dados estruturados. São o método mais fácil de realizar análises, permite grandes estruturas de dados, manipulação de tabelas numéricas e dados de séries temporais. Existem duas estruturas de dados no *Pandas* (*Series* - manipula e armazena em dados unidimensionais e *DataFrame* – manipula e armazena em dados bidimensionais)
- **Matplotlib:** O *Matplotlib* é outra biblioteca *Python* útil para visualização de dados. A análise descritiva e a visualização de dados são muito importantes para qualquer organização. Permite criar rapidamente gráficos de linhas, gráficos circulares, histogramas e outras figuras de nível profissional. Utilizando esta biblioteca é possível personalizar todos os aspetos de uma figura e possui recursos interativos, como zoom, planeamento e salvar gráfico em formato gráfico.
- **Scipy:** É popular para ciência de dados e computação científica. O *Scipy* oferece uma excelente funcionalidade para matemática científica e programação de computação. Contém submódulos para otimização, álgebra linear, integração, interpolação, funções especiais, FFT, processamento de sinais e imagens, solucionadores de ODE, *Statmodel* e outras tarefas comuns em ciência e engenharia.
- **Scikit-learn:** *Sklearn* é uma biblioteca *Python* para *Machine Learning* com vários algoritmos e funções que são usadas. Baseia-se no *NumPy*, *SciPy* e *matplotlib*, aprovisionando ferramentas fáceis e simples para mineração e análise de dados e um conjunto de algoritmos comuns de ML aos usuários por meio de uma interface consistente. Ajuda a implementar rapidamente algoritmos populares em conjuntos de dados e a resolver problemas do mundo real.

2.7. CONCLUSÃO DO ESTADO DA ARTE

A consulta de várias fontes durante a pesquisa bibliográfica permitiu efetuar o enquadramento desde o setor das águas em Portugal Continental, passando pela área comercial e tecnologias relevantes onde se insere o objeto de estudo desta dissertação.

Avultou-se em primeiro lugar o valor da água na qualidade de vida e no desenvolvimento socioeconómico do país, com um progresso favorável no SAA e SAR atualmente, em grande parte devido ao esforço de infraestruturação e à existência de fundos comunitários. Não obstante desta melhoria, ainda existe um longo trilha a percorrer no sentido de promover uma maior eficiência na utilização da água, tendo por base um plano estratégico com metas e objetivos operacionais em curso até ao próximo ano.

A relação entre EG e cliente ocorre maioritariamente na parte comercial que se ocupa de todas as atividades que decorrem no âmbito da gestão processual, de reclamações, do relacionamento com o cliente e do cadastro das infraestruturas de clientes. Recentemente, foram publicados dois regulamentos que seguramente terão impactos nas EG: o Regulamento de Relações Comerciais dos Serviços de Águas e Resíduos, que clarifica um conjunto de aspetos relativamente à legislação existente, e o Projeto de Regulamento Tarifário, que propõe a uniformização na constituição das tarifas e dos rendimentos tarifários dos serviços de abastecimento público de água e saneamento de águas residuais urbanas com a perspetiva de eficiência produtiva, sendo este último um alvo de contestação por parte da Câmara que se opõe a esta homogeneização que acredita penalizar os consumidores em vários níveis.

A satisfação do cliente é um fator importante na interação com as EG que se estende em diversas áreas, algo consagrado com o ESUAR 2017 que proporcionou resultados interessantes, essencialmente na satisfação global, perceção do serviço de apoio ao cliente e reclamações ou sugestões. A destacar, tem-se a satisfação global a marcar 43% e 45% no valor correspondente a 6 numa escala de avaliação de 0 a 6 para o AA e AR, respetivamente. Quanto à interface com o utilizador final, apenas 16% dos inquiridos contactou a EG por motivos relacionados com AA e 3% em questões de AR. A gestão de reclamações trata-se de um processo que requer análise meticulosa indo de encontro às expetativas do cliente em conformidade com as diretivas da EG, onde se sobressaem os campos da leitura, faturação e cobrança tanto no generalidade do país como no município do Porto.

No que diz respeito à Inteligência Artificial, esta dispõe de várias ferramentas relevantes ao desenvolvimento de um algoritmo adaptado ao panorama das solicitações que chegam via email. As vertentes de *Supervised Learning* e *Deep Learning* mostram-se como as mais apropriadas atendendo às variáveis de entrada presentes e assim obter as variáveis de saída utilizando a linguagem de programação *Python* e as suas bibliotecas necessárias no contexto do problema.

A principal dificuldade encontrada na recolha bibliográfica acerca deste tema foi a lacuna de informação relacionada com a classificação de solicitações e o processo de produção de resposta automática. Isto porque não se encontrou casos de estudo diretamente associados, apenas de categorização de texto ou resposta automática programada, por exemplo em Gmail, enquanto se pretendia para ambiente do *Microsoft Outlook*.

3

ÂMBITO E OBJETIVOS

3.1. ÂMBITO

O presente documento foi desenvolvido no âmbito da Unidade Curricular Dissertação que integra o plano de estudos do Mestrado Integrado em Engenharia do Ambiente, realizando-se em ambiente/colaboração empresarial cuja entidade promotora é a Águas do Porto, EM.

Após o enquadramento bibliográfico sobre a parte comercial e inteligência artificial, em concreto, com respeito à gestão de solicitações, a informação existente é diminuta em consequência da inexistência atual por parte das EG. Contudo, o panorama é outro no caso das reclamações em que são disponibilizados manuais de procedimentos para gestão de reclamações.

Este motivo suscitou o interesse na proposta de diretrizes para a classificação de pedidos provenientes de clientes e na integração do mecanismo de resposta automática num algoritmo classificador e ainda a exploração da geração de texto automático em LSTM. A longo prazo, seriam visíveis os resultados e uma poupança significativa para este indicador.

Pretende-se dar a conhecer com a composição deste estudo, as vantagens decorrentes da projeção de um código capaz de responder às solicitações, assim como a redução do tempo gasto na comunicação com os utilizadores finais.

Empregaram-se as informações desta EG a fim de aprimorar a fiabilidade da metodologia adotada nesta dissertação, nomeadamente, no desenvolvimento do conteúdo do próximo capítulo.

3.2. OBJETIVOS

Esta exposição escrita apresenta como objetivo principal a gestão de solicitações de clientes do município do Porto, aplicando ferramentas no campo da Inteligência Artificial. Os objetivos complementares identificados traduzem-se nos seguintes:

- Classificação de solicitações com base na metodologia praticada para a gestão de reclamações;
- Georreferenciação de solicitações por tipologia;
- Avaliação dos vários algoritmos classificadores de aprendizagem automática através de parâmetros, como a precisão, exatidão e medição F1;
- Preparação e geração da resposta automática;
- Conceção de um código em contexto de aprendizagem profunda com vista à produção de texto automático utilizando redes neuronais (LSTM).

4

CMPEA – EMPRESA DE ÁGUAS DO MUNICÍPIO DO PORTO, EM

4.1. APRESENTAÇÃO DA EMPRESA

Anteriormente designada de Serviços Municipalizados de Água e Saneamento (SMAS) do Porto desde 1927, a Águas do Porto, EM foi instaurada em outubro de 2006 como entidade municipal local munida de autonomia administrativa, financeira e estatutária, cujo acionista é a Câmara Municipal do Porto, sucedendo aos antigos serviços prestados onde se notabilizam novas áreas de atuação com vista à gestão integrada e sustentável de todo o ciclo urbano da água no município. Na Fig.4.1 encontram-se particularizadas as atividades que englobam este ciclo, sobretudo, a drenagem e tratamento de águas residuais; drenagem de águas pluviais; gestão de ribeiras; gestão da frente marítima e promoção de ações de educação ambiental e da participação pública.

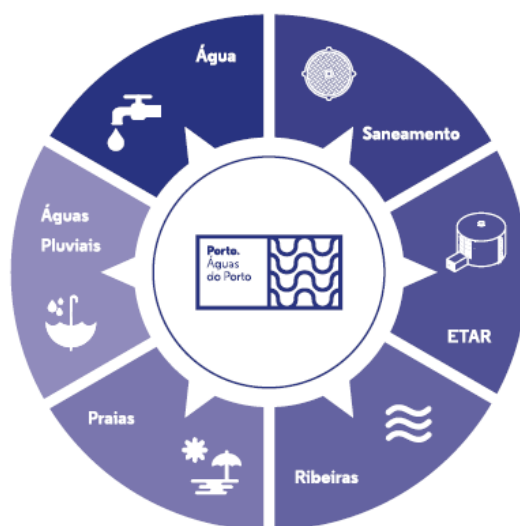


Fig.4. 1 – Objeto social (Águas do Porto, 2019b)

A AdPorto constitui uma das maiores empresas do setor, abastecendo 156 920 clientes numa conjuntura equivalente a 500 000 habitantes, o que corresponde a uma média de 47 520 m³/dia e atesta uma taxa de cobertura de 100%.

A água fornecida procede das Águas do Douro e Paiva, S.A, empresa de capitais públicos que arrecadou com a responsabilidade do abastecimento em alta à Área Metropolitana do Porto (AMP), consequente da necessidade de grandes investimentos ao nível de captação e tratamento de água. São três as captações que se localizam no rio Douro, na albufeira da barragem de Crestuma-Lever: Lever I - Gaia, Lever II - Porto e ETA de Lever.

O Laboratório de Análises da Águas do Porto é reconhecido pelo Instituto Português de Acreditação segundo a norma NP EN ISO/IEC 17025, onde se assevera o controlo analítico da qualidade da água em cumprimento com os padrões de qualidade rigorosos impostos pela legislação portuguesa e europeia. 40 671 análises são realizadas anualmente, das quais 33 536 provêm de rede pública e 7 135 de rede predial (Águas do Porto, 2019b).

4.2. UNIDADE ORGÂNICA DE GESTÃO DE CLIENTES

4.2.1. ENQUADRAMENTO

A estrutura organizacional da Águas do Porto, EM, está delimitada em cinco áreas operacionais e cinco áreas de suporte. Entre as áreas operacionais, notabiliza-se a Direção Comercial que se subdivide em duas unidades orgânicas: Gestão de Clientes e Gestão da Dívida, conforme ilustrado na Figura 4.2.

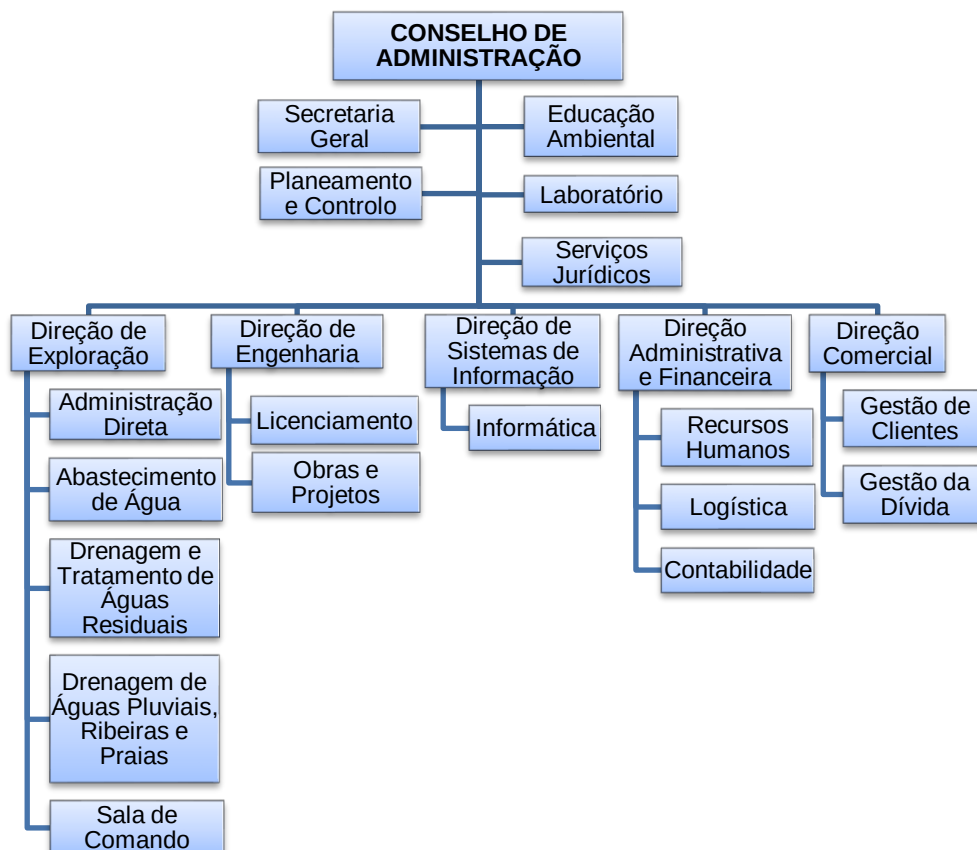


Fig.4. 2 - Organograma da empresa (Adaptado de Águas do Porto, 2019b)

A orientação estratégica vocacionada para a perspetiva do cliente assenta no desenvolvimento de um sistema integrado de gestão de clientes.

À UO de Gestão de Clientes compete o relacionamento da empresa com o utilizador final, composto por uma equipa multidisciplinar de colaboradores que se dedica desde o primeiro contacto do cliente com a empresa (celebração de contrato) à resolução de situações que geram insatisfação junto do cliente (gestão de reclamações).

Na figura seguinte surgem descritas as funções desempenhadas no contexto comercial.

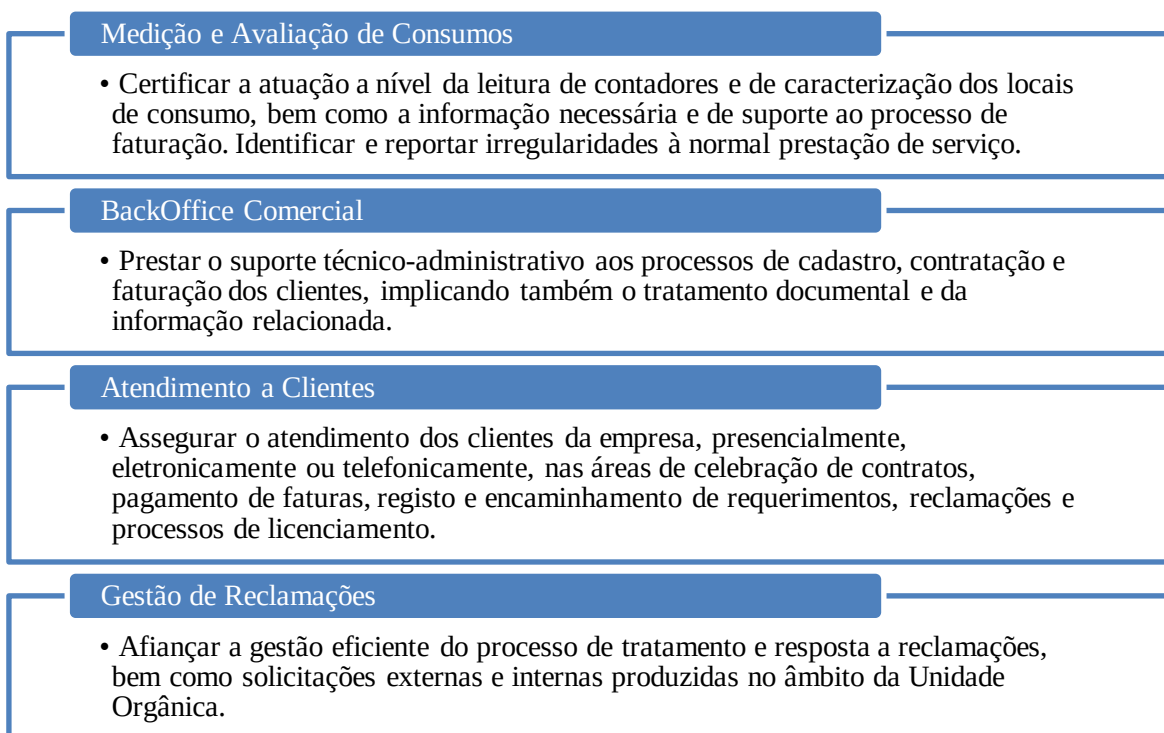


Fig.4. 3 - Principais linhas de atuação da UO (Adaptado de Águas do Porto, 2018a)

4.2.2. EVENTOS RELEVANTES EM 2018

No ano económico em análise, 376 668 clientes contactaram a empresa, tanto para apoio comercial como para apoio geral e técnico, sendo este último administrado pela Sala de Comando. Em média, cada cliente contactou 2,4 vezes ao longo do ano em busca de esclarecimentos, solicitação de serviços, reclamações e elogios, maioritariamente por telefone com 61,2% seguido pelo atendimento presencial com 24%. Quanto aos meios digitais, tais como correio eletrónico, Balcão Digital, plataforma de contratação e App, ainda apresentam um valor comparativamente baixo com 14,8%, o que se pode constatar pela Figura 4.4 (Águas do Porto, 2019b).

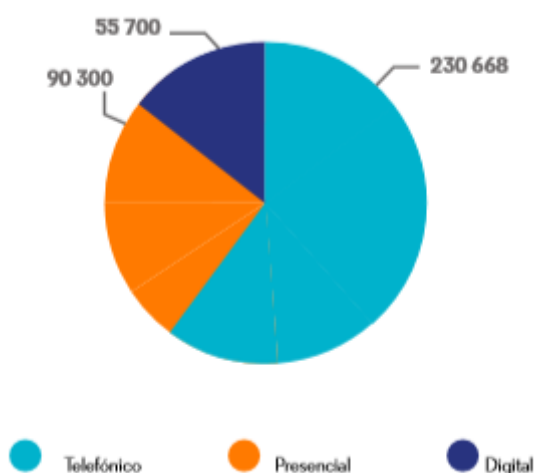


Fig.4. 4 - Meios de contacto utilizados pelos clientes em 2018 (Águas do Porto, 2019b)

4.2.2.1. Atendimento Comercial

Com base na Tabela 4.1, confirma-se que foram atendidos 193 091 clientes por contacto telefónico e presencial, valor esse que contradiz a propensão para diminuição dos últimos anos. Em virtude do acréscimo do contacto por via telefónica, ocorreu um incremento de 7,7% no número total de clientes atendidos confrontado com o ano transato, pois uma das metas para 2018 consistia exatamente pela translação de clientes do atendimento presencial para o telefónico (Águas do Porto, 2019b).

Tabela 4. 1 - Evolução do atendimento aos clientes (Adaptado de Águas do Porto, 2019b)

Tipo de Atendimento	2018	2017	Δ% (2017-2018)
Atendimento Telefónico			
Call Center	102 791	88 563	16,07%
Subtotal	102 791	88 563	16,07%
Atendimento Presencial			
Sede	57 059	53 638	6,38%
Gabinete do Município	33 241	37 138	-10,49%
Subtotal	90 300	90 776	-0,52%
Total	193 091	179 339	7,67%

No atendimento presencial, a quantidade de clientes foi praticamente constante, não obstante ambos os postos de atendimento registaram uma disparidade em que o número de clientes recebidos na sede cresceu aproximadamente 6,4% em relação a 2017, enquanto nos balcões do Gabinete do Município caiu cerca de 10,5% (Águas do Porto, 2019b).

Ao contrário do que era esperado, o aumento das chamadas telefónicas não se refletiu negativamente nos indicadores de tempo médio de espera (TME) e tempo médio de atendimento (TMA).

Aliás, como se pode visualizar na seguinte tabela, o TME decresceu 36,7% (Águas do Porto, 2019b).

Tabela 4. 2 - Evolução do TME e do TMA no *Call Center* (Adaptado de Águas do Porto, 2019b)

Atendimento Telefónico	2018	2017	Δ% (2017-2018)
Tempo médio de espera (TME)	00:00:38	00:01:00	-36,67%
Tempo médio de atendimento (TMA)	00:03:10	00:03:20	-5,00%

É demonstrado na Tabela 4.3 os TME e TMA nos dois locais de atendimento no intervalo em estudo.

Tabela 4. 3 - Crescimento do TME e TMA nos balcões de atendimento (Adaptado de Águas do Porto, 2019b)

Atendimento Presencial	2018	2017	Δ% (2017-2018)
TME			
Sede	00:12:08	00:09:51	23,18%
Gabinete do Município	00:20:44	00:17:56	15,61%
Total	00:15:14	00:13:05	15,61%
TMA			
Sede	00:08:16	00:08:26	-1,98%
Gabinete do Município	00:06:20	00:06:15	1,33%
Total	00:07:35	00:07:32	0,66%

Averiguou-se uma subida de 15,6% comparado ao ano anterior, ao passo que o TMA cresceu somente 0,7% à vista do período equivalente. O progresso do TME nos balcões de atendimento deveu-se à entrada dos processos de licenciamento, um novo serviço que passou a ser prestado ao público dada a necessidade de aglomerar todos os assuntos que envolvam o contacto presencial do cliente e limitar a uma única área da empresa (Águas do Porto, 2019b).

4.2.2.2. Interação Digital com os Utilizadores Finais

Um dos principais alvos da EG passa por favorecer a interação digital com os clientes, pelo que em novembro de 2018 entrou em vigor uma interface eletrónica de contratação, emergindo como alternativa à afluência aos balcões de atendimento, o que contribuiu para a celebração de sensivelmente 100 contratos. O pedido para realização de novo contrato também é possível efetuar via correio eletrónico, formalizando-se assim mais de 3 300 contratos durante o ano de 2018, correspondente a 20% do total dos contratos recentes (Águas do Porto, 2019b).

O novo Balcão Digital ao serviço do cliente irrompeu no final do exercício, mais prático, ágil e intuitivo. É tecnologicamente mais avançado, permitindo o acesso a várias funcionalidades, no entanto será

aprimorado num futuro próximo e com a expectativa de se tornar no meio preferencial de contacto dos clientes com a empresa a fim de simplificar a gestão dos contratos (Águas do Porto, 2019a).

Para além disso, no mesmo ano foi lançada uma nova campanha de adesão ao débito direto e à fatura eletrónica, repercutindo-se na poupança de tempo de espera, pagamento atempado, menor tempo de espera pela receção da fatura, à semelhança de uma maior consciencialização ambiental. Em 2018, observou-se uma ascensão de 32,2% no número de clientes aderentes à fatura eletrónica que passou dos 21 427 para 28 324, perfazendo 18,1% do total de utilizadores finais. Por outro lado, o número de clientes aderentes ao débito direto acumulou 4,5%, elevando de 54 778 para 57 260, verificando-se o encerramento do ano com 36,5% dos clientes a elegerem esta modalidade de pagamento (Águas do Porto, 2019b).

A título de curiosidade, em 2019 pretende-se ainda introduzir a chave móvel digital como uma forma suplementar de validação do titular do contrato, a juntar-se à atual por reconhecimento facial. Deseja-se similarmente atualizar o sistema de faturação comercial para um sistema de gestão integrada de toda a interface que incida nas três grandes temáticas da unidade comercial, detalhadamente, na contratação, liquidação e cobrança, onde se reconhecem os maiores constrangimentos (Águas do Porto, 2019b).

4.2.2.3. Atendimento Geral e Técnico

O atendimento geral e técnico é coordenado pela Sala de Comando, cuja atividade decorre de maneira intersectorial e transversalmente com a finalidade de obter uma maior operacionalidade e celeridade de resposta em todas as ocorrências na área de atuação da empresa. A sua atividade divide-se entre o atendimento geral (telefónico e eletrónico), o atendimento técnico (solicitações por parte dos clientes via eletrónica e telefónica, motivadas essencialmente por assuntos de natureza técnica) e a telegestão (gestão remota dos equipamentos em tempo real). Encontra-se em funcionamento todo o ano, de modo a servir continuamente, sendo situações frequentes os casos de roturas, falta de água, falta de pressão, avarias nos contadores ou obstrução de coletores (Águas do Porto, 2019b) (Águas do Porto, 2018b).

Em termos de números, foram atendidas 94 252 chamadas telefónicas no atendimento geral, representando um aumento de 9,5% relativamente a 2018 com uma espera média de 43 segundos, enquanto no atendimento técnico rececionaram-se 33 625 contactos para intervenção, registando-se um acréscimo de 2,5% e TME remontado a 39 segundos. Quanto ao correio eletrónico, foram recebidos cerca de 50 000 e-mails na conta institucional da empresa, porém a nível técnico deram entrada 2 300 pedidos (Águas do Porto, 2019b).

4.2.2.4. Gestão de Reclamações

Apesar das atividades do âmbito da empresa desenrolarem-se em prol do cliente, ocorrem situações suscetíveis de originar o descontentamento junto deste, sendo do interesse da empresa fomentar a sua resolução encarando como uma oportunidade de melhoria da qualidade do serviço.

Na sequência das reivindicações em 2018, assinalaram-se 2,75 reclamações redigidas por cada 1000 clientes. Em detalhe, totalizaram-se 388 os protestos em torno do AA e AR, mais 123 comparativamente com 2017, contrariando a tendência de decréscimo comprovada desde 2014. 51,5% das reclamações foram expostas no Livro de Reclamações em formato papel (disponível nos locais de atendimento) e online, depois são obrigatoriamente encaminhadas para a ERSAR que neste seguimento apresentou 32 pedidos de esclarecimento adicionais. Assim, o crescimento destas manifestações foi 62,6%, tendo subido 32,4% noutros meios como e-mail, correio normal, etc. (Águas do Porto, 2019b).

5

METODOLOGIA PARA GESTÃO DE SOLICITAÇÕES DE CLIENTES

5.1. PROCEDIMENTO ADOTADO PELA EMPRESA

De acordo com o Artigo 35.º do Regulamento de Relações Comerciais dos Serviços de Águas e Resíduos, um dos deveres de uma EG compreende a disponibilização de meios de atendimento com vista à resolução de situações que envolvam os serviços de abastecimento de água e de saneamento de águas residuais. A complementar, segue o Artigo 40.º no âmbito do atendimento ao público, onde se expressa visivelmente a necessidade de locais de atendimento ao público e de um serviço de atendimento telefónico e via *internet*, pelos quais os utilizadores finais podem contactar diretamente.

Conforme explorado no capítulo transato, a Águas do Porto, EM possui várias formas de interação com o cliente, nomeadamente, a partir dos balcões de atendimento presencial, por correio eletrónico, por contacto telefónico, pelo Balcão Digital, plataforma de contratação ou app que, embora ocupem uma posição de menor relevo com a percentagem mais baixa quanto ao meio de contacto, ainda assim são um meio em crescimento cada vez com mais aderência, atentando aos padrões da sociedade moderna.

Este caso de estudo incide essencialmente sobre o contacto por correio eletrónico que receciona inúmeras solicitações de temáticas diversas, no entanto o cenário atual exige constantemente a preparação da resposta e o ato de envio por parte de um colaborador assinalado para o propósito. Na Fig. 5.1 consta um fluxograma que representa o processo em torno da gestão de e-mails rececionados no endereço cliente@aguasdoporto.pt do domínio da área comercial.

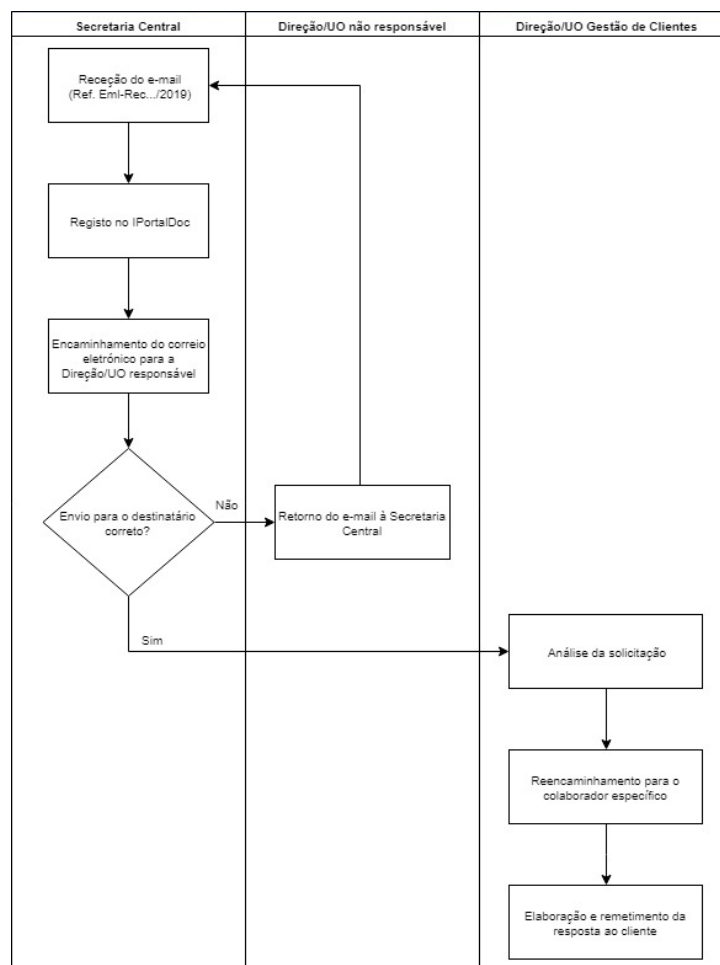


Fig. 5. 1 - Fluxograma das solicitações enviadas para o endereço cliente@aguasdoporto.pt

O correio eletrónico remetido pelo cliente dá entrada automaticamente na Secretaria Central, onde é atribuída uma designação com numeração sequencial por ano, “Ref.Eml-Rec.../2019” e registado no IPortalDoc, sendo inserido o local de procedência, para gestão documental e de processos. Posteriormente, um colaborador da Secretaria Central encaminha o respetivo correio para a Direção/UA responsável pelo assunto em causa, contudo, é considerado neste esquema a hipótese de o envio ser bem sucedido ou não. Como tal, caso este não alcance o destinatário correto retorna ao ponto de partida e volta a seguir os passos anteriores. Uma vez enviado o e-mail com êxito, é analisado pelo administrador do endereço que efetua o reencaminhamento para um colaborador específico, encarregue pela composição e envio da resposta final ao cliente. Não obstante todo o processo ocorrer para o endereço do serviço comercial, a solução é direcionada para o cliente a partir do endereço do colaborador em questão, pois ainda não está programado para uso direto dos funcionários.

Na representação seguinte, indicada pela Fig. 5.2, encontra-se detalhado o processo de gestão de e-mails que chegam ao endereço “geral@aguasdoporto.pt”, o endereço genérico da EG gerido pela Sala de Comando. Aqui são rececionados todo o tipo de rubricas concernentes às várias unidades orgânicas, ao contrário do endereço prévio onde apenas constam conteúdos do foro comercial, pelo que as fases de atuação diferem tenuemente do precedente.

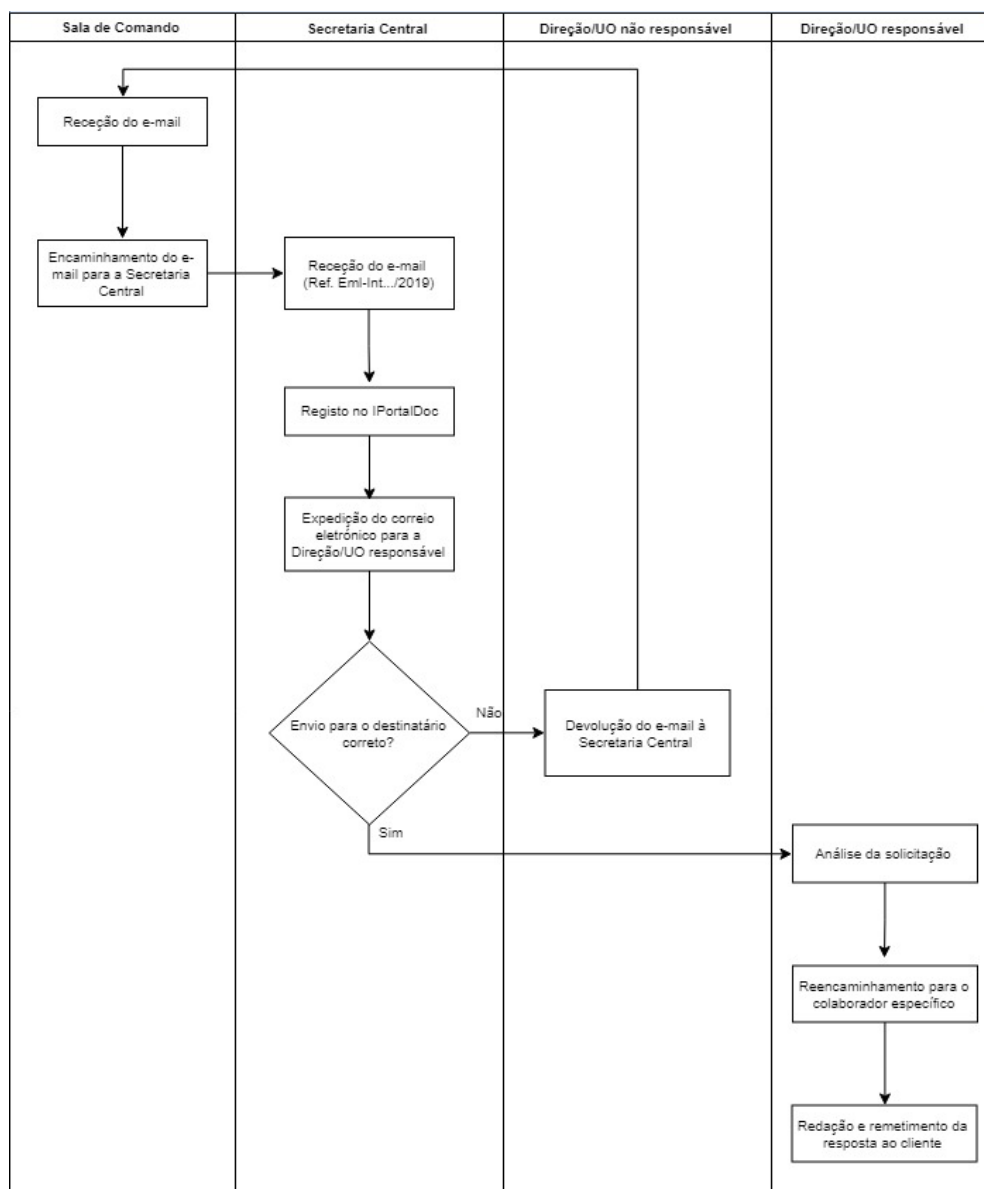


Fig. 5. 2 - Fluxograma das solicitações remetidas para o endereço geral@aguasdoporto.pt

O cliente envia um e-mail, onde é recebido na caixa de entrada do endereço corrente e conduzido para a Secretaria Central por um colaborador responsável pela gestão do correio eletrónico da Sala de Comando. Assim que ingressa, arrecada com uma numeração sequencial por ano sob o formato “Ref.Eml-Int.../2019” e registado no IPortalDoc com o preenchimento dos campos “entidade” e “tipo de entidade”. Depois prossegue para a Direção/UO incumbida, se o envio for efetivado para o destinatário correto, senão procede-se à devolução para a Secretaria Central que acaba por repetir as etapas pela ordem mencionada anteriormente.

Aquando da receção na Direção/UO respetiva, o correio eletrónico é reencaminhado para um colaborador designado após a submissão do conteúdo à observação por parte do administrador da conta. O colaborador em questão faz a redação do e-mail e expede a resposta diretamente ao cliente.

5.2. PROPOSTA DE CLASSIFICAÇÃO PARA SOLICITAÇÕES

Na sequência da criação do Guia de Reclamações intitulado “Gestão de Reclamações como Oportunidade de Melhoria” pelo Grupo de Trabalho de Gestão de Clientes da APDA em 2015 e depois adaptado à realidade da Águas do Porto, EM no ano presente com a denominação de “Gestão de Reclamações, Sugestões e Elogios – Manual de Procedimentos”, procedeu-se à conceção de um método aplicado de maneira análoga às solicitações assente no último documento.

Pretende-se consolidar um conjunto de orientações com a finalidade de homogeneizar as definições e conceitos articulados com muitos aspetos do relacionamento entre os clientes e entidade gestora com principal enfoque para as solicitações. Importa frisar que os termos “esclarecimento”, onde o cliente requer informações sobre a prestação do serviço, e “solicitação”, como um pedido realizado à EG que demanda uma ação da mesma, foram considerados ambos solicitações, visto que se trata no fundo de um pedido apresentado pelo cliente.

O tratamento de rogativas visa apresentar e potenciar futura discussão das linhas gerais, definições e objetivos, constituindo um instrumento de melhoria da qualidade do serviço e da imagem da empresa (APDA, 2015).

5.2.1. TIPOLOGIAS

Relativamente à classificação de solicitações, recomenda-se uma metodologia com o intuito de categorizar os assuntos respeitantes a uma solicitação, ponderando os distintos ramos que se dedicam à prestação do serviço de abastecimento de água e drenagem de águas residuais, tais como: Atendimento; Contratação; Faturação, Leitura e Tarifário; Ligação e Disponibilidade e Restantes Categorias, onde se integra a Qualidade da Água, Prestação de Serviço Técnico, Contador, Regularização do Corte de Água, Tampa/Grelha, Fuga de Água, Danos – Inundações e/ou Infiltrações, Pavimentos, Obras, Ribeiras e Praias, ETAR e Outras, conferindo-lhes uma classificação principal, secundária e descrição correspondente.

Seguem-se os tópicos mais frequentes de pedidos, repartidos em conformidade com o contexto das áreas anteriormente indicadas.

5.2.1.1. Atendimento

O Atendimento ao Público numa Entidade Gestora de Abastecimento de Água e Saneamento marca o início da conexão com o cliente, cuja prestação de serviço se formaliza através da contratação. Nos últimos anos, a qualidade de serviço e imagem externa têm adquirido maior peso na área do atendimento, resultando num investimento adicional, de modo a acompanhar a evolução de novas tecnologias e as necessidades dos clientes. Os distintos meios de contacto entre a EG e os utilizadores finais (presencial, telefónica ou *internet*) são fruto da aposta em ferramentas que proporcionam a otimização e melhoria contínua neste campo. Na Tabela 5.1 mostram-se exemplos de possíveis solicitações que se inserem na tipologia do atendimento.

Tabela 5. 1 - Solicitações do tipo atendimento

Classificação Principal	Classificação Secundária	Descrição
ATENDIMENTO	Acesso pelo cliente aos meios de contato com a EG (telefônicos/eletrônicos) Ex: Sítio <i>Internet</i> , Balcão Digital, etc.	Acesso aos meios de contato com a EG. Ex: Contato telefónico da EG; acesso à plataforma Balcão Digital e <i>website</i> .
	Acesso a informação do interesse do cliente veiculada nos meios eletrônicos de contacto com EG	Informação complementar ou esclarecimento de informação existente.
	Espaço físico e acessibilidade aos serviços de atendimento	Melhoramento das instalações. Facultação de acessos para população com mobilidade condicionada/reduzida, etc.
	Horário de atendimento (presencial/telefónico)	Horário de abertura ou encerramento dos serviços de atendimento (telefónico/presencial) ou encerramento antecipado da distribuição de senhas.
	Atendimento personalizado	Atendimento conduzido sob diretrizes específicas.
	Livro de Reclamações	Acesso ao Livro de Reclamações.
	Informações respeitantes ao atendimento	Funcionamento do atendimento. Inclui atendimento telefónico. Não inclui informação prévia à contratação.
	Tempo de resposta a pedidos de informação e esclarecimento	Tempo de espera para obtenção de resposta a pedido de informação ou esclarecimento; retribuição de contato solicitado pelo cliente.
	Tempos de espera no atendimento (presencial/telefónico)	Ordem das senhas e prioridade no atendimento; tempo de espera em linha.

5.2.1.2. Contratação

A contratação para a prestação de serviço de água e saneamento baseia-se num acordo entre ambas as partes, sustentada no conhecimento total das condições contratuais que contribuem para o serviço em análise. Eis algumas situações de possíveis pedidos sobre a contratação que podem ser constatadas na seguinte tabela:

Tabela 5. 2 - Solicitações do tipo contratação

Classificação Principal	Classificação Secundária	Descrição
CONTRATAÇÃO	Informação contratual	Comunicação das condições contratuais.
	Celebração de contrato Alteração de titularidade	Contratação do serviço, apresentado a documentação necessária.
	Correção de dados contratuais	Retificação de dados respeitantes ao cliente (Ex: NIF, morada de envio de correspondência, etc.).
	Prorrogação do contrato com a tarifa “obras”	Prolongamento do contrato sujeito a apresentação de documento comprovativo (Prorrogação de Prazo para Obras).
	Rescisão ou suspensão de abastecimento a pedido	Efetivação da cessação do contrato.
	Alteração de grupo tarifário; Atribuição de tarifário	Alteração da tarifa em vigor. Imputação de tarifa (doméstica, comércio, familiar, etc.).
	Disponibilização do serviço de abastecimento de água (contador)	Prestação do serviço depois de contratado (colocação de contador).
	Contratação aquando da indisponibilidade do serviço	Contratação antes do término da execução do ramal.
	Outras questões de índole contratual	Questões alusivas à contratação não abrangidas pelas demais classificações.

5.2.1.3. Faturação, Leitura e Tarifário

A faturação e leitura possui um vasto campo de aplicação, pois recai sobre todos assuntos referentes ao processo de faturar, a informação transparente e concisa contida na fatura. Abrange também os tópicos em redor das leituras, desde a sua obtenção à utilização do valor obtido, enquanto o tarifário expressa-se por uma combinação de valores unitários, outros parâmetros e regras de cálculo, que permitem determinar o montante exato a pagar pelo cliente à Entidade Gestora, como contrapartida do serviço prestado. Os valores cobrados servirão para asseverar a viabilidade e melhoria dos serviços prestados. Seguem alguns exemplos de rogativas mais habituais:

Tabela 5. 3 - Solicitações do tipo faturação, leitura e tarifário

Classificação Principal	Classificação Secundária	Descrição
FATURAÇÃO, LEITURA E TARIFÁRIO	Informação prestada na fatura	Esclarecimentos de elementos presentes ou em falta na fatura.
	Periodicidade das faturas	Intervalo da emissão entre faturas, mensal/bimestral ou outra, a pedido ou imposta pela EG.
	Periodicidade das leituras e estimativas de consumo	Periodicidade das leituras, utilização das leituras comunicadas, valor adotado para estimativa de consumo por ausência de leitura.
	Retificação de valores faturados	Correção de valores faturados, acertos e distribuição do consumo pelos escalões. Correção das leituras faturadas, seja leituras efetuadas pela EG ou fornecidas pelo cliente. Correção de serviços faturados e no entender do cliente não prestados ou considerados ilegítimos. Correção de faturação de consumos estimados e não medidos pelo contador ou por ausência do mesmo.
	Meios de pagamento/recebimento	Meios de pagamentos disponibilizados para o pagamento de faturas, ativação do débito direto e todos os aspetos relacionados com o mesmo ou referências multibanco; emissão de notas de crédito, débito ou recibos de cobrança. Inclui forma de realização, de acertos de pagamento (compensação na fatura seguinte ou emissão de nota de crédito).
	Envio de faturas	Faturas requisitadas pelo cliente, desde 2ª via a faturas específicas não rececionadas ou extraviadas.
	Fatura eletrónica	Adesão à fatura eletrónica, confirmação da adesão, alteração do endereço de correio eletrónico, esclarecimentos alusivos à adesão, cancelamento da receção por via eletrónica, assim como outros pedidos no âmbito da fatura eletrónica.

	Acordo de pagamento	Realização de um plano de pagamento em prestações. Envolve esclarecimentos ou pedido de dados para pagamento.
	Extrato de conta corrente	Pedido do extrato de conta corrente correspondente a um determinado período.
	Tarifário	Todas as solicitações que se refiram à estrutura tarifária, tarifa de serviços complementares, valor das principais tarifas, tarifários especiais (condições de adesão à Tarifa Social, Tarifa Familiar, etc., regras e valores atribuídos), valores de impostos e taxas faturados (IVA, TRH, TGR, entre outras integradas na fatura) e a alterações do tarifário.
	Outras situações referentes à faturação e leitura	Questões relativas à faturação e leitura não incluídas pelas demais classificações.

5.2.1.4. Ligação e Disponibilidade

A ligação à rede pública de abastecimento de água e saneamento é de carácter indispensável, quando a mesma esteja disponível para tal. A disponibilidade pode ser encarada como uma garantia da prossecução do serviço prestado, regular e em qualidade, assegurando a boa gestão dos sistemas públicos de aa e saneamento. Comprovam-se algumas das situações de possível pedido para ligação às redes públicas de abastecimento de água e drenagem de águas residuais:

Tabela 5. 4 - Solicitações do tipo ligação e disponibilidade

Classificação Principal	Classificação Secundária	Descrição
LIGAÇÃO E DISPONIBILIDADE	Disponibilização do serviço de abastecimento de água	Requisição da ligação à rede pública de abastecimento, uma vez disponível para servir fisicamente.
	Disponibilização do serviço de saneamento de águas residuais	Pedido de ligação à rede pública de saneamento, dada a sua disponibilidade para servir fisicamente.
	Disponibilização do serviço de águas pluviais	Exigência de ligação à rede de águas pluviais, fisicamente disponível para servir o prédio.
	Condições de ligação dos ramais às redes públicas	Emissão de plantas topográficas. Agendamento vistoria/fiscalização a obras. Inclui outras informações acerca

		das condições de ligação, bem como a necessidade de instalação de bombas de elevação.
	Instalação e alterações do local do contador	Abrange obras relacionadas com o nicho do contador, alterações do local do nicho do contador e acessórios do nicho do contador (exclui definições ou alterações do calibre do contador).
	Outras questões associadas à ligação e disponibilidade	Questões de ligação e disponibilidade que não se inserem nas demais classificações.

5.2.1.5. Prestação de Serviço Técnico

Contém solicitações sobre a Prestação de Serviço Técnico, onde se apuram situações normalmente decursivas do funcionamento das redes públicas e ainda as solicitações em que se confirme uma ação por parte da EG no equipamento de medição de consumos, consoante a tabela que se segue:

Tabela 5. 5 - Solicitações do tipo prestação de serviço técnico

Classificação Principal	Classificação Secundária	Descrição
PRESTAÇÃO DE SERVIÇO TÉCNICO	Execução de serviços domiciliários no âmbito do contrato de abastecimento	Prestação do serviço.
	Serviço de saneamento	Limpeza de fossas sépticas ou funcionamento da rede de saneamento.
	Serviço de águas pluviais	Atuação com a intenção de repor o normal funcionamento da rede de águas pluviais.
	Interrupção do abastecimento devido a roturas e/ou obras na rede	Atuação em roturas na rede sujeita a interrupção prévia; pedido de aviso oportuno; informação sobre a interrupção de abastecimento.
	Prazo previsto para regularização do fornecimento	Informação acerca do período de interrupção e eventual reposição.
	Solução para falta de água	Intervenção por motivos alusivos à rede de distribuição e problemas decorrentes da rede predial;

	Pressão da rede de abastecimento	esclarecimentos e/ou intervenção na pressão da rede: para reduzir ou aumentar.
	Contador	Todas as solicitações associadas à instalação ou retirada do contador, bem como intervenção em contador danificado, funcionamento anómalo, avaria, contador inoperacional e definição do tipo de calibre.
	Outras questões relacionadas com a prestação de serviço técnico	Questões de prestação de serviço técnico que não se identificam nas restantes classificações.

5.2.1.6. Restantes Categorias

As categorias remanescentes são agrupadas numa só categoria onde se introduzem os assuntos referentes à Qualidade da Água, que engloba pedidos de intervenção consequentes de transgressões no âmbito da qualidade da água; ao Corte de Água, em que o cliente solicita a regularização do serviço de AA; às Fugas de Água, em que é requisitada a intercessão da EG neste tipo de ocorrências; aos Danos decorrentes de Inundação e/ou Infiltrações; à Reparação/Reposição de Tampa/Grelha; à Reparação/Reposição de Pavimentos; às Obras, que refletem as solicitações dos clientes quanto ao modo como decorrem as obras numa perspetiva de impacto na zona pública; à atividade em Ribeiras e Praias; à ETAR, que passa pela caracterização dos fatores que proporcionem possíveis solicitações a respeito das atividades decorridas no seu espaço e por último, outras solicitações de carácter variado que não se encaixam em nenhuma das classes citadas.

Tabela 5. 6 - Solicitações das categorias excedentes

Classificação Principal	Classificação Secundária	Descrição
RESTANTES CATEGORIAS	Qualidade da Água	Verificação da qualidade da água abastecida e intervenção por parte da EG. Compreende também pedidos de intervenção em situações de descargas de águas residuais na via pública, rios, ribeiras ou água (insalubridade) e ainda no âmbito da qualidade da água de ribeiras e praias.
	Corte de Água	Solicitações para restabelecimento do serviço de abastecimento de água (ordem de religação) e esclarecimentos acerca do motivo que levou ao corte de água.

	Fuga de Água	Requisição de atuação em fuga de água ocorrida em equipamentos de caráter público (Ex: fontes, bebedouros e hidrantes), assim como fuga de água sucedida em arruamento público ou acessório da rede de abastecimento de água e em fuga de água no interior de um imóvel.
	Danos – Inundação e/ou Infiltrações	Resolução para danos provocados a terceiros devido a inundação originada por rotura na rede de distribuição no decorrer de intervenções por serviços prestados pela instalação, substituição e retirada de contadores, tal como, por obstrução da rede de saneamento e águas pluviais); resolução para danos provocados a terceiros por infiltração, sem inundação, produzida por rotura de conduta ou por obstrução da rede de saneamento e/ou águas pluviais.
	Reparação/Reposição de Tampa/Grelha	Todas as solicitações que envolvam acessórios da rede de infraestruturas.
	Reparação/Reposição de Pavimentos	Todas as solicitações relacionadas com reparação/reposição de pavimentos em mau estado sequente das infraestruturas da Águas do Porto.
	Obras	Plena sinalização completa e visível de obras na via pública que englobam ainda a delimitação visível da localização do estaleiro e definição visível da circulação pedonal; esclarecimentos acerca de obras a decorrer; reposição das condições iniciais ou atraso na reposição de pavimentos; reparação de infraestruturas deterioradas por intervenções (fugas na rede de distribuição, redes de rega, hidrantes, etc.).
	Ribeiras e Praias	Pedidos alusivos à atividade da Águas do Porto que se enquadrem no contexto das ribeiras e praias.
	ETAR	Atuação por parte da EG relativamente à presença de odores nauseabundos e descargas de águas residuais urbanas; explanações acerca da legislação

		aplicável e/ou conduta adotada pela EG neste âmbito (Ex: Licença de Rejeição).
	Outras	Solicitações possíveis ou não de estar relacionadas com infraestruturas da Águas do Porto e solicitações que não se enquadram em nenhum dos contextos anteriores.

Não foram consideradas as categorias de “Acidentes pessoais” e “Sinistro com veículos” em conformidade com o Manual de Procedimentos da AdPorto, já que se trata de situações pontuais com rara incidência ou até mesmo nula, no entanto poderão encaixar-se no contexto de “Outras categorias”.

5.3. GEORREFERENCIAÇÃO DE SOLICITAÇÕES E RECLAMAÇÕES PROVENIENTES DE CORREIO ELETRÓNICO NA CIDADE DO PORTO

Dada a apresentação das várias categorias de solicitações, constata-se a relevância dos tipos de consumidores de modo a entender o panorama geral em que se introduz este estudo. Os utilizadores finais encontram-se organizados por distintas tipologias: Doméstico, Social, Empresarial, Público, Autárquico e Próprios, sendo diferenciados pelas características de consumo e tarifários praticados. A distribuição dos clientes é revelada na Tabela 5.7 que se resume aos últimos dois anos de funções da empresa.

Tabela 5. 7 - Evolução do número de clientes por tipologia (Adaptado de Águas do Porto, 2019b)

Cliente	2018		2017		Δ% 2017-2018
	N.º	%	N.º	%	
Doméstico	127 103	81,0	128 498	82,5	-1,09%
Social	440	0,2	467	0,3	-5,78%
Empresarial	27 908	17,8	25 251	16,2	10,52%
Público	341	0,2	346	0,2	-1,45%
Autárquico	1 081	0,7	1 081	0,7	0,00%
Próprios	47	0,03	24	0,02	95,83%
Total	156 920	100	155 667	100	0,80%

Atendendo aos dados da tabela anterior, é perceptível a maioria dos consumidores afeta ao tipo doméstico com 81,0%, seguindo-se o tipo empresarial com 17,8%. Houve uma variação significativa neste período de observação, em que os consumos próprios e empresarial representaram acréscimos na ordem dos 98,53% e 10,52% respetivamente na contratação, enquanto a maior quebra foi registada para o consumo do tipo social, cuja percentagem foi de -5,78%.

Todos os tipos de cliente foram notados numa primeira análise que se debruçou sobre ambos os endereços eletrónicos consignados no início deste capítulo.

O e-mail “geral@aguasdoporto.pt” foi analisado no intervalo de 01 de janeiro de 2019 a 31 de março de 2019, num total de 4992 itens, ao passo que a pesquisa na conta “cliente@aguasdoporto.pt” decorreu no tempo de 17 de junho de 2019 a 13 de agosto de 2019, com um total de 1046 itens, assim perfazendo um total de 6038 de itens examinados nos dois endereços. O desfasamento apurado deveu-se ao facto do último endereço ter sido concebido ainda durante o período escolhido, contribuindo posteriormente com mais exemplos para a efetuação deste estudo.

No espaço da investigação foram categorizadas as solicitações presentes no corpo de texto desses e-mails sob diversos assuntos. Findo essa fase, procedeu-se à subdivisão dos pedidos pelas categorias genéricas, conforme demonstrado dantes, e repartiu-se cada uma das tipologias por folhas de *Excel* constituindo a base de dados necessária à representação das solicitações na cidade do Porto.

5.3.1. REPRESENTAÇÃO DAS SOLICITAÇÕES DA CATEGORIA ATENDIMENTO

As solicitações do tipo atendimento foram identificadas como todas aquelas em que o cliente requereu esclarecimentos ou ajuda em questões relativas ao Balcão Digital. Os pedidos mais frequentes foram o acesso à plataforma, bem como apoio aquando do registo e visualização do histórico de faturação por dificuldades técnicas, somando um total de 132 instalações. Na Figura 5.3. é possível observar a georreferenciação das várias solicitações de carácter atendimento através dos códigos de identificação local (CIL) dos clientes reconhecidos aquando da pesquisa.

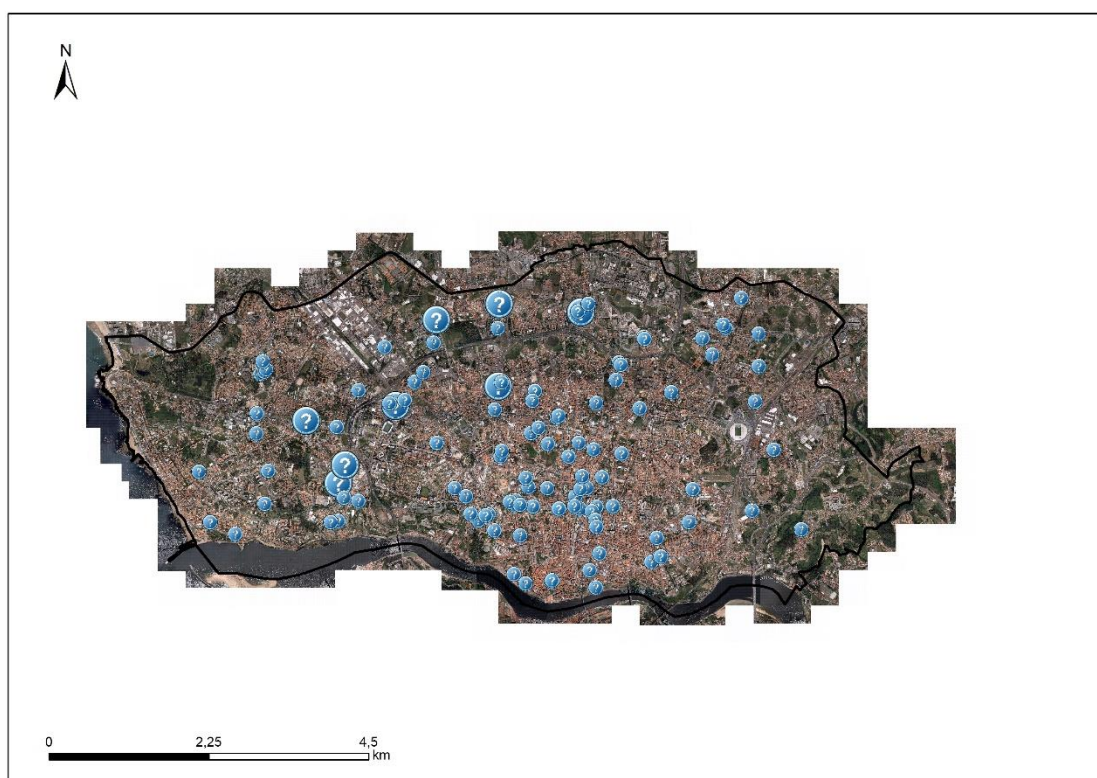


Fig. 5. 3 – Georreferenciação das solicitações do tipo atendimento

Pela figura pode averiguar-se a existência de instalações marcadas com mais ou menos relevo, o que se deve à repetição dos mesmos locais de consumo no caso de maiores dimensões. Isto traduz-se pelo facto de alguns clientes se manifestarem mais vezes no cerne da mesma tipologia ou até mesmo por reiteração de um pedido semelhante.

Quanto às zonas de consumo adotadas pela Águas do Porto, não é denotado um padrão na tendência deste tipo de pedidos, pelo que se estende ao longo da cidade do Porto com proeminência na parte mais central, porém com a zona de Francos a destoar nesta descrição.

5.3.2. REPRESENTAÇÃO DAS SOLICITAÇÕES DA CATEGORIA CONTRATAÇÃO

Apresenta-se como a segunda maior categoria com solicitações validadas de natureza contratual que compreende principalmente o processo em torno da celebração e denúncia de contratos, alteração de dados contratuais e esclarecimentos acerca do procedimento para novos contratos ou rescisões, resultando na recolha de 2001 locais de consumo, mas 15 não estão georreferenciados, pois não têm número de polícia no SIG. Na figura seguinte encontram-se representadas as instalações decorrentes das solicitações desta categoria no município do Porto.

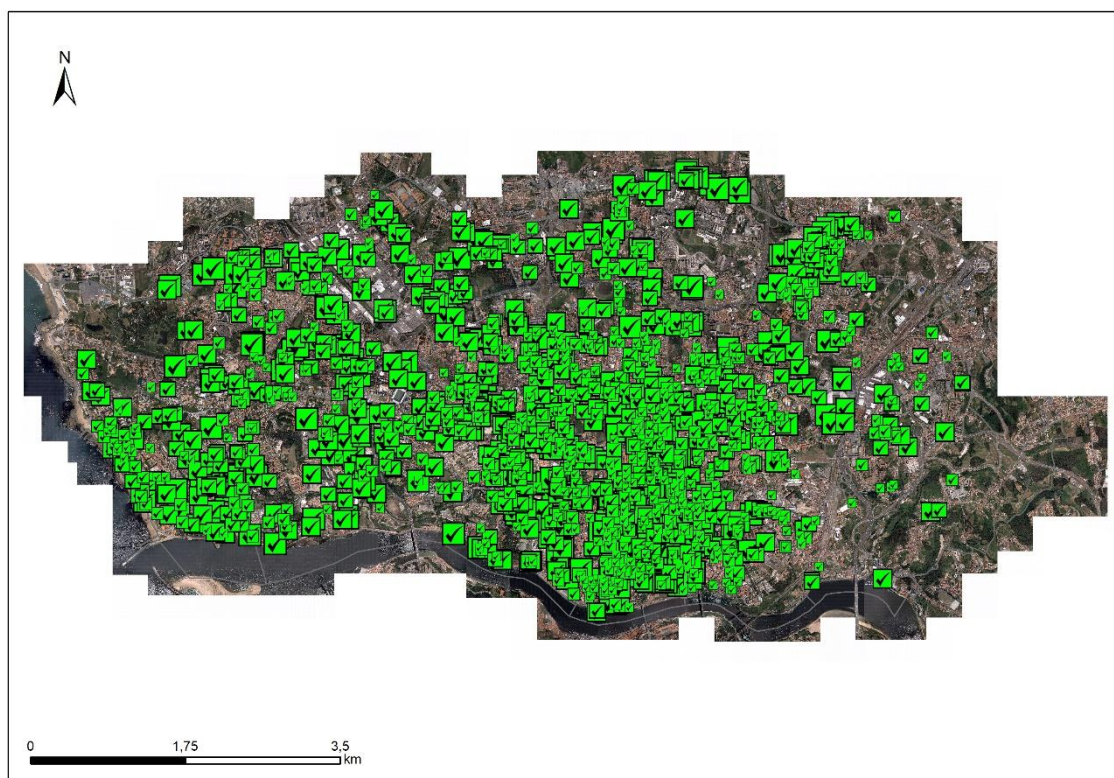


Fig. 5. 4 - Georreferenciação das solicitações do tipo contratação

Aqui verificam-se os locais de consumo espalhados um pouco por toda a parte e particularmente centrados na baixa da cidade dada a enorme afluência de turismo que se tem verificado nos últimos anos incentivando à expansão de alojamentos locais, o que significa mais contratos realizados, com exceção das zonas mais periféricas da Areosa e Cerco/Azevedo que possuem menos ou nenhuns. Visualizam-se

alguns marcadores com mais relevo que se deve especialmente à celebração ou cancelamento de contratos na mesma instalação.

5.3.3. REPRESENTAÇÃO DAS SOLICITAÇÕES DA CATEGORIA FATURAÇÃO, LEITURA E TARIFÁRIO

Esta é a categoria com maior número de solicitações registadas, onde são tratadas todas as questões em redor da faturação, leitura e tarifário. Os pedidos mais habituais foram a 2ª via de faturas, retificação de valores faturados, extrato de conta corrente e ressarcimento de valores em nota de crédito, reunindo um total de 2517 instalações em que 55 não possuem qualquer georreferenciação.

Com apenas 4 solicitações registadas no período em questão, o tarifário é uma das categorias menos apelantes, passando normalmente por esclarecimentos relativamente ao tarifário em vigor e sobre as condições de adesão à tarifa familiar. A representação é feita na figura seguinte.

Na Figura 5.5. comprova-se a representação das instalações atribuídas a esta categoria de solicitações ao longo da cidade.

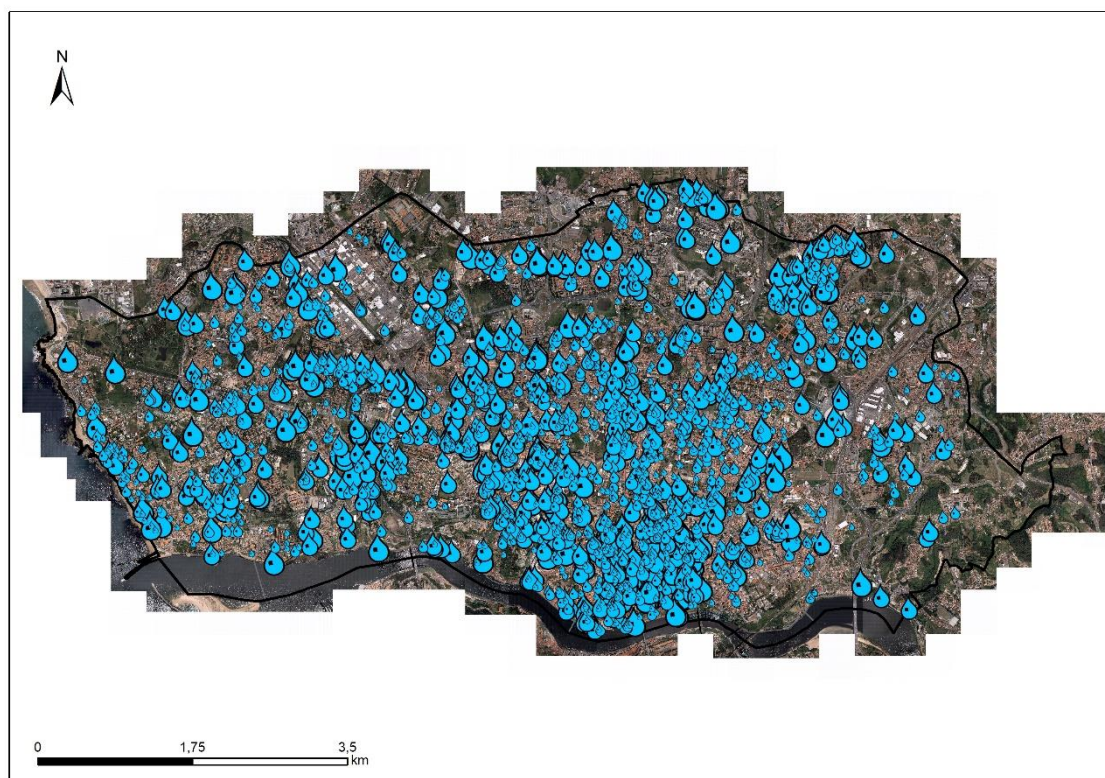


Fig. 5. 5 - Georreferenciação das solicitações do tipo faturação, leitura e tarifário

Atentando à figura, verifica-se uma distribuição de pedidos aproximadamente uniforme com mais incidência na baixa da cidade e menor na zona do Cerco/Azevedo. Quanto à temática de tarifário, são muito poucos os pedidos e encontram-se dispersos pela cidade do Porto.

É de salientar que vários clientes recorrem ainda ao correio eletrónico como meio preferencial para comunicação de leitura, pelo que o código projetado no passo seguinte desta dissertação seria um elemento dissuasor conduzindo a outras formas de envio. Contudo, por razões técnicas ou impercetibilidade do cliente mantém-se como uma alternativa, de modo a concretizar a submissão do valor correspondente ao consumo obtido para um determinado período.

5.3.4. REPRESENTAÇÃO DAS SOLICITAÇÕES DA CATEGORIA LIGAÇÃO E DISPONIBILIDADE

Os pedidos de ligação e disponibilidade correspondem a solicitações, maioritariamente, de emissão de plantas topográficas, requerimentos para executar ligações às redes de abastecimento de águas, águas residuais domésticas e águas pluviais, assim como verificações técnicas de AA. Para o período de investigação, obtiveram-se 58 CIL georreferenciados na íntegra, segundo se pode provar pela figura que se segue.

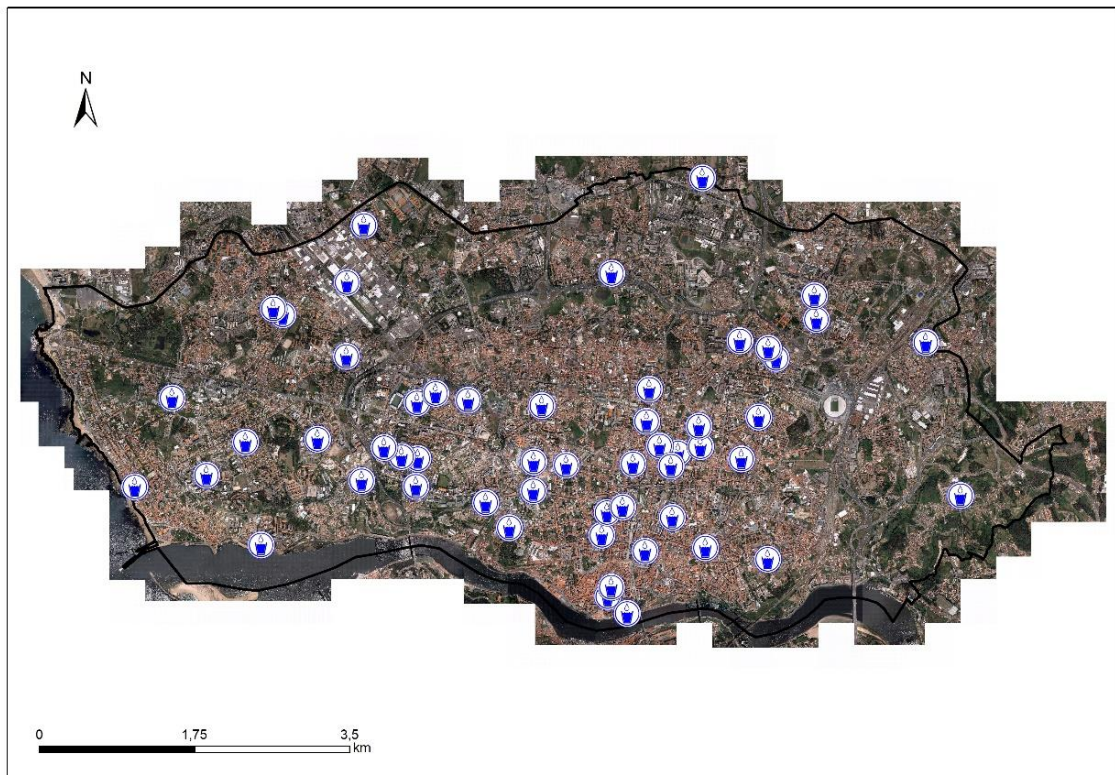


Fig. 5. 6 - Georreferenciação das solicitações do tipo ligação e disponibilidade

Apesar de na realidade se ter verificado inúmeros pedidos desta categoria, a maioria deles não dispunham de qualquer informação conducente à instalação pela simples circunstância de que, no caso de emissão de plantas topográficas e execução de ramais, o requerente não ter ainda uma instalação no local. Essa abundância de solicitações, particularmente registadas na baixa da cidade do Porto, deve-se precisamente ao elevado número de obras de reabilitação de casas antigas em virtude do incremento do turismo. Esta observação é discutida posteriormente no ponto 5.3.8.

5.3.5. REPRESENTAÇÃO DAS SOLICITAÇÕES DA CATEGORIA PRESTAÇÃO DE SERVIÇO TÉCNICO

Surge como outra das tipologias com algumas solicitações comuns, devido ao funcionamento dos serviços de saneamento e águas pluviais. Das 75 instalações anotadas, 2 não são georreferenciadas por ausência do número de polícia. Inserem-se também os assuntos relacionados com o contador que demonstra um número significativo de incidências, cujos principais motivos devem-se à colocação de um novo contador e aferição do equipamento, em que se reuniram 77 CIL durante a análise, surgindo 2 sem qualquer georreferenciação. A figura que se segue demonstra as várias instalações marcadas no município.

Na Figura 5.7 encontra-se a representação geográfica dos pedidos desta categoria.



Fig. 5. 7 - Georreferenciação das solicitações do tipo prestação de serviço técnico

De acordo com a figura, é bastante perceptível o reforço de alguns locais de consumo, cujo motivo passa pelos pedidos reiterados por intervenção urgente. Contudo, o número de solicitações rececionadas pela via eletrónica é visivelmente menor comparado com as solicitações efetuadas por contacto telefónico, visto que os clientes preferem a última pela simples razão do atendimento e posterior acionamento dos meios ocorrer de forma mais rápida.

5.3.6. REPRESENTAÇÃO DAS SOLICITAÇÕES DAS RESTANTES CATEGORIAS

Esta categoria resulta da aglomeração de quatro subcategorias que tiveram menor incidência: Danos - Inundação e/ou Infiltrações; Corte de Água; Fuga de Água e Qualidade da Água.

A primeira subcategoria caracteriza-se pelos danos resultantes de inundações e ou infiltrações provocadas a terceiros, em que o cliente habitualmente solicita uma solução para ocorrências deste teor e foram 11 as instalações recolhidas.

Quanto às fugas de águas, estas podem ocorrer em equipamentos públicos, arruamentos ou no interior das habitações. No entanto, o mais observado para esta conjuntura foi em equipamentos como bocas de incêndio e bebedouros, sendo 8 as instalações obtidas.

Relativamente ao corte de água, 8 foi o número de instalações de onde provieram pedidos, nomeadamente, alusivos a esclarecimentos acerca da causa de corte de água ou solicitações para restabelecimento do serviço. Confirmou-se 1 local de consumo sem georreferenciação no SIG.

No que toca à qualidade da água, ostenta-se como uma das subcategorias menos requisitadas, em que a insalubridade constituiu o aspeto que mais se identificou neste assunto em que se solicitou atuação por parte da EG quanto a descargas ilegais de águas residuais em via pública. Para o mesmo intervalo de tempo, indicaram-se 7 locais de consumo associados a estas ocorrências.

Na figura abaixo é possível verificar a representação geográfica das solicitações obtidas para todos os tópicos previamente descritos.

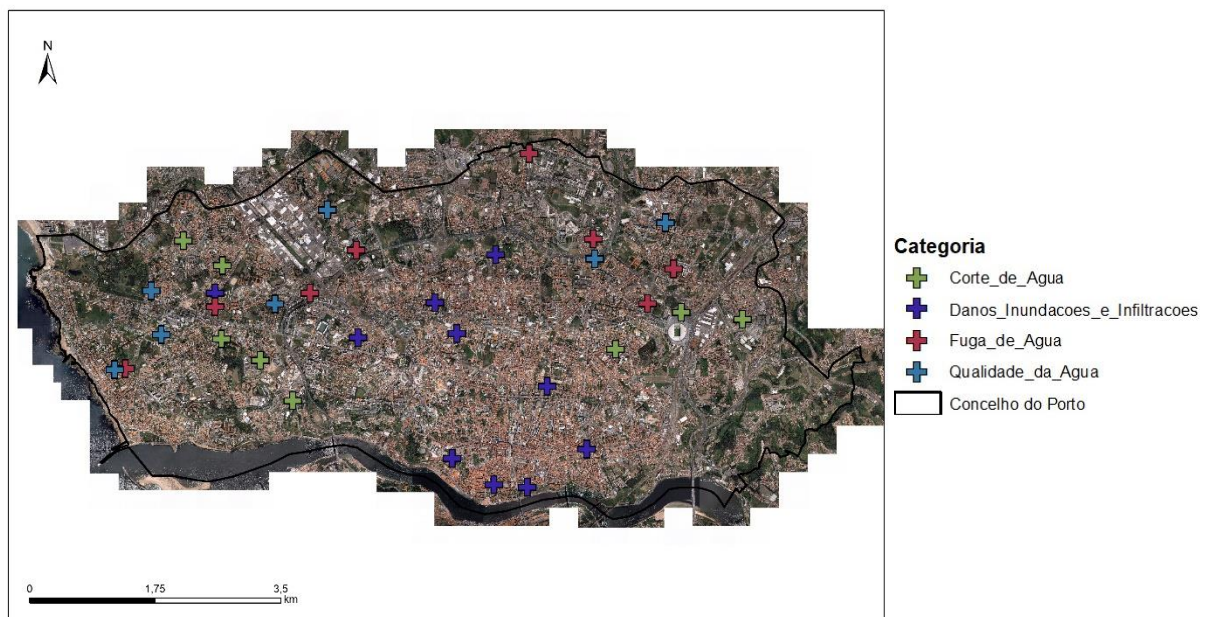


Fig. 5. 8 - Georreferenciação das solicitações do tipo restantes categorias

Não seguem propriamente uma tendência, no entanto na zona mais a oeste é onde se constam mais os pedidos efetuados. Segundo a figura, não há quaisquer pedidos na zona mais central da cidade referentes a cortes de água cuja maioria situa-se na zona do Fluvial carregada de bairros como o é o caso da Pasteleira, Aleixo, Mouteira, Rainha Dona Leonor, entre outros, sendo o mais recorrente nestes locais.

5.3.7. REPRESENTAÇÃO DE RECLAMAÇÕES DAS VÁRIAS CATEGORIAS

Para além da análise de solicitações, também se efetuou uma distribuição das reclamações recebidas via correio eletrónico pelas várias categorias com base no Manual de Procedimentos para a Gestão de Reclamações da Águas do Porto, documento esse mencionado no início deste capítulo.

Foram representadas 115 reclamações recolhidas ao longo do caso prático. O procedimento envolveu todas as interações que tivessem como assunto “Reclamação” e/ou indícios de indignação no relato do cliente. Todavia, foram excluídas 2 reclamações por dados insuficientes do contestante e 5 reclamações como consequência da consulta dos pressupostos admitidos pela empresa, porque não são consideradas para efeitos de classificação as situações expostas a outras entidades, tais como a DECO, Provedor da Justiça, PSP, Portal da Queixa, Câmara Municipal do Porto, etc., pedidos de serviço, relatos de meros prejuízos ou documentos não autenticados por falta de assinatura ou identificação do reclamante. Dos casos identificados, 2 foram desclassificados por exibição à DECO e 3 por se dirigirem ao Balcão de Atendimento Virtual da CMP.

Na Figura 5.9 apresentam-se as reclamações sinalizadas por CIL no município.

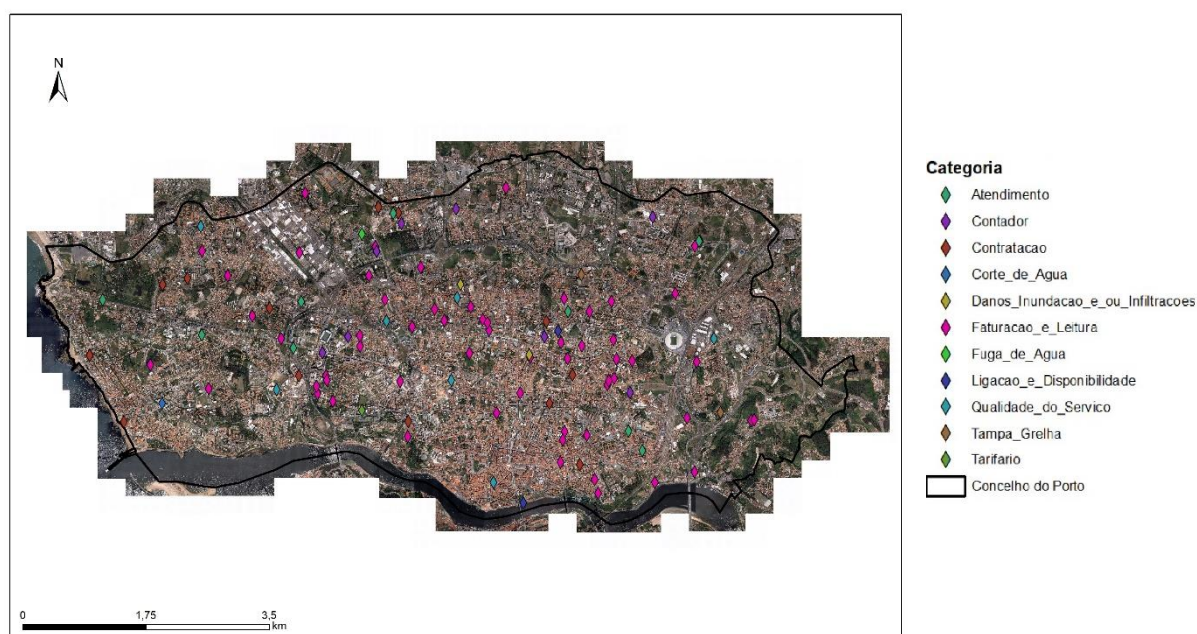


Fig. 5. 9 - Georreferenciação de reclamações de várias tipologias

A categoria de faturação e leitura salta logo à primeira vista com mais reclamações indicadas, seguindo-se aquelas respeitantes à contratação. Habitualmente, o que motiva o cliente a apresentar as reclamações em torno destas áreas são a discordância dos valores faturados, no caso da faturação e leitura, e a tarifa aplicada, no caso da contratação.

5.3.8. SOLICITAÇÕES NÃO COBERTAS PELO SIG

Aquando do processo de análise de emails, constataram-se inúmeras situações que obrigou a um registo temporário na base de dados dada a falta de informação fornecida pelos clientes no contacto com a empresa, sendo depois complementada consoante a classificação proposta.

Consistiram em solicitações de assuntos diversos, uns sendo exequível a sua categorização e respetiva contabilização, enquanto outros foram anotados consoante o pedido e entidade específica.

Na Tabela 5.8 apresentam-se as informações obtidas de emails recebidos durante o período de estudo das solicitações via correio eletrónico.

Tabela 5. 8 – Número de solicitações das várias tipologias rejeitadas por dados insuficientes

Categoria	Número de solicitações
Faturação e Leitura	55
Ligação e Disponibilidade	541
Prestação de Serviço Técnico	33
Tarifário	5
Qualidade da Água	3
Contador	45
Pavimentos	1
Tampa/Grelha	13
Fuga de Água	7
Danos – Inundação e ou Infiltrações	8
Obras	35
ETAR	2
Outras	2

Averiguou-se um elevado número de pedidos efetuados no âmbito da ligação e disponibilidade sem quaisquer dados que permitissem alcançar os dados dos clientes, uma vez que corresponderiam a locais sem quaisquer contratos ou por omissão do requerente em que maioritariamente os pedidos são de plantas topográficas e de ligação às redes públicas de abastecimento de água, águas residuais domésticas e águas pluviais.

Quanto à faturação e leitura, muitas das rogativas correspondem a aspetos de faturação da parte de ligação e disponibilidade, nomeadamente, faturas e pagamento no que concerne às PT, deferimento de projetos e reembolso dos depósitos de garantia subsequente da aprovação de projetos de redes prediais.

Na tipologia de contadores, o seu alto número inscrito deve-se ao levantamento e colocação de contadores, peculiarmente em situações de obras.

Para a prestação de serviços técnicos, observou-se um número significativo de pedidos normalmente relacionados com intervenções em via pública, pelo que os requisitantes não providenciaram outro tipo de informação para além da sua identificação civil.

A quantificação das solicitações do tipo obras foram fundamentalmente em termos de regularizações de anomalias no pavimento consequente de intervenções por parte da Águas do Porto e esclarecimentos de obras conduzidas pela mesma.

Para além das solicitações tipificadas, surgiram outras de natureza diversa que apesar de não conterem informações acerca do cliente, acabaram por ser contabilizadas para uma análise mais pormenorizada. Na Tabela 5.9 reúnem-se os vários assuntos abordados pelos clientes ou outras entidades.

Tabela 5. 9 - Número de solicitações sem categoria atribuída

Sem categoria definida	Número de solicitações
Identificação insuficiente do cliente	367
Ponto de situação/reforço do pedido	327
Cliente de outro município	15
Pedido de informação da Domus Social	31
Pedidos CMP	7
Pedidos de informação de autoridades (GNR, PJ e PSP)	7
Pedidos de informação do Tribunal Judicial e Procuradoria da República da Comarca do Porto	2
Reclamações	3
Elogios	4

Dos dados recolhidos, ressalta o elevado número de pedidos em que o requerente não se identificou de forma adequada (por ausência do CIL, NIF, n.º cliente, n.º de conta ou informação escassa) perfazendo na íntegra 367 as solicitações, em que constaram 12 instalações sem registo ou sem contratos ativos. Nas primeiras semanas do período analisado, a maioria das solicitações que não continham dados eram essencialmente acerca do acesso ao Balcão Digital (altura em que o conceito foi renovado), no entanto ao longo de tempo também se verificou para outros temas como a celebração de contrato, rescisão, débito direto e fatura eletrónica.

Os pedidos de informação da Domus Social efetuaram-se com o intuito desta entidade obter informações sobre contratos ativos em determinadas moradas geridas no âmbito da sua atividade.

Algumas das solicitações provaram ser ambíguas ou incoerentes, obrigando por vezes a uma atenção redobrada dada a complexidade do teor das mensagens. Vários emails expuseram ainda no seu conteúdo mais de 1 solicitação.

Para além disso, foram anotados elogios relativamente ao atendimento telefónico, aos serviços de saneamento e de abastecimento de águas, assim como à Sala de Comando. Num cenário menos favorável, assinalaram-se 2 reclamações relativos a ligações às redes AA e AR e ao orçamento requerido para uma execução de ramal AA e outras 2 foram descartadas por carência de informação do cliente – indicação de n.º cliente incorreto ou nenhum contato ativo para o local.

Muitos clientes pediram por diversas ocasiões confirmação da recção do email enviado, pelo que a fase seguinte desta dissertação apresenta-se como uma solução ao tempo despendido na interação do cliente com a empresa e vice-versa. Isto também poderá ter sido estimulado pelo contacto a solicitar o ponto de situação de um processo ou a reforçar um pedido realizado anteriormente, registando-se cerca de 327 mensagens deste carácter.

Outra situação observada foi o facto de terem sido estabelecidos contactos em inglês e espanhol, normalmente, em representação de clientes do setor empresarial, contudo foi uma porção residual de um cosmo mais amplo.

Ainda se comprovou por diversas vezes solicitações de clientes que pretendiam dirigir-se a um colaborador específico da AdPorto, no seguimento de contactos previamente efetuados com a empresa.

5.4. CÓDIGO DESENVOLVIDO EM APRENDIZAGEM SUPERVISIONADA PARA RESPONDER A SOLICITAÇÕES VIA E-MAIL

5.4.1. APLICAÇÃO DOS VÁRIOS CLASSIFICADORES COM ATRIBUIÇÃO DE RESPOSTAS TIPO GENÉRICAS

5.4.1.1. Conceção da base de dados

A base de dados que serviu de suporte à elaboração deste código foi composta pelos emails analisados no mesmo período da georreferenciação de solicitações. De um universo de 6038 emails que se encontram divididos pelos endereços `geral@aguasdoporto.pt` e `cliente@aguasdoporto.pt`, foram eleitas 449 mensagens das mais variadas matérias e criou-se 3 colunas com as denominações “Nome_ficheiro”, “Conteúdo” e “Categoria” correspondente.

Para esta aplicação, optou-se por um modelo sintético em concordância com a proposta para a classificação das solicitações, de maneira a otimizar o código sendo recomendável a utilização de poucas categorias com base no tamanho do *dataset*. Isto seria mais flexível se fossem trabalhados milhares de exemplos, algo comum no campo da ciência de computação. A tipologia “Tarifário” foi assim adicionada à categoria de Faturação e Leitura (Cat3) e agrupou-se todas as categorias de menor incidência numa só, entre elas, as categorias de Qualidade da Água, Corte de Água, Tampa/Grelha, Fuga de Água, Danos, Pavimentos, Obras, Ribeiras e Praias, ETAR e Outras (Cat6). A última categoria refere-se às reclamações que surge individualmente dada a sua abrangência.

Na Tabela 5.10 encontram-se expostas as categorias e a sua identificação na construção do código.

Tabela 5. 10 - Agrupamento das categorias e respetiva designação

Identificação da categoria	Categoria
Cat1	Atendimento
Cat2	Contratação
Cat3	Faturação, Leitura e Tarifário
Cat4	Prestação de Serviços Técnicos
Cat5	Ligação e Disponibilidade
Cat6	Todas as outras Categorias
Cat7	Reclamações

A tabela anterior serviu de base ao longo deste projeto, cujo objetivo consistiu em verificar se é possível, através da análise do texto de cada mensagem, atribuir cada mensagem a uma categoria. Depois

construiu-se um classificador numa última fase da classificação do texto antes de avançar propriamente para a preparação da resposta automatizada.

5.4.1.2. Análise de dados

O primeiro passo da análise de dados passou por implementar várias bibliotecas de *software* essenciais à fase inicial do código. A biblioteca “pandas” aplica-se à manipulação e análise de dados, o “matplotlib” para plotagem produzindo figuras de alta qualidade, o módulo “pickle” é empregado na conversão de um objeto *Python* (lista, dicionário, etc.) num fluxo de caracteres, o “seaborn” como uma biblioteca de visualização de dados alicerçada no “matplotlib” e por último, o “warnings” para suprimir avisos que possam surgir no programa. Na linha “sns.set_style(“whitegrid”)” é fixado um tema branco no painel dos gráficos representados:

```
import pandas as pd
import matplotlib.pyplot as plt
import pickle
import seaborn as sns
sns.set_style("whitegrid")
import warnings
warnings.filterwarnings("ignore")
```

De seguida, é invocado o *dataset* e aprontam-se os parâmetros imprescindíveis à representação do gráfico com o número de emails distribuídos por cada categoria, estabelecendo os eixos e as características das barras apresentados na Figura 5.10:

```
plt.figure(figsize=(12.8,6))
sns.countplot(df['Categoria'].sort_values())
plt.ylabel('Número de emails')
plt.xlabel('Categoria')
plt.title('Número de emails em cada categoria')
plt.show()
```

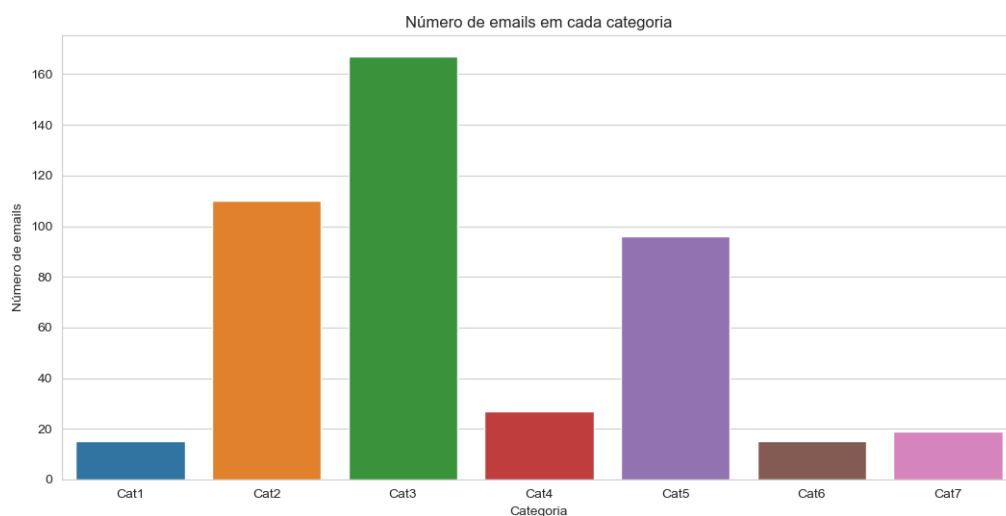


Fig. 5. 10 - Representação do número de emails afetos a cada categoria

Como se pode notar, a categoria mais popular é a Cat3 alusiva à faturação e leitura (incluindo também o tarifário), seguida pela Cat 2 que diz respeito à contratação.

Elaborou-se um outro gráfico, desta vez com a percentagem de emails que se inserem em cada categoria. Primeiro, a contagem das colunas foi gerada para cada linha e seguiu-se a definição dos parâmetros fundamentais à representação do gráfico (Figura 5.11):

```
df['Count'] = 1
df2 = pd.DataFrame(df.groupby('Categoria').count()['Count']).reset_index()
df2['Percentage'] = df2['Count'] / df2['Count'].sum()
plt.figure(figsize=(12.8,6))
sns.barplot(x=df2['Categoria'], y=df2['Percentage'])
plt.ylabel('% de emails em cada categoria')
plt.xlabel('Categoria')
plt.title('Percentagem de emails em cada categoria')
plt.show()
```

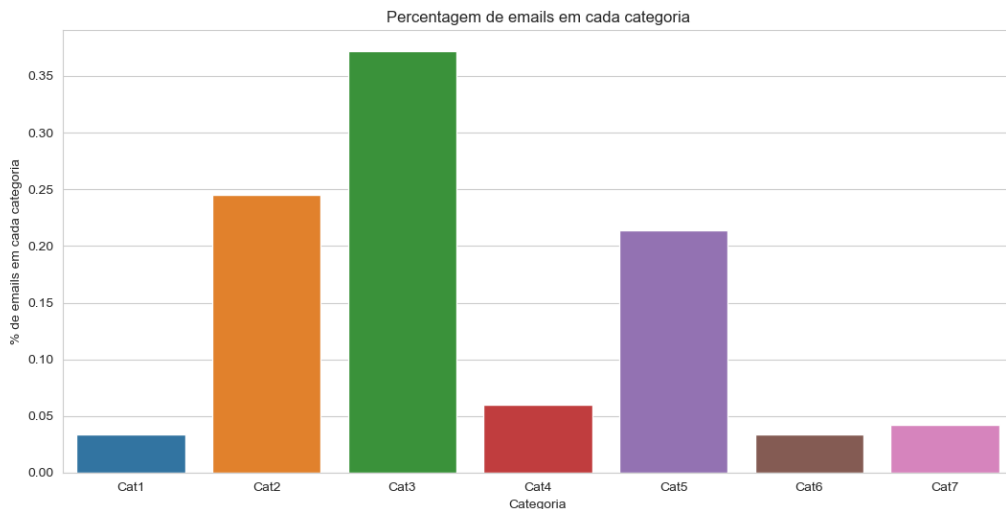


Fig. 5. 11 - Percentagem de emails repartidos pelas categorias

Aqui cingiu-se pela mesma tendência que o gráfico anterior, verificando-se novamente Cat3 e Cat2 como as mais usuais nas solicitações analisadas. Resolveu-se então mostrar a distribuição das dimensões dos emails sob outro formato (Figura 5.12). Para tal, foram configurados os novos parâmetros e passou-se à respetiva demonstração. No último passo desta etapa, o “pickle” desempenhou a função de guardar objetos:

```
df['Tamanho_emails'] = df['Conteúdo'].str.len()
plt.figure(figsize=(12.8,6))
sns.distplot(df['Tamanho_emails'])
plt.ylabel('Distribuição')
plt.xlabel('Tamanho dos emails')
plt.title('Distribuição do tamanho dos emails')
plt.show()
```

```
with open('df.pickle', 'wb') as output:
    pickle.dump(df, output)
```

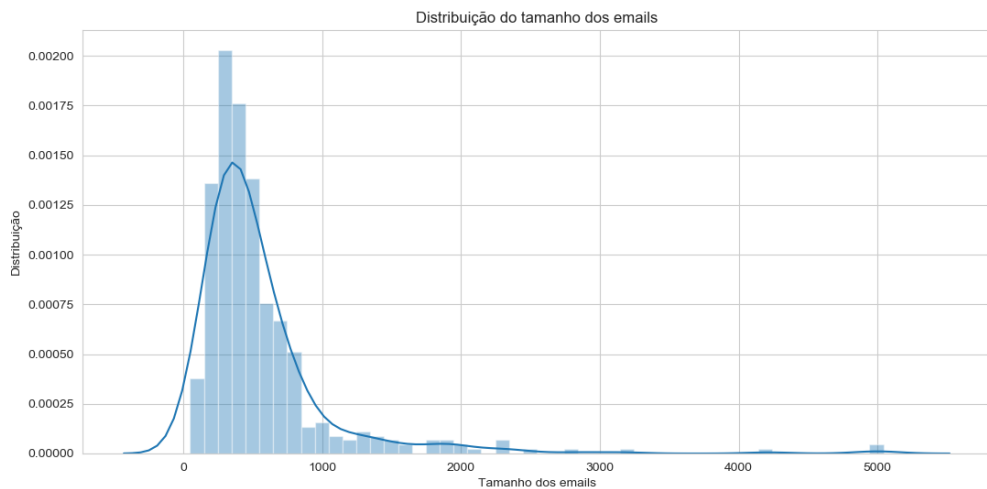


Fig. 5. 12 - Distribuição dos tamanhos dos emails presentes na base de dados

Logo foi numa distribuição normal onde se averiguaram os tamanhos de emails entre os 200-500 caracteres como os mais recorrentes, pelo que a distribuição mais elevada registou um valor na ordem dos 0,00200.

5.4.1.3. Tratamento de texto

Neste estágio começou-se por importar as bibliotecas precisas ao longo da mesma, surgindo de novo o “pickle” e o “nltk”, uma ferramenta que realiza o processamento de linguagem natural. Ocorreu a lematização que geralmente faz uso de um vocabulário e análise morfológica de palavras, ou seja, uma normalização das palavras e o conceito de “TfidfVectorizer” passa pela conversão de texto em vetores de frequência de palavras. Ainda foi usado um “train_test_split” para separar matrizes em subconjuntos de teste e treino e procedeu-se à abertura do objeto guardado em “pickle”:

```
import pickle
import nltk
from nltk.stem import WordNetLemmatizer
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split

path_df = 'df.pickle'

data = open(path_df, 'rb')
df = pickle.load(data)
data.close()
```

Depois realizou-se uma preparação e limpeza do texto retirando todos os espaços em branco e parágrafos e sucedeu-se a transformação de letras maiúsculas em minúsculas de forma a uniformizar o texto, assim como a remoção de quaisquer sinais de pontuação:

```

dff['Conteúdo_Analisado_1'] = df['Conteúdo'].str.replace('\r', ' ')
dff['Conteúdo_Analisado_1'] = df['Conteúdo_Analisado_1'].str.replace('\n', ' ')
dff['Conteúdo_Analisado_1'] = df['Conteúdo_Analisado_1'].str.replace(' ', '')

dff['Conteúdo_Analisado_2'] = dff['Conteúdo_Analisado_1'].str.lower()

sinais_pontuação = list("?!.,;")
dff['Conteúdo_Analisado_3'] = dff['Conteúdo_Analisado_2']

for sinal_pontuação in sinais_pontuação:
    dff['Conteúdo_Analisado_3'] = dff['Conteúdo_Analisado_3'].str.replace(sinal_pontuação, "")

dff['Conteúdo_Analisado_4'] = dff['Conteúdo_Analisado_3'].str.replace("'", "")

```

Foi descarregado o “punkt” e “wordnet” do complemento NLTK, que executa a tokenização dividindo o texto numa lista de frases. Guardou-se o lematizado sob a forma de objeto, gerou-se uma lista vazia que continha as palavras lematizadas e zelou-se o texto juntamente com as suas palavras sob a forma de objeto. Foram iteradas as palavras para lematizar, sendo depois adicionadas à lista criada e posteriormente acrescentadas à lista que continha os textos lematizados:

```

nltk.download('punkt')
print("-----")
nltk.download('wordnet')

wordnet_lematizador = WordNetLemmatizer()

n_linhas = len(df)
lista_texto_lematizado = []

for linha in range(0, n_linhas):

    lista_lematizada = []

    texto = df.loc[linha]['Conteúdo_Analisado_4']
    palavras_texto = texto.split(" ")

    for palavra in palavras_texto:
        lista_lematizada.append(wordnet_lematizador.lemmatize(palavra, pos="v"))

    texto_lematizado = " ".join(lista_lematizada)

    lista_texto_lematizado.append(texto_lematizado)

dff['Conteúdo_Analisado_5'] = lista_texto_lematizado

```

Descarregou-se a lista das “stopwords” em português (ver Anexo 1), palavras essas sem significado aparente que seguramente podem ser excluídas, não afetando de todo a coerência da frase:


```

nltk.download('stopwords')

stop_words = nltk.corpus.stopwords.words('portuguese')
stop_words[0:10]
len(stop_words)
'a' in stop_words
print(stop_words)

df['Conteúdo_Analisado_6'] = df['Conteúdo_Analisado_5']

for stop_word in stop_words:
    exp_stopword = r"\b" + stop_word + r"\b"
    df['Conteúdo_Analisado_6'] = df['Conteúdo_Analisado_6'].str.replace(exp_stopword, "")

df.loc[5]['Conteúdo']
df.loc[5]['Conteúdo_Analisado_1']
df.loc[5]['Conteúdo_Analisado_2']
df.loc[5]['Conteúdo_Analisado_3']
df.loc[5]['Conteúdo_Analisado_4']
df.loc[5]['Conteúdo_Analisado_5']
df.loc[5]['Conteúdo_Analisado_6']
df.head(1)
lista_colunas = ["Nome_ficheiro", "Categoria", "Conteúdo", "Conteúdo_Analisado_6"]
df = df[lista_colunas]

df = df.rename(columns={'Conteúdo_Analisado_6': 'Conteúdo_Analisado'})

```

O texto foi representado com base na selecção de parâmetros, sujeitando-se a uma transformação a nível do teste e do treino. O método “fit_transform” aprende o vocabulário e a frequência inversa do documento (IDF) retornando a matriz documento-termo e, por outro lado, “transform” metamorfoseia documentos numa matriz documento-termo:

```

corpus_train, corpus_test, y_train, y_test = train_test_split(df['Conteúdo_Analisado'],
df['Categoria'], test_size=0.2, random_state=8, stratify=df['Categoria'])

ngram_range = (1,1)
min_df = 1
max_df = 1.
max_features = None

tfidf = TfidfVectorizer(encoding='utf-8', ngram_range=ngram_range, stop_words=None,
lowercase=False, max_df=max_df, min_df=min_df, max_features=max_features, norm='l2',
sublinear_tf=False)

X_train = tfidf.fit_transform(corpus_train).toarray()
print(X_train.shape)

X_test = tfidf.transform(corpus_test).toarray()
print(X_test.shape)

```

No final, os parâmetros “y_train”, “y_test”, “x_train” e “x_test” foram guardados sob o modo “pickle”, tal como o objeto TF-IDF, conforme se pode averiguar no código presente.

Do conjunto de dados de treino, existiam 359 elementos com 5628 características (extraídas para realizar o TF-IDF que correspondem a um grupo de palavras do *dataset*). Quanto aos dados de teste, foram 90 elementos com as mesmas características.

5.4.1.4. Classificadores

A plataforma *sklearn* dispõe de várias tipologias de classificadores. Apesar de constarem inúmeros na documentação existente, foram abordados os seguintes: *Linear SVC*, *Bagging*, *RandomForest*, *Multi-layer Perceptron*, *Stochastic Gradient Descent*, *DecisionTree*, *ExtraTree*, *KNeighbors*, *BernoulliNB*, *Gaussian NB*, *MultinomialNB* e *ComplementNB*.

Primeiro escolheu-se a Classificação de Suporte Vetorial Linear (*LinearSVC*) para exemplificar a apresentação de resultados na consola. Constitui um dos módulos da Máquina de Vetores de Suporte (SVM), cuja característica é a elevada flexibilidade na escolha de penalidades e em funções de perda e considerado o melhor em termos de resultados segundo a literatura:

```
import pickle
from sklearn.svm import LinearSVC
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
import matplotlib.pyplot as plt
import seaborn as sns
```

Seguidamente, foram abertos os vários objetos “pickle” dos parâmetros guardados antes. Introduziu-se o seguinte comando para aplicar o classificador tendo em conta às suas especificações:

```
clf = LinearSVC(penalty='l2', loss='squared_hinge', dual=True, tol=0.0001, C=1.0,
multi_class='ovr', fit_intercept=True, intercept_scaling=1, class_weight=None, verbose=0,
random_state=None, max_iter=1000)
clf.fit(X_train, y_train)
```

Testou-se o classificador, o que consistiu em treiná-lo com um conjunto de treino e aplicá-lo seguidamente a um conjunto de teste. Neste caso, está-se perante uma classificação multi-classe que se distingue da binária, do tipo mais simples de problema envolvendo ML, à qual compete a tarefa de classificar os elementos de uma base de dados em duas classes. Por exemplo, olhos azuis ou olhos não azuis, verdadeiro ou falso, estão entre outros exemplos mais comuns.

Assim, foi possível efetuar a previsão e, em seguida, originaram-se os resultados de classificação obtidos quanto à:

- Exatidão, que se apresenta sob a forma $VP + FV / TP + FP + FN + TN$, em que VP é o número de Verdadeiros Positivos e FV o número de Falsos Verdadeiros;
- Precisão, que se expressa pela razão $VP / (VP + FP)$ onde VP é o número de Verdadeiros Positivos e FP o número de Falsos Positivos;
- Sensibilidade (*Recall*, em inglês), que se traduz pela relação entre $VP / VP + FN$ em que VP é o número de Verdadeiros Positivos e FN o número de Falsos Negativos;
- Medida F1, que consta como média ponderada da precisão e da sensibilidade em que $2 \times (Sensibilidade \times Precisão) / (Sensibilidade + Precisão)$.

```

clf_pred = clf.predict(X_test)

print("Exatidão do treino: ")
print(accuracy_score(y_train, clf.predict(X_train)))

print("Exatidão do teste: ")
print(accuracy_score(y_test, clf_pred))

print("Relatório de classificação")
print(classification_report(y_test, clf_pred))

```

Tabela 5. 11 - Performance do classificador Linear SVC

Categoria	Precisão	Sensibilidade	F1	Exemplos de suporte
Cat1	1.00	0.67	0.80	3
Cat2	0.75	0.68	0.71	22
Cat3	0.69	0.91	0.78	34
Cat4	0.00	0.00	0.00	5
Cat5	0.85	0.89	0.87	19
Cat6	1.00	0.33	0.50	3
Cat7	0.50	0.25	0.33	4
Micro-média	-	-	0.74	90
Macro-média	0.68	0.53	0.57	90
Média ponderada	0.71	0.74	0.71	90

Segundo os resultados apresentados na consola, a exatidão do treino rondou os 1.0 (100%) e a exatidão do teste foi avaliada no equivalente a 74%. De notar que os melhores valores correspondem a 1 e os piores a 0.

O método da micro-média refere-se ao somatório individual dos verdadeiros positivos, falsos positivos e falsos negativos do sistema para diferentes conjuntos, enquanto a macro-média define-se como a média da precisão e da sensibilidade do sistema em diferentes conjuntos. O primeiro método foi o que mostrou melhores resultados com 0.74.

As várias categorias revelaram na generalidade bom desempenho, todavia as tipologias Cat4 e Cat7 tiveram resultados menos favoráveis, o que pode estar correlacionado com o facto de terem menor número de exemplos na base de dados. Como se constata, os exemplos de suporte foram, respetivamente, 5 e 3.

A matriz que se segue na Figura 5.13 representa uma matriz confusão para a classificação multi-classe, culminando no resguardo do ficheiro “pickle”:

```

labels = df['Categoria'].drop_duplicates().sort_values()
conf_matrix = confusion_matrix(y_test, clf_pred)
plt.figure(figsize=(12,8))
sns.heatmap(conf_matrix, annot=True, xticklabels=labels, yticklabels=labels, cmap="Blues")
plt.ylabel('Verdadeiro')
plt.xlabel('Previsto')
plt.title('Matriz de confusão')
plt.show()

```

```

with open('clf.pickle', 'wb') as output:
    pickle.dump(clf, output)

```

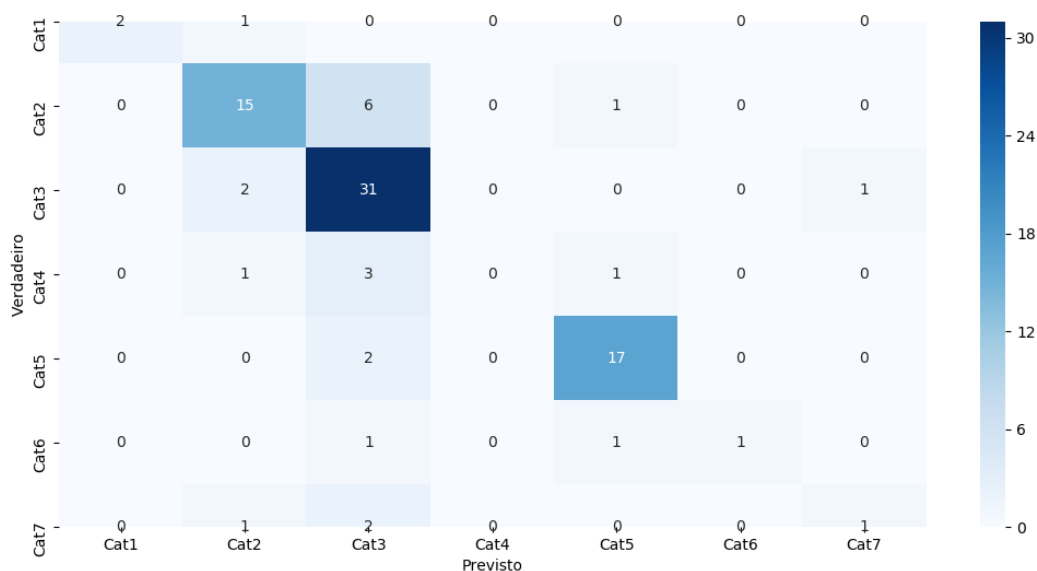


Fig. 5. 13 - Matriz de confusão obtida para o classificador *Linear SVC*

A interpretação da matriz de confusão pode ser feita da seguinte forma: os resultados previstos são lidos no eixo dos xx, isto é, em cada coluna, enquanto os resultados reais ou verdadeiros constam do eixo dos yy, ou seja, são consultados em cada linha. Portanto, na diagonal onde se cruzam as categorias similares, encontram-se os resultados que corresponderam às expectativas (Verdadeiro Positivo – TP). Quanto mais escura a célula, maior o valor obtido, tal como se averigua para Cat2 – Cat2 com 15 exemplos, Cat3 – Cat3 com 31 exemplos e Cat5 – Cat5 com 17 exemplos assinalados. Para a Cat4 não foram sinalizadas quaisquer amostras bem sucedidas, devido ao número reduzido de solicitações recolhidas para o treino e teste do *dataset*.

Identificaram-se vários Falsos Positivos (FP), em que através da categoria verdadeira acaba por se prever a categoria errada. São exemplos disso as associações entre Cat4 (Verdadeiro) – Cat2 (Previsto) que resultou em 1 amostra mal atribuída e Cat3 (Verdadeiro) – Cat2 (Previsto) que somou 2 incorretas. As amostras com resultados Falso Verdadeiro (TN) e Falso Negativo (FN) não foram demonstradas na matriz de confusão, já que dificilmente se detetam. No primeiro caso, foi impossível demonstrar pois

não é aplicável à situação. Se a classificação fosse binária, aí facilmente seriam revelados por haverem unicamente duas classes.

Comparado com o esboço do código transato, a única linha que se alterou no momento da importação foi precisamente onde surge o classificador “BaggingClassifier”:

```
import pickle
from sklearn.ensemble import BaggingClassifier
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
import matplotlib.pyplot as plt
import seaborn as sns
```

Este caracteriza-se por constituir um meta-estimador de conjunto que se encaixa nos classificadores base, cada um em subconjuntos aleatórios do conjunto de dados original e, em seguida, agrega as suas previsões individuais (por votação ou por média) para formar uma previsão final. É tipicamente usado como uma maneira de reduzir a variação de um estimador de caixa preta (por exemplo, uma árvore de decisão), introduzindo a randomização no seu procedimento de construção e elaborando um conjunto a partir do mesmo. Este algoritmo advém da junção de várias obras presentes na literatura.

Outro excerto de código afeto ao classificador corresponde aos parâmetros que o caracterizam:

```
clf = BaggingClassifier(base_estimator=None, n_estimators=10, max_samples=1.0,
max_features=1.0, bootstrap=True, bootstrap_features=False, oob_score=False,
warm_start=False, n_jobs=None, random_state=None, verbose=0)
```

Executou-se o código atendendo ao classificador *Bagging* e obteve-se o seu relatório de classificação. Nisto foi possível constatar que apresenta um desempenho razoável, cuja exatidão do treino foi de aproximadamente 98% e a exatidão do teste alcançou os 58%. Verificou-se que na categoria 7 não conseguiu prever nada, já a categoria 4, apesar do valor baixo de precisão, ainda ultrapassou a nulidade nos parâmetros de precisão, sensibilidade e medida F1 com base em 5 exemplos. A precisão do teste, que se traduz por um conjunto de mensagens que o sistema colocou corretamente numa categoria de texto entre todas as mensagens foi de 100%, 57%, 52%, 25%, 83%, 100% e 0%, respetivamente, enquanto a medida F1, apresentou valores de 80%, 56%, 62%, 22%, 65%, 50% e 0%, respetivamente. É de referir que o número de exemplos de suporte mantém-se inalterado para todos os classificadores, pelo que será o mesmo de acordo com o disposto na Tabela 5.11.

O classificador que se aplicou de seguida foi o de Floresta Aleatória (*RandomForestClassifier*), pelo que se modificou o campo associado à importação deste elemento:

```
import pickle
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
import matplotlib.pyplot as plt
import seaborn as sns
```

Comporta um meta-estimador (tem como função a combinação de modelos base para melhorar o desempenho) que ajusta vários classificadores da árvore de decisão em várias subamostras do conjunto de dados, usa a média para melhorar a precisão preditiva e controla o ajuste excessivo. O tamanho da subamostra é sempre o mesmo do tamanho da amostra de entrada original, mas as amostras são coletadas com a substituição, se a condição padrão for “bootstrap = True”:

```
clf = RandomForestClassifier(n_estimators='warn', criterion='gini', max_depth=None,
min_samples_split=2, min_samples_leaf=1, min_weight_fraction_leaf=0.0, max_features='auto',
max_leaf_nodes=None, min_impurity_decrease=0.0, min_impurity_split=None, bootstrap=True,
oob_score=False, n_jobs=None, random_state=None, verbose=0, warm_start=False,
class_weight=None)
```

Assim, “correu-se” o código com as especificações do classificador em questão e o relatório desta ação encontra-se definido na tabela seguinte. Com base nestes resultados, observou-se uma exatidão de treino de 99% e uma exatidão de teste de 57%. As categorias 4 e 7 alcançaram valores nulos, atentando ao reduzido número de exemplares e ao fraco desempenho do próprio classificador nesse cenário, enquanto por outro lado, as categorias 1 e 6 obtiveram 100% na precisão do teste.

Depois, selecionou-se o classificador *Multi-layer Perceptron (MLP)* que envolve redes neurais, efetuando a respetiva importação:

```
import pickle
from sklearn.neural_network import MLPClassifier
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
import matplotlib.pyplot as plt
import seaborn as sns
```

Este modelo otimiza a perda logarítmica usando o algoritmo de Broyden-Fletcher-Godfarb-Shanno (da sigla LBFGS) ou a descida de gradiente estocástico, ou seja, um gradiente dedicado ao cálculo das probabilidades. Deste modo, apresentam-se os parâmetros que o integram:

```
clf = MLPClassifier(hidden_layer_sizes=(200, ), activation="relu", solver="adam", alpha=0.0001,
batch_size="auto", learning_rate="constant", learning_rate_init=0.001, power_t=0.5,
max_iter=200, shuffle=True, random_state=None, tol=0.0001, verbose=False, warm_start=False,
momentum=0.9, nesterovs_momentum=True, early_stopping=False, validation_fraction=0.1,
beta_1=0.9, beta_2=0.999, epsilon=1e-08, n_iter_no_change=10)
```

Com estes parâmetros, obteve-se o seu relatório de classificação onde se comprovou uma exatidão de treino de 100% e uma exatidão de teste a rondar os 71%. Uma vez mais, as categorias 4 e 7 voltaram a adquirir valores nulos, sendo as melhores precisões correspondentes às categorias 1 e 6.

O próximo classificador em foco foi a Descida de Gradiente Estocástico (*Stochastic Gradient Descent, SGD*) que consiste numa abordagem simples, mas muito eficiente, para a aprendizagem discriminativa de classificadores lineares sob funções de perda convexa, como as Máquinas de Vetores de Suporte

```
import pickle
from sklearn.linear_model import SGDClassifier
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
import matplotlib.pyplot as plt
import seaborn as sns
```

(lineares) e Regressão Logística. As suas vantagens são a eficiência e a facilidade de implementação (muitas oportunidades para ajuste de código), todavia requer vários hiperparâmetros, como o parâmetro de regularização e o número de iterações, e é sensível ao dimensionamento de características.

Esta implementação trabalha com dados representados como matrizes densas ou esparsas de valores de ponto flutuante para as características. O modelo adequado pode ser controlado com o parâmetro de perda. Por norma, ele encaixa-se numa máquina de vetores de suporte linear (SVM).

O regularizador é uma penalidade adicionada à função de perda que reduz os parâmetros do modelo em relação ao vetor zero usando a norma euclidiana quadrada L2 ou a norma absoluta L1 ou uma combinação de ambas (*Elastic Net*). Importou-se então o classificador em questão:

```
clf = SGDClassifier(loss='hinge', penalty='l2', alpha=0.0001, l1_ratio=0.15, fit_intercept=True,
max_iter=1000, tol=0.001, shuffle=True, verbose=0, epsilon=0.1, n_jobs=None,
random_state=None, learning_rate='optimal', eta0=0.0, power_t=0.5, early_stopping=False,
validation_fraction=0.1, n_iter_no_change=5, class_weight=None, warm_start=False,
average=False)
```

Assim, obteve-se uma exatidão de treino de 100% e uma exatidão de teste de aproximadamente 69%. Novamente, voltam a surgir as categorias 4 e 7 com valores nulos, o que revela uma certa dificuldade por parte de vários classificadores em reconhecer o conteúdo associado às mesmas. O valor mais elevado para a precisão registou-se apenas no caso da categoria 1.

Desta vez, o classificador empregado foi o de Árvore de Decisão (*DecisionTree*) que se traduz num método de aprendizagem supervisionada não paramétrica usado para classificação e regressão. A importação fez-se da seguinte maneira:

```
import pickle
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
import matplotlib.pyplot as plt
import seaborn as sns
```

O objetivo é criar um modelo que prediz o valor de uma variável de destino, aprendendo regras simples de decisão inferidas a partir das características dos dados:

```
clf = DecisionTreeClassifier(criterion='gini', splitter='best', max_depth=None, min_samples_split=2,
min_samples_leaf=1, min_weight_fraction_leaf=0.0, max_features=None, random_state=None,
max_leaf_nodes=None, min_impurity_decrease=0.0, min_impurity_split=None,
class_weight=None, presort=False)
```

Atentando à sua implementação, resultou numa exatidão de treino de 100% e exatidão de teste de 49%. Apenas a categoria 6 atingiu a melhor precisão com 100%, enquanto as categorias 1 e 7 assinalaram valores nulos.

O classificador Árvore Extra implementa um meta-estimador que se encaixa em várias árvores de decisão aleatórias (as tais árvores extra) em várias subamostras do conjunto de dados e usa a média para

melhorar a precisão preditiva e controlar o erro de modelação que ocorre quando uma função é muito ajustada para um conjunto limitado de pontos de dados. Para tal, importou-se o classificador:

```
import pickle
from sklearn.tree import ExtraTreeClassifier
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
import matplotlib.pyplot as plt
import seaborn as sns
```

Um conjunto diversificado de classificadores é criado pela introdução da aleatoriedade na construção do classificador e a previsão do conjunto é dada como a previsão média dos classificadores individuais:

```
clf = ExtraTreeClassifier(criterion='gini', splitter='best', max_depth=None, min_samples_split=2,
min_samples_leaf=1, min_weight_fraction_leaf=0.0, max_features=None, random_state=None,
max_leaf_nodes=None, min_impurity_decrease=0.0, min_impurity_split=None,
class_weight=None, presort=False)
```

Logrou-se uma exatidão de treino de 100% e uma exatidão de teste de 52%. A única categoria em que se precisou 100% foi a categoria 6, porém todas as outras conseguiram obter valores maioritariamente medíocres.

A classificação baseada em vizinhos (*KNearestNeighbors*) está associada a um tipo de aprendizagem em que não tenta construir um modelo interno geral, mas que simplesmente armazena instâncias dos dados de treino. A classificação é calculada a partir de um voto maioritário simples dos vizinhos mais próximos de cada ponto: um ponto recebe a classe de dados que tem mais representantes nos vizinhos mais próximos do ponto. Como tal, sucedeu-se a importação do classificador em questão:

```
import pickle
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
import matplotlib.pyplot as plt
import seaborn as sns
```

A escolha ideal do valor *k* depende muito dos dados, geralmente um maior valor suprime os efeitos do ruído, mas torna os limites da classificação menos distintos. Em algumas circunstâncias, é melhor ponderar os vizinhos, de modo que os vizinhos mais próximos contribuam mais para o ajuste, o que pode ser efetivado pelo parâmetro “weight” que atribui pesos uniformes a cada vizinho:

```
clf = KNeighborsClassifier(n_neighbors=5, weights='uniform', algorithm='auto', leaf_size=30, p=2,
metric='minkowski', metric_params=None, n_jobs=None)
```

Logo, a exatidão do treino foi cerca de 77% e a exatidão do teste de 68%. Várias categorias apresentaram valores nulos, exceto as categorias 1 (melhor precisão), 2, 3 e 5.

Os últimos 4 classificadores inserem-se nos métodos de *Naive Bayes* que formam um conjunto de algoritmos de aprendizagem supervisionada baseados na aplicação do teorema de *Bayes* com a suposição simples de independência condicional entre cada par de características, dado o valor da variável de classe.

O classificador *BernoulliNB* implementa algoritmos de treino e classificação de Bayes para dados distribuídos, de acordo com distribuições multivariadas de *Bernoulli*, ou seja, pode haver várias características, mas cada uma delas é considerada uma variável com valor binário. Portanto, esta classe exige que as amostras sejam representadas como vetores de características com valor binário e o seu desempenho é melhor em alguns conjuntos de dados, especialmente aqueles com documentos mais curtos. A importação deste classificador foi feita do seguinte modo:

```
import pickle
from sklearn.naive_bayes import BernoulliNB
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
import matplotlib.pyplot as plt
import seaborn as sns
```

O parâmetro “binarize” pode ser determinado pelo *input* dado a este classificador, encontrando-se entre outros:

```
clf = BernoulliNB(alpha=1.0, binarize=0.0, fit_prior=True, class_prior=None)
```

Após a execução do código, constatou-se no relatório de classificação que a exatidão do treino foi de 67% e a exatidão do teste de 43%. A melhor precisão verificou-se para a categoria 5 e a maioria das categorias obtiveram valores nulos.

O classificador que se segue é o *GaussianNB* que se apresenta como um algoritmo gaussiano de *Naive Bayes* para classificação, onde a probabilidade das características assume-se como gaussiana. A importação ocorreu da seguinte maneira:

```
import pickle
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
import matplotlib.pyplot as plt
import seaborn as sns
```

Seguiu-se então a sua configuração:

```
clf = GaussianNB(priors=None, var_smoothing=1e-09)
```

O relatório de classificação resultante desta implementação detetou uma exatidão de treino de 100%, seguida por uma exatidão de teste de 54%. As categorias 4 e 7 retornaram aos valores nulos, constatando-se a melhor precisão para as categorias 1 e 6.

O classificador *MultinomialNB* é outro integrante dos *Naive Bayes* destinado a modelos multinomiais. É adequado para classificação com características discretas (por exemplo, contagem de palavras para classificação de texto), sendo que na distribuição multinomial normalmente são requeridas contagens de características inteiras. No entanto, na prática, contagens fracionárias como *TF-IDF* também podem funcionar:

```
import pickle
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
import matplotlib.pyplot as plt
import seaborn as sns

(...)

clf = MultinomialNB(alpha=1.0, fit_prior=True, class_prior=None)
```

Depois de se “correr” o código, obteve-se o relatório de classificação respeitante a este classificador, cuja exatidão de treino foi de 81% e exatidão de teste de, sensivelmente, 48%. Quanto às precisões, verificou-se apenas para as categorias 3, com 0.42, e categoria 5, com 1.00.

Para terminar o estudo dos classificadores e ainda no âmbito dos métodos de *Naive Bayes*, usou-se o classificador *ComplementNB* que aplica o algoritmo de complemento *Naive Bayes* (CNB). É uma adaptação do algoritmo multinomial NB que é particularmente apropriado para conjuntos de dados desequilibrados e usa estatísticas do complemento de cada classe para calcular os pesos do modelo. Empiricamente, as estimativas de parâmetros para o CNB são mais estáveis do que as do modelo multinomial, superando-o regularmente por uma margem considerável nas tarefas de classificação de texto, pois foi projetado com a fim de corrigir as suas limitações. Este foi importado, de acordo com o excerto que se segue:

```
import pickle
from sklearn.naive_bayes import ComplementNB
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
import matplotlib.pyplot as plt
import seaborn as sns

(...)

clf = ComplementNB(alpha=1.0, fit_prior=True, class_prior=None, norm=False)
```

Os resultados da classificação evidenciaram uma exatidão de treino de quase 100% e uma exatidão de teste de 72%, notoriamente superior ao valor obtido no classificador transato. Apresenta no geral ótimos valores de precisão, salvo as categorias 4 e 7 que indicaram dificuldades na previsão.

Com base na aplicação e análise efetuada para todos os classificadores, encontram-se discriminados os resultados principais na tabela seguinte.

Tabela 5. 12 - Perspetiva geral do desempenho dos vários classificadores

Classificador	Exatidão do teste	Macro F1
<i>Linear SVC</i>	0.74	0.57
<i>Bagging</i>	0.58	0.48
<i>RandomForest</i>	0.57	0.40
<i>Multi-layer Perceptron</i>	0.71	0.51
<i>Stochastic Gradient Descent</i>	0.69	0.47
<i>DecisionTree</i>	0.49	0.32
<i>ExtraTree</i>	0.52	0.41
<i>KNeighbors</i>	0.68	0.43
<i>BernoulliNB</i>	0.43	0.15
<i>GaussianNB</i>	0.54	0.39
<i>MultinomialNB</i>	0.48	0.18
<i>ComplementNB</i>	0.72	0.47

Como já era expectável, *Linear SVC* foi o classificador com melhores resultados obtidos na generalidade dos casos expostos, tanto pelo valor da exatidão do teste como pelo parâmetro macro F1. Não tão atrás ficaram os classificadores *ComplementNB* e *Multi-layer Perceptron* com 0.72 e 0.71, respetivamente, pelo que também seriam alternativas igualmente plausíveis. Todavia, deparou-se com uma exatidão de teste de 0.43 para o classificador *BernoulliNB* que revelou claramente não ser o mais adequado ao *dataset* utilizado, dadas as suas características.

5.4.1.5. Resposta automática

As bibliotecas importadas inicialmente, “pandas” e “pickle”, voltam a surgir de novo, seguindo por outras como “smtplib” que se dedica à função de envio à semelhança dos módulos de emails necessários. SMTP, em inglês *Simple Mail Transfer Protocol* compreende o protocolo padrão para receção e entrega de correio eletrónico através da Internet, cujo servidor aplicado a este procedimento foi o da FEUP:

```
import pandas as pd
import pickle
import smtplib
from email.mime.multipart import MIMEMultipart
from email.mime.text import MIMEText
from getpass import getpass
from sys import argv

host = 'smtp.fe.up.pt'
```

Carregou-se o objeto “pickle” e realizou-se a leitura, transformação e respetiva previsão da resposta a enviar:

```
def load_pickle(filename):
   .pkl = open(filename, 'rb')
    data = pickle.load(pk)
   .pkl.close()
    return data

def predict(corpus):
    vectorizer = load_pickle('tfidf.pickle')
    clf = load_pickle('clf.pickle')
    X = vectorizer.transform(corpus)
    y_pred = clf.predict(X)
    return y_pred
```

Para enviar o email, começou-se por definir o assunto (“subject”), destinatário (“toaddrs”), remetente (“fromaddr”) e especificações da porta do servidor, tal como o formato da mensagem:

```
def send_emails(fromaddr, password, y_pred, templates):
    subject = 'Example subject'
    toaddrs = ['ega12024@fe.up.pt']
    toaddrs_str = ', '.join(toaddrs)
    server = smtplib.SMTP(host=host, port=587)
    server.ehlo()
    server.starttls()
    server.ehlo()
    server.login(fromaddr, password)
    for pred in y_pred:
        msg = MIME multipart()
        msg['Subject'] = subject
        msg['From'] = fromaddr
        msg['To'] = toaddrs_str
        body = templates[pred]
        msg.attach(MIMEText(body, 'plain'))
        text = msg.as_string()
        server.sendmail(fromaddr, toaddrs, text)
    server.quit()
```

Aqui foi buscar parâmetros e as respostas tipo a uma base de dados que constam no Anexo 2. Depois, instruiu-se o programa para pedir a palavra-chave e desencadeou a mensagem prevista para a respetiva classe de solicitação:

```
def main():
    if len(argv) != 4:
        print("Usage: python3 %s <excel_file> <excel_column> <source_email>" % (argv[0]))
        return
    excel_file = argv[1]
    excel_column = argv[2]
    fromaddr = argv[3]
    templates = pd.read_excel('templates_das_categorias.xlsx')
    templates = dict([(templates['Categoria'][i], templates['Template'][i]) for i in
range(len(templates))])
    print("Por favor, coloque a palavra-chave.")
    password = getpass()
    df = pd.read_excel(excel_file)
    corpus = df[excel_column]
    y_pred = predict(corpus)
    send_emails(fromaddr, password, y_pred, templates)

if __name__ == "__main__":
    main()
```

Para testar o mecanismo da resposta automática, foram introduzidos os diretórios dos ficheiros que constituem o código por ordem de execução na linha de comandos:

```
Anaconda Prompt - python _4_Resposta_Automatica.py teste_emails_2.xlsx "Conteúdo" "ega12024@fe.up.pt"
(359, 5628)
(90, 5628)
Exatidão do treino:
1.0
Exatidão do teste:
0.7444444444444445
Relatório de classificação
C:\Users\leandrarocha.AGUASDOPORTO\AppData\Local\Continuum\anaconda3\envs\leandra_thesis\lib\site-packages\sklearn\me
trics\classification.py:1437: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in l
abels with no predicted samples.
'precision', 'predicted', average, warn_for)
precision    recall  f1-score   support

   Cat1       1.00    0.67    0.80         3
   Cat2       0.75    0.68    0.71        22
   Cat3       0.69    0.91    0.78        34
   Cat4       0.00    0.00    0.00         5
   Cat5       0.85    0.89    0.87        19
   Cat6       1.00    0.33    0.50         3
   Cat7       0.50    0.25    0.33         4

 accuracy          0.74         90
 macro avg       0.68    0.53    0.57         90
weighted avg       0.71    0.74    0.71         90

(leandra_thesis) C:\Users\leandrarocha.AGUASDOPORTO\Desktop\Tese Leandra\Tese Python>python _4_Resposta_Automatica.py
teste_emails_2.xlsx "Conteúdo" "ega12024@fe.up.pt"
Por favor, coloque a palavra-chave.
Password:
```

Fig. 5. 14 - Execução do comando para gerar resposta automática

Após a colocação da palavra-passe afeta ao endereço fixado como remetente, foi acionada a ação do envio do correio eletrónico:



Fig. 5. 15 - Receção do email no destinatário indicado

O código pode ser consultado na íntegra no Anexo 3.

5.4.2. APLICAÇÃO DE VÁRIOS CLASSIFICADORES COM ATRIBUIÇÃO DE RESPOSTAS TIPO PERSONALIZADAS

5.4.2.1. Conceção da base de dados

Criou-se um novo *dataset* com o mesmo conjunto de emails, atribuindo as novas classes idealizadas em redor das temáticas mais solicitadas pelos clientes. Assim, dentro das classes pré-definidas distinguem-se várias subclasses. A saber: Cat11 (Balcão Digital); Cat12 (Resposta genérica para a Categoria 1); Cat21 (Celebração de contrato); Cat22 (Alteração de titularidade); Cat23 (Rescisão de contrato); Cat24 (Alteração da morada de envio de faturação); Cat25 (Resposta genérica para a Categoria 2); Cat30 (Resposta genérica para a Categoria 3); Cat31 (Adesão ao débito direto); Cat32 (Alteração de IBAN/NIB); Cat33 (Adesão à fatura eletrónica); Cat34 (Alteração de email); Cat35 (2ª via de faturas); Cat36 (Análise/Correção de faturas); Cat37 (Comunicação de leituras); Cat38 (Extrato de conta corrente); Cat39 (Nota de crédito); Cat310 (Regularização de pagamentos); Cat41 (Serviço de saneamento); Cat42 (Instalação de contador); Cat43 (Serviço de abastecimento de água); Cat44 (Serviço de águas pluviais); Cat51 (Planta topográfica); Cat52 (Autorização de utilização das redes prediais); Cat53 (Projeto de redes prediais); Cat54 (Ligações às redes públicas); Cat55 (Comunicação de início de obras); Cat56 (Resposta genérica para a Categoria 5); Cat57 (Verificação de abastecimento de água/contador); Cat61 (Tampa/Grelha); Cat62 (Pavimentos); Cat63 (Obras); Cat64 (Religação na sequência de corte de água); Cat65 (Fuga de água); Cat66 (Danos relacionados com inundações e/ou infiltrações); Cat67 (Resposta genérica para a Categoria 6) e por fim, Cat70 (Reclamações).

5.4.2.2. Análise de dados

Neste tópico recorreu-se de novo ao código anterior, em que apenas diferem alguns passos executados. Entretanto alterou-se o *dataset* para o seguinte:

```
df_path = 'teste_emails_11.xlsx'
```

De seguida, fez-se um ajuste dos parâmetros dos gráficos em função do elevado número de categorias, aumentando o comprimento da figura de modo a facilitar a visualização de todas as classes abrangentes. Então modificou-se o comprimento de 12.8 para 30, cujas representações podem ser consultadas no Anexo 4.

Uma vez que a fase de tratamento de texto não sofreu quaisquer alterações, passou-se imediatamente para a fase de classificação.

5.4.2.3. Classificadores

À semelhança do procedimento realizado anteriormente, apenas foi alterado no código o excerto correspondente à importação e mais à frente no estágio de classificação. Visto que as linhas de importação mantiveram-se, procedeu-se à parte experimental que envolve todos os classificadores já conhecidos.

Os resultados dos relatórios de classificação para cada classificador encontram-se especificados na Tabela 5.13.

Tabela 5. 13 - Resultados obtidos para os diversos classificadores

Classificador	Exatidão do treino	Categorias com melhor precisão	Categorias com parâmetros nulos
<i>Linear SVC</i>	1.0	25, 31, 32, 33, 39, 41, 54, 55	12, 42, 44, 52, 53, 65, 67
<i>Bagging</i>	0.99	310, 39, 55	12, 25, 33, 36, 42, 44, 52, 53, 54, 57, 65, 67
<i>RandomForest</i>	0.99	32, 38, 41, 55, 57	12, 25, 30, 34, 37, 42, 44, 52, 53, 54, 65, 67
<i>Multi-layer Perceptron</i>	1.0	11, 23, 24, 25, 33, 37, 38, 39, 41, 55	12, 34, 42, 44, 52, 53, 54, 56, 57, 65, 67
<i>Stochastic Gradient Descent</i>	1.0	22, 310, 37, 39, 41, 55, 56	12, 21, 25, 30, 33, 34, 36, 38, 42, 44, 52, 53, 54, 57, 65, 67

<i>DecisionTree</i>	1.0	24, 41, 55	11, 12, 25, 30, 3, 34, 35, 36, 37, 44, 52, 53, 57, 65, 67
<i>ExtraTree</i>	1.0	55	11, 12, 21, 310, 32, 34, 37, 39, 42, 44, 52, 53, 54, 57, 65, 67, 70
<i>KNeighbors</i>	0.58	37, 39, 54, 55, 57	12, 21, 25, 30, 32, 33, 42, 44, 52, 53, 65, 67
<i>BernoulliNB</i>	0.16	Nenhuma	Todas nulas, excepto 51 com 0.08 de precisão
<i>GaussianNB</i>	1.0	11, 24, 32, 33, 38, 41, 55	12, 21, 25, 310, 34, 36, 37, 42, 44, 52, 53, 54, 56, 57, 65, 67
<i>MultinomialNB</i>	0.64	11, 24, 32, 33, 38, 41, 55	11, 12, 24, 25, 30, 310, 32, 33, 34, 35, 36, 37, 38, 39, 42, 44, 52, 53, 54, 55, 56, 57, 65, 67
<i>ComplementNB</i>	1.0	11, 24, 33, 37, 38, 39, 41, 55, 56	12, 25, 30, 34, 42, 44, 52, 53, 54, 65, 67

Pela análise da tabela anterior percebe-se que uma grande maioria dos classificadores granjeou um bom valor de exatidão de treino, afora *KNeighbors*, *BernoulliNB* e *MultinomialNB*. No que diz respeito às precisões, averiguou-se algumas categorias com maior propensão para falhar na previsão, entre elas, as Cat12, Cat44, Cat53, Cat65 e Cat67, pois sempre apresentaram valores nulos. Das categorias com melhores previsões, a Cat55 foi a única que se conseguiu prever em todas as situações. É de referir que, apesar das categorias 62 e 63 possuírem a respetiva resposta tipo, não surgiu qualquer exemplo no *dataset* empregado.

Quanto aos critérios exatidão e macro F1 demonstrados para cada um dos classificadores, os mesmos estão dispostos na tabela posterior.

Tabela 5. 14 - Vista geral do desempenho dos vários classificadores atendendo às inúmeras categorias

Classificador	Exatidão do teste	Macro F1
<i>Linear SVC</i>	0.53	0.46
<i>Bagging</i>	0.39	0.30
<i>RandomForest</i>	0.32	0.28
<i>Multi-layer Perceptron</i>	0.40	0.37
<i>Stochastic Gradient Descent</i>	0.31	0.24
<i>DecisionTree</i>	0.33	0.26
<i>ExtraTree</i>	0.23	0.18
<i>KNeighbors</i>	0.36	0.31
<i>BernoulliNB</i>	0.08	0.00
<i>GaussianNB</i>	0.27	0.23
<i>MultinomialNB</i>	0.21	0.09
<i>ComplementNB</i>	0.41	0.36

Como é possível comprovar pela tabela prévia, o melhor valor para a exatidão do teste reafirmou-se novamente como sendo o *Linear SVC* apenas com 53% e todos os outros indicaram valores extremamente baixos, pelo que esta metodologia acaba por perder alguma fiabilidade. Comparado com os resultados adquiridos no ponto 5.4.1, estes mostraram-se relativamente inferiores.

5.4.2.4. Resposta automática

Tal como anteriormente, esta fase suportou ligeiras mudanças a começar pelo “host”:

```
host = 'smtp.office365.com'
```

Dado que no processo de resposta automática do tipo genérica utilizou-se o servidor da FEUP, optou-se por testar o da Águas do Porto. Para tal, introduziu-se o endereço respetivo e o modo de *login* no servidor. Aproveitou-se também para modificar o assunto do email gerado:

```
subject = 'Teste de envio'
toaddrs = ['leandra.rocha@aguasdoporto.pt']
server.login('relaySmtpt@aguasdoporto.pt', password)
```

No fim do código, foi necessário permutar o ficheiro que contém os novos *templates*. Estes estão dispostos no Anexo 5:

```
templates = pd.read_excel('templates_das_categorias_2.xlsx')
```

Para executar a ação de resposta automática, colocaram-se os diretórios dos ficheiros que compõem o código por ordem de cumprimento na linha de comandos, como se pode observar pela Figura 5.16.

```

Anaconda Prompt
Cat35    0.60    0.75    0.67    4
Cat36    0.50    0.25    0.33    4
Cat37    1.00    0.50    0.67    2
Cat38    1.00    0.50    0.67    2
Cat39    1.00    0.33    0.50    3
Cat41    1.00    0.50    0.67    2
Cat42    0.00    0.00    0.00    2
Cat44    0.00    0.00    0.00    1
Cat51    0.29    1.00    0.45    7
Cat52    0.00    0.00    0.00    1
Cat53    0.00    0.00    0.00    1
Cat54    0.00    0.00    0.00    3
Cat55    1.00    1.00    1.00    1
Cat56    1.00    0.20    0.33    5
Cat57    0.50    0.50    0.50    2
Cat65    0.00    0.00    0.00    1
Cat67    0.00    0.00    0.00    1
Cat70    0.33    0.50    0.40    4

accuracy    0.41    90
macro avg   0.44    0.36    0.36    90
weighted avg 0.45    0.41    0.38    90

(leandra thesis) C:\Users\leandrarocha.AGUASDOPORTO\Desktop\Tese Leandra\Tese Python>python _4_Resposta_Automatica.py teste_emails_2.xlsx "Conteúdo" "noreply@aguasdoporto.pt"
Por favor, coloque a palavra-chave.
Password:
(leandra thesis) C:\Users\leandrarocha.AGUASDOPORTO\Desktop\Tese Leandra\Tese Python>

```

Fig. 5. 16 - Produção da resposta automática através da linha de comando Anaconda

Assim que foi validada a senha correspondente ao endereço definido, gerou-se duas respostas automáticas nesse preciso momento para a caixa de correio da mesma conta com base no *dataset* “teste_emails_2.xlsx” que contém três exemplares. Para todos eles, a atribuição foi bem-sucedida, tal como se pode visualizar na figura que se segue.



Fig. 5. 17 – Exemplo de duas respostas acionadas

5.5. GERAÇÃO DE TEXTO AUTOMÁTICO IMPLEMENTANDO UMA ARQUITETURA DE REDE NEURONAL RECORRENTE ARTIFICIAL (LSTM)

Segundo a pesquisa bibliográfica, LSTM é uma rede neuronal artificial aplicada no campo da aprendizagem profunda (*Deep Learning*) projetada para reconhecer padrões em sequências de dados, como texto, genomas, caligrafia, palavra falada ou dados numéricos de séries temporais emanados de sensores, mercados de ações e agências governamentais. Esses algoritmos têm em consideração o tempo e a sequência, e possuem uma dimensão temporal.

As redes neurais recorrentes podem ser usadas como modelos generativos, o que significa que, além de serem usados em modelos preditivos, eles são capazes de aprender as sequelas de um problema e, em seguida, gerar sequências plausíveis totalmente novas para o domínio do problema. Modelos generativos são úteis não apenas para estudar quão bem um modelo aprendeu um problema, mas para aprender mais sobre o próprio domínio do problema. De seguida, é demonstrado o procedimento para a criação um modelo generativo para texto, caracter por caracter, usando redes neurais recorrentes LSTM com a biblioteca *Keras* de código aberto escrita em *Python*, que se destina a desenvolver e avaliar modelos de aprendizagem profunda.

5.5.1. IMPORTAÇÃO DOS MÓDULOS E CONVERSÃO DE TEXTO

Começou-se por importar todas as bibliotecas necessárias ao código no contexto de uma rede maior de LSTM e de geração de texto que se pretendeu usar para treinar o modelo:

```
import sys
import numpy
from keras.models import Sequential
from keras.layers import Dense
from keras.layers import Dropout
from keras.layers import LSTM
from keras.callbacks import ModelCheckpoint
from keras.utils import np_utils
from pandas import read_excel
from sys import argv
```

No passo seguinte foram verificados os argumentos indispensáveis à previsão e carregou-se o *dataset* e as respostas tipo utilizadas no ponto 5.4.1, uma vez que se obteve melhores resultados. Depois o texto presente em ambos os ficheiros foi convertido em letras minúsculas, em que se esperou o carácter ‘\0’ somente ao separar emails de *templates*:

```
df1 = read_excel('teste_emails_1.xlsx')
emails = [email for email in df1['Conteúdo']]
categories = [c for c in df1['Categoria']]
df2 = read_excel('templates_das_categorias.xlsx')
templates = dict((df2['Categoria'][i], df2['Template'][i]) for i in range(len(df2['Categoria'])))
raw_text = []
for email, category in zip(emails, categories):
    template = templates[category]
    raw_text.append(email)
    raw_text.append(template)
raw_text = '\0'.join(raw_text)
raw_text = raw_text.lower()
```

5.5.2. MODELAÇÃO DA REDE NEURONAL

Tendo o texto carregado, preparou-se os dados para modelação pela rede neural. Não é possível modelar os caracteres diretamente, portanto em vez disso deve-se converter os caracteres em números inteiros. Gerou-se primeiro um conjunto de todos os caracteres diferenciados do texto e, em seguida, o mapeamento de cada caracter para um número inteiro único e o inverso deste. O texto foi de novo carregado e o mapeamento preparado, pelo que se pôde resumir o conjunto de dados:

```
chars = sorted(list(set(raw_text)))
char_to_int = dict((c, i) for i, c in enumerate(chars))
int_to_char = dict((i, c) for i, c in enumerate(chars))

n_chars = len(raw_text)
n_vocab = len(chars)
print("Total Characters: ", n_chars)
print("Total Vocab: ", n_vocab)
```

A execução do código neste ponto produziu a seguinte saída:

```
Total Characters: 577266
Total Vocab: 105
```

Constatou-se que o texto tem mais de 500 000 caracteres (número de elementos do texto bruto) e que, quando convertido para letras minúsculas, existiam apenas 105 caracteres distintos no vocabulário (número de caracteres únicos) para a rede aprender, muito mais do que os 26 do alfabeto. Precisou-se depois de definir os dados de treino para a rede.

O texto foi dividido em subsequências com um comprimento fixo de 100 caracteres (tamanho arbitrário). Os dados podem ser facilmente repartidos por frases, preenchidos nas sequências mais curtas e truncados nas mais longas. Cada padrão de treino da rede é composto por 100 etapas de tempo de um caracter (X) seguido por uma saída de caracter (y). Ao criar essas sequências, percorreu-se o painel de visualização ao longo do texto, um caracter de cada vez, permitindo que cada caracter aprenda com os 100 caracteres que o precederam exceto os 100 primeiros.

```
seq_length = 100
dataX = []
dataY = []
for i in range(0, n_chars - seq_length, 1):
    seq_in = raw_text[i:i + seq_length]
    seq_out = raw_text[i + seq_length]
    dataX.append([char_to_int[char] for char in seq_in])
    dataY.append(char_to_int[seq_out])
n_patterns = len(dataX)
print("Total Patterns: ", n_patterns)
```

A execução do código até este ponto expôs que, quando se dividiu o conjunto de dados em dados de treino para a rede neuronal, deparou-se com mais de 500 000 padrões de treino. Excetuando os 100 primeiros caracteres, existiu um padrão de treino para prever cada um dos caracteres restantes:

```
Total Patterns: 577166
```

Dispondo dos dados de treino, foi preciso transformá-los para que fossem compatíveis com *Keras*. Primeiro, modificou-se a lista de sequências de entrada no formato [amostras, etapas de tempo, características] esperado por uma rede LSTM. Em seguida, redimensionou-se os números inteiros no intervalo de 0 a 1 para facilitar a aprendizagem dos padrões pela rede LSTM que usa a função de ativação sigmoide por padrão. Finalmente, converteu-se os padrões de saída apresentados como caracteres únicos convertidos em números inteiros numa única codificação quente (*hot encoding*), ocorrendo para que fosse possível configurar a rede para prever a probabilidade de cada um dos 105 caracteres diferentes no vocabulário, em vez de tentar forçá-la a prever com precisão o próximo caracter.

Todas essas etapas foram implementadas como se averigua abaixo.

```
X = numpy.reshape(dataX, (n_patterns, seq_length, 1))
X = X / float(n_vocab)
y = np_utils.to_categorical(dataY)
```

Definiu-se um modelo LSTM com duas camadas, em que cada camada consta 256 neurónios. A camada de saída utilizou a função de ativação *softmax* para gerar uma previsão de probabilidade para cada um dos 105 caracteres entre 0 e 1.

Tratou-se de problema de classificação de caracter único com 105 classes e, como tal, foi definido como otimizar a perda (entropia cruzada), justapondo aqui o algoritmo de otimização *adam* para a velocidade:

```
model = Sequential()
model.add(LSTM(256, input_shape=(X.shape[1], X.shape[2]), return_sequences=True))
model.add(Dropout(0.2))
model.add(LSTM(256))
model.add(Dropout(0.2))
model.add(Dense(y.shape[1], activation='softmax'))

if mode == '--train':
    model.compile(loss='categorical_crossentropy', optimizer='adam')
```

Não houve uma base de dados de teste, pois modelou-se todo o conjunto de dados de treino para aprender a probabilidade de cada caracter numa sequência. Uma vez que o objetivo não passou por obter o modelo mais preciso através da precisão de classificação do conjunto de dados de treino, o que se pretendeu mesmo foi a generalização do conjunto de dados que minimizasse a função de perda escolhida. Procurou-se um equilíbrio entre generalização e sobre ajuste, mas com pouca memorização e um treino da rede vagaroso. Devido à morosidade e aos requisitos de otimização, pretendeu-se utilizar o ponto de verificação do modelo para registar todos os pesos de rede a serem arquivados cada vez que uma melhoria na perda for observada no final da época e o melhor conjunto de pesos (menor perda) para fundamentar o modelo generativo na próxima secção:

```
filepath="weights-improvement-{epoch:02d}-{loss:.4f}-bigger.hdf5"
checkpoint = ModelCheckpoint(filepath, monitor='loss', verbose=1, save_best_only=True,
mode='min')
callbacks_list = [checkpoint]
```

Ajustou-se então o modelo aos dados, utilizando um número finito de 50 épocas e um tamanho *batch* de 64 padrões (valor arbitrário), fixados com vista a dar uma oportunidade à rede para ser atualizada e aprender:

```
model.fit(X, y, epochs=50, batch_size=64, callbacks=callbacks_list)
```

Os resultados são diferentes devido à natureza estocástica do modelo e porque é difícil fixar a semente aleatória para os modelos LSTM para obter resultados 100% reproduzíveis. Depois de se executar o exemplo, é aconselhável ter vários arquivos de ponto de verificação de peso no diretório local, excluindo todos eles exceto aquele com o menor valor de perda.

5.5.3. USO DO MODELO DE TREINO PARA GERAÇÃO DE TEXTO

Na próxima fase verificou-se como usar o modelo de treino para gerar novas sequências de texto.

Imediatamente abaixo encontra-se o ponto de verificação com a menor perda que se atingiu, sendo que a partir das 37 épocas em diante se deteriorou a função de perda.

```
weights-improvement-37-0.8640-bigger.hdf5
“(…)
Epoch 00036: loss did not improve from 0.91622
Epoch 37/50
577166/577166 [=====] - 1388s 2ms/step - loss: 0.8640

Epoch 00037: loss improved from 0.91622 to 0.86401, saving model to weights-improvement-37-
0.8640-bigger.hdf5 (…)”
```

Em primeiro lugar, carregou-se os dados e estabeleceu-se a rede exatamente da mesma maneira, exceto os pesos da rede neuronal que foram carregados a partir de um arquivo de ponto de verificação e a rede não necessitou de ser treinada.

```
filename = argv[2]
model.load_weights(filename)
model.compile(loss='categorical_crossentropy', optimizer='adam')
```

5.5.4. PREVISÃO DA SAÍDA DE TEXTO

Passou-se assim à realização de previsões. A metodologia do modelo *Keras* comporta as tentativas para gerar 1000 caracteres, onde foram fornecidos 100 caracteres do email posteriormente normalizado e gera a sequência seguinte ao longo de um ciclo que ocorreu 1000 vezes. Foi exequível a escolha de um padrão de entrada aleatório como sequência inicial e, logo depois, publicou-se os caracteres produzidos à medida que iam sendo concebidos, graças ao uso do ficheiro “.flush()” que permitiu mostrar caracter a caracter na consola em vez de aguardar a ação do código até ao fim:

```
start = numpy.random.randint(0, len(dataX)-1)
pattern = dataX[start]
print("Seed:")
print("\n", ".join([int_to_char[value] for value in pattern]), "\n")

#null_count = 0
#while True:
for i in range(1000):
    x = numpy.reshape(pattern, (1, len(pattern), 1))
    x = x / float(n_vocab)
    prediction = model.predict(x, verbose=0)
    index = numpy.argmax(prediction)
    result = int_to_char[index]
    seq_in = [int_to_char[value] for value in pattern]
    sys.stdout.write(result)
    sys.stdout.flush()
    pattern.append(index)
    pattern = pattern[1:len(pattern)]
    #if result == '\0':
    #    null_count = null_count + 1
    #    if null_count >= 2:
    #        break
print("\nDone.")
else:
    raise ValueError("Only the options '--train' and '--predict' are accepted.")
```

O primeiro *output* correspondeu à semente aleatória selecionada, ou seja, partiu de 100 caracteres aleatórios do *dataset*, e depois os caracteres seguintes que foram sendo gerados. Apresentou-se assim os resultados de uma execução deste gerador de texto:

```

Selecionar Anaconda Prompt
(base) C:\Users\Asus>cd "C:\Users\Asus\Desktop\Tese Leandra\Tese Python\text_generator_LSTM"

(base) C:\Users\Asus\Desktop\Tese Leandra\Tese Python\text_generator_LSTM>conda activate text_generator

(text_generator) C:\Users\Asus\Desktop\Tese Leandra\Tese Python\text_generator_LSTM>python ./text_generator_LSTM.py --
predict ./weights-improvement-37-0.8640-bigger.hdf5
Using TensorFlow backend.
Total Characters: 577266
Total Vocab: 105
Total Patterns: 577166
2019-10-10 17:29:55.166861: I tensorflow/core/platform/cpu_feature_guard.cc:142] Your CPU supports instructions that t
his TensorFlow binary was not compiled to use: AVX AVX2
2019-10-10 17:29:55.174926: I tensorflow/stream_executor/platform/default/dso_loader.cc:42] Successfully opened dynami
c library nvcuda.dll
2019-10-10 17:29:55.668093: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1640] Found device 0 with properties:
name: GeForce 930MX major: 5 minor: 0 memoryClockRate(GHz): 0.9015
pciBusID: 0000:01:00:00
2019-10-10 17:29:55.674741: I tensorflow/stream_executor/platform/default/dlopen_checker_stub.cc:25] GPU libraries are
statically linked, skip dlopen check.
2019-10-10 17:29:55.679830: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1763] Adding visible gpu devices: 0
2019-10-10 17:29:56.822276: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1181] Device interconnect StreamExecuto
r with strength 1 edge matrix:
2019-10-10 17:29:56.827478: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1187]      0
2019-10-10 17:29:56.829518: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1200] 0:  N
2019-10-10 17:29:56.832268: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1326] Created TensorFlow device (/job:l
ocalhost/replica:0/task:0/device:GPU:0 with 1391 MB memory) -> physical GPU (device: 0, name: GeForce 930MX, pci bus i
d: 0000:01:00:00, compute capability: 5.0)
Seed:
" sada sexta feira vieram à obra fazer limpeza da fossa, no entanto só tínhamos pedido para o dia 11 d "
WARNING:tensorflow:From C:\Users\Asus\Anaconda3\envs\text_generator\lib\site-packages\keras\backend\tensorflow_backend
.py:422: The name tf.global_variables is deprecated. Please use tf.compat.v1.global_variables instead.

e contador de contador de contador de fontador de fontador de contador de contador de contador de contador de contador
de contador de contador de aontador de contador de contador de contador de contador de contador de contador de aontad
or de contador de contador de contador de contador de contador de contador de aontador de contador de contador de co
ntador de contador de contador de aontador de contador de contador de contador de contador de contador de aontador de
contador de contador de contador de aontador de contador de contador de aontador de contador de contador de aontador
de contador de contador de aontador de contador de contador de contador de contador de contador de contador de aontador
de contador de contador de aontador de contador de contador de aontador de contador de aontador de aontador de aontador
de contador de contador de aontador de aontador de co
Done.

```

Fig. 5. 18 - Demonstração do processo de geração de texto de uma porção de email aleatório

Conforme se visualiza pela figura prévia, o comprimento fixado para a sequência foi de 100, o que originou a produção de parte de um email com base num fio de texto aleatório e deixou de poder prever desde que reconheceu o termo “fatura”, gerando sucessivamente a palavra “contador” e outras aproximadas a esta sem qualquer nexos. Apresentou nitidamente mais dificuldade ao produzir este tipo de texto após o carregamento da rede neuronal treinada, visto que existem inúmeras sequências ao qual o código se foi adaptando ao longo do treino.

Segue outra amostra do que foi apresentado na linha de comandos na Figura 5.19.


```

Anaconda Prompt
2019-10-10 17:08:06.776712: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1187] 0
2019-10-10 17:08:06.779442: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1200] 0: N
2019-10-10 17:08:06.782437: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1326] Created Ten
sorFlow device (/job:localhost/replica:0/task:0/device:GPU:0 with 1391 MB memory) -> physical GP
U (device: 0, name: GeForce 930MX, pci bus id: 0000:01:00:0, compute capability: 5.0)
Seed:
" lhores cumprimentos,

serviço ao cliente águas do porto

contactos:
email geral/avarias: g "
WARNING:tensorflow:From C:\Users\Asus\Anaconda3\envs\text_generator\lib\site-packages\keras\back
end\tensorflow_backend.py:422: The name tf.global_variables is deprecated. Please use tf.compat.
v1.global_variables instead.

eral@aguasdoporto.pt
email área comercial: cliente@aguasdoporto.pt
linha verde comercial: 800 206 235 (2.ª a 6.ª feira, das 9h00m às 17h). boa tarde,

venho por este meio solicitar a alteração de contador de fontador de fontador de fontador de con
tador de contador de contador de contador de contador de contador de aontador de con
tador de contador de contador de contador de contador de contador de aontador de contador de con
tador de contador de contador de contador de contador de aontador de contador de contador de con
tador de contador de contador de contador de aontador de contador de contador de contador de con
tador de contador de aontador de contador de contador de contador de contador de contador de con
tador de aontador de contador de contador de contador de contador de contador de contador de aon
tador de contador de contador de contador de contador de contador de contador de aon
Done.

```

Fig. 5. 19 - Exemplificação da criação de texto envolvendo uma resposta tipo genérica e parte de um email aleatório

Aqui ainda se conseguiu prever razoavelmente bem a resposta tipo designada, isto porque se deveu à normalização do texto dados os poucos *templates* existentes e análogos entre si. No final, produziu parte de um email aleatório que, alcançando os 100 caracteres, não foi capaz de conceber além desse ponto e assumiu o restante texto em torno do termo “contador”, tal como verificado no exemplo anterior.

Apesar dos resultados não terem sido favoráveis, este código facilmente pode ser melhorado em vários aspetos, entre os quais, o ajuste do número máximo de caracteres previstos para 1 000; o aumento do número de épocas de treino para 100 ou muitas centenas; testar à volta do tamanho *batch* o acréscimo das unidades de memória às camadas (mais LSTM) e/ou juntar mais camadas, isto é, adequar a uma estrutura de rede sofisticada, e ainda o treino do modelo usando frases preenchidas em alternativa às sequências aleatórias de caracteres. De aludir que este é um processo demorado que requer muitas horas ou até mesmo dias aquando da modificação de parâmetros, no entanto para que se possa otimizar e obter resultados aceitáveis e realistas, exige um estudo aprofundado com base em diversas experiências.

As RNNs processam dados sequenciais, uma vez que se podem atender às últimas entradas por via de uma memória interna, gerando texto ao estilo de um autor específico. Alcançam um elevado desempenho em praticamente todos os problemas sequenciais e são usados pela maioria das grandes empresas.

Existe um modelo de RNNs, com fator multiplicativo, que aprende surpreendentemente bem modelos de linguagem usando apenas 1 500 unidades ocultas que, ao contrário de outras abordagens, é facilmente flexível ao longo de várias dimensões. Após um treino árduo com este tipo de rede maior, com milhões de unidades e bilhões de conexões, é possível que seja suficiente para atingir um padrão de desempenho ainda mais alto, no entanto obrigará consideravelmente a mais poder computacional.

O código completo elaborado em LSTM encontra-se no Anexo 6.

5.6. ANÁLISE CONTABILÍSTICA DE SOLICITAÇÕES

A contabilidade analítica apresenta um papel fundamental na criação de valor para a empresa no âmbito de uma gestão que tenciona lograr uma maior eficiência principalmente na produtividade dos seus recursos, enquanto instrumento de gestão e de avaliação do desempenho económico nas áreas de negócio e das atividades nas vertentes organizacionais.

Assim, recorreu-se ao Relatório de Contabilidade Analítica – Resultados 2018 para análise de solicitações recebidas no ano transato.

5.6.1. UO GESTÃO DE CLIENTES

Esta UO atua num conjunto de atividades com 11 indicadores de gestão, sendo as solicitações de clientes as mais importantes para o estudo. Foram registados 13 584 pedidos, cujos gastos diretos (Pessoal, Fornecimento e Serviços Externos, Depreciações e Outros gastos) totalizaram um custo para empresa de 119 414 €. Atendendo ao gasto médio por solicitações, o valor total de gastos direto seria menor com a implementação de um código robusto capaz de agilizar um maior número de processos que entram, respondendo de forma clara e concisa. Isto refletir-se-ia claramente nos gastos com colaboradores, pois a ação de resposta automática dispensaria grande parte das vezes a intervenção humana.

5.6.2. UO SALA DE COMANDO

A UO Sala de Comando é responsável pela coordenação e gestão de recursos, onde a atividade de comunicação eletrónica perfez um gasto direto de 15 988 €, traduzindo-se em gastos com Pessoal, FSE, Depreciações e Outros tipos de gastos, dada a receção de 50 000 emails. A partir do gasto direto por unidade de volume, prevê-se um decréscimo do valor total com a utilização de um código preparado para responder a vários tipos de solicitações e, consequentemente, da constante necessidade de um funcionário afeto a este tipo de serviço.

6

CONCLUSÕES E RECOMENDAÇÕES

6.1. CONCLUSÕES

A gestão de solicitações é uma realidade pouco ou nada explorada nas EG que obriga o cliente a entrar a contacto com a prestadora de serviços em diversos momentos, sendo que por vezes não vê a concretização do seu pedido ou este não é solucionado.

A presente dissertação centrou-se principalmente no desenvolvimento de uma proposta para classificação de solicitações através da utilização da linguagem de programação *Python* na construção de um algoritmo capaz de responder de forma genérica a todo o tipo de situações que chegam por correio eletrónico.

Foram 3 os estágios do projeto: no primeiro, apresentou-se um método de classificação que se divide em várias categorias, consoante as temáticas dos pedidos e baseado no Manual de Procedimentos para a Gestão de Reclamações existente na empresa; depois, analisou-se 6038 mensagens rececionadas em duas contas de email, passando pela categorização e elaboração de mapas que contêm a georreferenciação de todos os pedidos anotados com sucesso; no terceiro e último passo, projetou-se um algoritmo que classifica e gera uma resposta automática.

Aquando da análise e recolha de exemplos de emails, deparou-se com alguma dificuldade na consulta de contratos de arrendamento ou de cadernetas prediais urbanas, pois os CIL não corresponderam ao descrito devido à localização da entrada do prédio/casa nem sempre coincidir com o registo no sistema UBS. Contudo, acabaram por não ser considerados ou após uma extensa pesquisa realmente foi possível obter os dados, tendo pouca incidência.

Após a classificação das solicitações, fez-se a sua representação geográfica que levantou uma série de questões pelo facto de haver instalações sem georreferenciação (não disponibilizados no SIG) e a arduidade na marcação dos pontos que se encontraram repetidos no ficheiro de apoio, respeitante a pedidos efetuados dentro da mesma categoria, pela cidade do Porto. Foi necessário adaptar as categorias geradas ao ambiente da programação, mas indispensável atendendo à otimização dos resultados obtidos pelo algoritmo. Quanto menos as classes, mais fiável é a classificação, agrupando-se assim em sete categorias (Cat6 junta todas as tipologias pequenas e Cat7 destinada às reclamações para a separar das restantes).

Apesar da dificuldade demonstrada na execução do código dada a sua complexidade e campo de aplicação distinto que trabalhou um leque variado de 449 e-mails do mesmo período de estudo, obtiveram-se resultados favoráveis relativamente às precisões de teste para o classificador *Linear SVC* entre os vários classificadores ensaiados, cuja exatidão de teste foi de 74%. No que concerne aos *templates*, estes foram formulados com o objetivo de serem apenas 1 por categoria, caso contrário

afetaria a *performance* do algoritmo. A partir da matriz de confusão, verificou-se a distribuição bem-sucedida ao longo da diagonal onde se encontram os verdadeiros positivos (cruzamento dos exemplos reais com os exemplos previstos).

De seguida, reformulou-se um outro cenário com o mesmo código, porém com diversas subcategorias e, consequentemente, mais respostas tipo. Todavia, os resultados logrados para todos os classificadores foram inferiores comparativamente com o caso anterior, o que não se refletiu muito na execução de resposta automática que conseguiu atribuir a resposta correta várias vezes ao longo da experimentação. Claro está que se sujeitou sempre a alguma margem de erro.

Na última fase do problema, criou-se um código para geração de texto automático utilizando redes neurais (LSTM) com duas camadas, cada uma com 256 neurónios. Configurou-se um comprimento máximo de 100 caracteres que testou durante 50 épocas, onde a menor função de perda foi detetada à época 37 tendo por base o *dataset* e os *templates* iniciais. Aquando da execução deste código na linha de comandos, observou-se a reprodução de 100 caracteres de parte de um email, no entanto o restante texto foi gerado sem que fizesse qualquer sentido. Portanto, propôs-se essencialmente a transposição para 1.000 caracteres dada a impossibilidade da sua concretização.

Em suma, este projeto desenvolvido em torno de ambos os códigos, constitui uma mais valia a pôr em funcionamento pelas EG, que pode ser ajustado a qualquer cenário e integrado nas suas estratégias tendo em vista a melhoria do relacionamento com o cliente. Eleva o valor do serviço prestado e conduz a numa poupança significativa nos gastos diretos com as comunicações por meio de correio eletrónico, tal como promove a diminuição do tempo médio de resposta.

6.2. RECOMENDAÇÕES PARA TRABALHOS FUTUROS

Face à impossibilidade para desenvolver outros projetos dado o tempo despendido na análise de solicitações de clientes via e-mail e na conceção dos códigos idealizados para gerar resposta às mesmas, expõem-se as seguintes propostas de melhoramento e de complemento ao tema em causa:

- Arranjo e implementação de um algoritmo mais robusto preparado para responder de forma personalizada a qualquer tipo de conteúdo, quer seja reclamação ou solicitação e ainda adaptação ao modelo próprio do *Microsoft Outlook* numa perspetiva de maior proximidade da interface com esta ferramenta;
- Treino do código em LSTM para 1.000 caracteres com base em respostas reais e a realização de inúmeros testes, de modo a otimizar os resultados;
- Integração do pedido de referência gerado por resposta automática aos e-mails no Balcão Digital, para que o cliente possa acompanhar a resolução da sua interação com a empresa;
- Sugestão de novos serviços aos utilizadores finais, a partir da interseção dos consumos que permitam alcançar os maiores consumidores e potenciais clientes interessados na instalação de contadores ultrassónicos como novas fontes de receitas;

REFERÊNCIAS BIBLIOGRÁFICAS

- AEP - ÁguaGlobal. AEP, 2014. Disponível em WWW: <http://aguaglobal.aeportugal.pt/Documents/Encerramento/Setor%20Portugu%C3%AAs%20da%20%C3%81gua.pdf>. ISBN/ISSN:
- Águas do Porto - Organograma. Águas do Porto, EM: 2018a. ISBN/ISSN:
- Águas do Porto - Telegestão e Telemetria. Porto: Águas do Porto, 2018b. Disponível em WWW: <https://www.aguasdoporto.pt/distribuicao-de-agua/telegestao-e-telemetria>. ISBN/ISSN:
- Águas do Porto - Novo Balcão Digital da Águas do Porto é mais prático e intuitivo. Porto: Águas do Porto, EM, 2019a. ISBN/ISSN:
- Águas do Porto - Relatório & Contas 2018. Porto, Portugal: 2019b. ISBN/ISSN:
- Águas do Porto - Distribuição de Água. Porto: 2014. Disponível em WWW: <https://www.aguasdoporto.pt/areas-de-intervencao/distribuicao-de-agua>. ISBN/ISSN:
- Águas do Porto - Relatório & Contas 2017. Águas do Porto, EM. Porto: Águas do Porto, EM, 2018. ISBN/ISSN:
- Alecrim, Emerson - Machine learning: o que é e por que é tão importante 2018. Disponível em WWW: <https://tecnoBlog.net/247820/machine-learning-ia-o-que-e/>. ISSN:
- APA - Plano Estratégico de Abastecimento de Água e Saneamento de Águas Residuais 2020. s.d. Disponível em WWW: <https://apambiente.pt/index.php?ref=16&subref=7&sub2ref=9&sub3ref=1098>. ISBN/ISSN:
- APDA - Gestão de Reclamações como Oportunidade de Melhoria. Lisboa: 2015. ISBN/ISSN:
- Bhatia, Rahul - When not to use Neural Networks. Medium, 2018. Consult. em 23 de setembro de 2019. Disponível em WWW: <https://medium.com/datadriveninvestor/when-not-to-use-neural-networks-89fb50622429>. ISBN/ISSN:
- Brownlee, Jason - Supervised and Unsupervised Machine Learning Algorithms 2016. Disponível em WWW: <https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/>. ISSN:
- CNCS - A Internet das Coisas (IoT – Internet of Things). s.d. Disponível em WWW: <https://www.cncs.gov.pt/a-internet-das-coisas-iot-internet-of-things/>. ISBN/ISSN:
- Codecademy - Why Learn Python? Codecademy, 2019. Disponível em WWW: <https://www.codecademy.com/learn/learn-python-3>. ISBN/ISSN:
- Creekside Center for Earth Observation - GIS Layers Edit. 2012. Consult. em 20 de agosto de 2019. Disponível em WWW: https://creeksidescience.com/people/gis_layers_edit/. ISBN/ISSN:
- Cunha, Sara - O SIG ao Serviço do Ordenamento do Território: Modelo de Implementação. Faculdade de Letras da Universidade do Porto: Universidade do Porto, 2009. ISBN/ISSN:
- DataRobot - Unsupervised Machine Learning. 2019. Disponível em WWW: <https://www.datarobot.com/wiki/unsupervised-machine-learning/>. ISBN/ISSN:
- ERSAR - Relação das entidades gestoras com os utilizadores dos serviços de águas e resíduos Entidade Reguladora dos Serviços de Águas e Resíduos Lisboa: 2012. ISBN: 978-989-8360-14-4
- ERSAR - Estudo de satisfação dos utilizadores dos serviços de águas e resíduos – Estudo piloto. Entidade Reguladora dos Serviços de Águas e Resíduos 2016. ISBN/ISSN:
- ERSAR - Estudo aos utilizadores dos serviços de abastecimento de água, saneamento de águas residuais urbanas e gestão de resíduos urbanos, sobre a satisfação dos serviços prestados pelas várias entidades destes setores: Relatório Global Nacional. Entidade Reguladora dos Serviços de Águas e Resíduos 2017. ISBN/ISSN:
- ERSAR - Projeto de Regulamento Tarifário dos Serviços de Águas 2018a. Consult. em 30 de maio de 2019. ISBN/ISSN:
- ERSAR - Relatório Anual dos Serviços de Águas e Resíduos em Portugal Entidade Reguladora dos Serviços de Águas e Resíduos 2018b. ISBN/ISSN:
- ERSAR - Abastecimento de Água. 2019a. Consult. em 27 de fevereiro de 2019. Disponível em WWW: <http://www.ersar.pt/pt/setor/caracterizacao/abastecimento-de-agua>. ISBN/ISSN:
- ERSAR - Gestão de Reclamações. ERSAR, 2019b. Disponível em WWW: <http://www.ersar.pt/pt/o-que-fazemos/gestao-de-reclamacoes>. ISBN/ISSN:
- ERSAR - O setor da água e do saneamento de águas residuais. s.d. ISBN/ISSN:

- ESRI - What is ArcMap? Environmental Systems Research Institute. 2016a. Disponível em WWW: <<http://desktop.arcgis.com/en/arcmap/10.3/main/map/what-is-arcmap-.htm>>. ISBN/ISSN:
- ESRI - What is raster data? 2016b. Disponível em WWW: <<http://desktop.arcgis.com/en/arcmap/10.3/manage-data/raster-and-images/what-is-raster-data.htm>>. ISBN/ISSN:
- ESRI - O que são os SIG? s.d. Disponível em WWW: <<http://www.esriportugal.pt/O-que-sao-os-SIG->>. ISBN/ISSN:
- Figueiredo, Marisa - E se o seu médico de família fosse um robô? Lisboa: Montepio Geral - Associação Mutualista, 2018. ISBN/ISSN:
- GeeksforGeeks - Python for Data Science. s.d. Disponível em WWW: <<https://www.geeksforgeeks.org/python-for-data-science/>>. ISBN/ISSN:
- Jornal de Notícias - Câmara do Porto contra proposta de regulamento tarifário dos serviços de águas. *Jornal de Notícias*. (2019). Consult. em 14 de março. Disponível em WWW: <<https://www.jn.pt/local/noticias/porto/porto/interior/camara-do-porto-contraproposta-de-regulamento-tarifario-dos-servicos-de-aguas-10674266.html>>. ISSN:
- Olaia, Ana - Gestão de Sistemas de Abastecimento de Água através de Modelação Hidráulica. Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa: Universidade Nova de Lisboa, 2012. ISBN/ISSN:
- PENSAAR 2020 - Uma Estratégia ao Serviço da População: Serviços de Qualidade a um Preço Sustentável. 2015. ISBN/ISSN:
- Python Software Foundation - What is Python? Executive Summary. 2001. Disponível em WWW: <<https://www.python.org/doc/essays/blurbl/>>. ISBN/ISSN:
- Salesforce - Como a IA ajuda na experiência dos clientes? s.d.-a. Disponível em WWW: <<https://www.salesforce.com/br/products/einstein/guia-da-inteligencia-artificial/#>>. ISBN/ISSN:
- Salesforce - Inteligência Artificial x Experiência do Cliente. s.d.-b. Disponível em WWW: <<https://www.salesforce.com/br/products/einstein/guia-da-inteligencia-artificial/#>>. ISBN/ISSN:
- SAS - Deep Learning. 2018. Disponível em WWW: <https://www.sas.com/pt_br/insights/analytics/deep-learning.html>. ISBN/ISSN:
- Sinfic - Fotografia Aérea e Ortofotomapas. s.d. Disponível em WWW: <<http://www.sinfic.pt/sig/displayconteudo.do2?numero=34539>>. ISBN/ISSN:
- Singh, Maneesh - Python Features 2016. Disponível em WWW: <<https://www.geeksforgeeks.org/python-features/>>. ISSN:
- Sousa, Susana; Vilhena, Ana; Frada, Joana; Colarejo, José; Fernandes, José; Peças, José; Moreira, Lino; Branco, Luís; Marques, Marta; Martins, Raquel; Martins, Susana - A Experiência do Cliente 2017. ISSN:
- Teixeira, Joana - Redução das Perdas Aparentes em Sistemas de Abastecimento de Água através da Detecção e Controlo de Consumos Ilícitos Faculdade de Engenharia da Universidade do Porto: Universidade do Porto, 2014. ISBN/ISSN:
- Water & Wastewater International - Drinking water and wastewater coverage and treatment. 2011. Consult. em 27 de fevereiro de 2019. ISBN/ISSN:

ANEXOS

1. LISTA DAS STOPWORDS UTILIZADAS

['de', 'a', 'o', 'que', 'e', 'é', 'do', 'da', 'em', 'um', 'para', 'com', 'não', 'uma', 'os', 'no', 'se', 'na', 'por', 'mais', 'as', 'dos', 'como', 'mas', 'ao', 'ele', 'das', 'à', 'seu', 'sua', 'ou', 'quando', 'muito', 'nos', 'já', 'eu', 'também', 'só', 'pelo', 'pela', 'até', 'isso', 'ela', 'entre', 'depois', 'sem', 'mesmo', 'aos', 'seus', 'quem', 'nas', 'me', 'esse', 'eles', 'você', 'essa', 'num', 'nem', 'suas', 'meu', 'às', 'minha', 'numa', 'pelos', 'elas', 'qual', 'nós', 'lhe', 'deles', 'essas', 'esses', 'pelas', 'este', 'dele', 'tu', 'te', 'vocês', 'vos', 'lhes', 'meus', 'minhas', 'teu', 'tua', 'teus', 'tuas', 'nosso', 'nossa', 'nossos', 'nossas', 'dela', 'delas', 'esta', 'estes', 'estas', 'aquele', 'aquela', 'aqueles', 'aquelas', 'isto', 'aquilo', 'estou', 'está', 'estamos', 'estão', 'estive', 'esteve', 'estivemos', 'estiveram', 'estava', 'estávamos', 'estavam', 'estivera', 'estivéramos', 'esteja', 'estejamos', 'estejam', 'estivesse', 'estivéssemos', 'estivessem', 'estiver', 'estivermos', 'estiverem', 'hei', 'há', 'havemos', 'hão', 'houve', 'houvermos', 'houveram', 'houvera', 'houvéramos', 'haja', 'hajamos', 'hajam', 'houvesse', 'houvéssemos', 'houvessem', 'houver', 'houvermos', 'houverem', 'houverei', 'houverá', 'houveremos', 'houverão', 'houveria', 'houveríamos', 'houveriam', 'sou', 'somos', 'são', 'era', 'éramos', 'eram', 'fui', 'foi', 'fomos', 'foram', 'fora', 'fôramos', 'seja', 'sejamos', 'sejam', 'fosse', 'fôssemos', 'fossem', 'for', 'formos', 'forem', 'serei', 'será', 'seremos', 'serão', 'seria', 'seríamos', 'seriam', 'tenho', 'tem', 'temos', 'tém', 'tinha', 'tínhamos', 'tinham', 'tive', 'teve', 'tivemos', 'tiveram', 'tivera', 'tivéramos', 'tenha', 'tenhamos', 'tenham', 'tivesse', 'tivéssemos', 'tivessem', 'tiver', 'tivermos', 'tiverem', 'terei', 'terá', 'teremos', 'terão', 'teria', 'teríamos', 'teriam']

2. RESPOSTAS TIPO GENÉRICAS PROJETADAS PARA O CÓDIGO

2.1. TEMPLATE PARA A CAT 1

Caro Cliente,

Acusamos a receção do seu email que mereceu a nossa melhor atenção, e em resposta, informamos que iremos atender ao seu pedido.

Para mais informações, consulte os links:

https://www.aguasdoporto.pt/paginas/contactos_7

<https://www.aguasdoporto.pt/cliente/regulamentos>

<https://www.aguasdoporto.pt/balcao-digital/c/o-que-e-o-balcao-digital>

Colocamo-nos ao seu dispor para qualquer esclarecimento adicional.

Com os melhores cumprimentos,

Serviço ao Cliente Águas do Porto



Para mais informações, por favor contacte-nos:

Email Geral/Avárias: geral@aguasdoporto.pt

Email Área Comercial: cliente@aguasdoporto.pt

Linha Verde Comercial: 800 206 235 (2.ª a 6.ª feira, das 9h00m às 17h).

2.2. TEMPLATE PARA A CAT2

Caro Cliente,

Na sequência da sua comunicação, informamos que o seu pedido foi rececionado nas nossas instalações e encontra-se em processamento.

Para mais informações, consulte as seguintes hiperligações:

https://www.aguasdoporto.pt/cliente/novo-contrato#form_personal_data

<https://www.aguasdoporto.pt/cliente/rescisao-contratual>

<https://www.aguasdoporto.pt/cliente/clausulas-do-contrato>

Colocamo-nos ao seu dispor para qualquer esclarecimento adicional.

Com os melhores cumprimentos,

Serviço ao Cliente Águas do Porto



Contactos:

Email Geral/Avarias: geral@aguasdoporto.pt

Email Área Comercial: cliente@aguasdoporto.pt

Linha Verde Comercial: 800 206 235 (2.ª a 6.ª feira, das 9h00m às 17h).

2.3. TEMPLATE PARA A CAT3

Caro Cliente,

Em resposta à sua solicitação, confirmamos a sua receção e contactá-lo-emos com a maior brevidade possível.

Para mais informações, consulte as seguintes páginas do nosso website:

<https://www.aguasdoporto.pt/cliente/fatura-electronica>

<https://www.aguasdoporto.pt/cliente/debito-direto-sepa>

<https://www.aguasdoporto.pt/cliente/locais-de-pagamento>

<https://www.aguasdoporto.pt/cliente/leitura-contador>

<https://www.aguasdoporto.pt/cliente/tarifario>

Colocamo-nos ao seu dispor para qualquer esclarecimento adicional.

Com os melhores cumprimentos,

Serviço ao cliente Águas do Porto



Contactos:

Email Geral/Avarias: geral@aguasdoporto.pt

Email Área Comercial: cliente@aguasdoporto.pt

Linha Verde Comercial: 800 206 235 (2.ª a 6.ª feira, das 9h00m às 17h).

2.4. TEMPLATE PARA A CAT4

Caro Cliente,

No seguimento do seu email, informamos que o seu pedido está a ser analisado pelos nossos colaboradores. Contactá-lo-emos com a máxima brevidade possível, de modo a agendar uma visita a efetuar pela equipa técnica ao local em questão.

Colocamo-nos ao seu dispor para qualquer esclarecimento adicional.

Com os melhores cumprimentos,

Serviço ao cliente Águas do Porto



Contactos:

Email Geral/Avarias: geral@aguasdoporto.pt

Email Área Comercial: cliente@aguasdoporto.pt

Linha Verde Comercial: 800 206 235 (2.ª a 6.ª feira, das 9h00m às 17h).

2.5. TEMPLATE PARA A CAT5

Caro Cliente,

Em resposta ao seu email, informamos que o seu pedido foi rececionado nos nossos serviços e prosseguirá para análise. Entraremos em contacto consigo brevemente.

Para mais informações, consulte os links:

<https://www.aguasdoporto.pt/licenciamento/licenciamento-digital>

<https://www.aguasdoporto.pt/licenciamento/c/redes-publicas>

Colocamo-nos ao seu dispor para qualquer esclarecimento adicional.

Com os melhores cumprimentos,

Serviço ao Cliente Águas do Porto



Para mais informações, por favor contacte-nos:

Email Geral/Avarias: geral@aguasdoporto.pt

Email Área Comercial: cliente@aguasdoporto.pt

Linha Verde Comercial: 800 206 235 (2.ª a 6.ª feira, das 9h00m às 17h).

2.6. TEMPLATE PARA A CAT6

Caro Cliente,

No seguimento do seu pedido, informamos que o mesmo deu entrada nas nossas instalações e encontra-se de momento em análise. Será contacto brevemente por um dos nossos colaboradores.

Para seguir as intervenções em curso, basta aceder a: <https://www.aguasdoporto.pt/paginas/quadro-de-ocorrencias>.

Colocamo-nos ao seu dispor para qualquer esclarecimento adicional.

Com os melhores cumprimentos,

Serviço ao cliente Águas do Porto



Contactos:

Email Geral/Avarias: geral@aguasdoporto.pt

Email Área Comercial: cliente@aguasdoporto.pt

Linha Verde Comercial: 800 206 235 (2.ª a 6.ª feira, das 9h00m às 17h).

2.7. TEMPLATE PARA A CAT7

Estimado(a) Cliente,

Na sequência da comunicação de V. Ex^a, cumpre-nos informar que a sua exposição foi rececionada e iremos contactá-lo em breve.

Para mais informações, por favor consulte o website: <https://www.livroreclamacoes.pt/inicio>

Colocamo-nos ao seu dispor para qualquer esclarecimento adicional.

Com os melhores cumprimentos,

Serviço ao cliente Águas do Porto



Contactos:

Email Geral/Avárias: geral@aguasdoporto.pt

Email Área Comercial: cliente@aguasdoporto.pt

Linha Verde Comercial: 800 206 235 (2.ª a 6.ª feira, das 9h00m às 17h).

3. CÓDIGO DE APRENDIZAGEM SUPERVISIONADA CONCEBIDO EM PYTHON

3.1. ANÁLISE DE DADOS

```
import pandas as pd
import matplotlib.pyplot as plt
import pickle
import seaborn as sns
sns.set_style("whitegrid")

import warnings
warnings.filterwarnings("ignore")

df_path = 'teste_emails_1.xlsx'
```

```
df = pd.read_excel(df_path)
```

```
plt.figure(figsize=(12.8,6))
sns.countplot(df['Categoria'].sort_values())
plt.ylabel('Número de emails')
plt.xlabel('Categoria')
plt.title('Número de emails em cada categoria')
plt.show()
```

```
df['Count'] = 1
df2 = pd.DataFrame(df.groupby('Categoria').count()['Count']).reset_index()
df2['Percentage'] = df2['Count'] / df2['Count'].sum()
plt.figure(figsize=(12.8,6))
sns.barplot(x=df2['Categoria'], y=df2['Percentage'])
plt.ylabel('%')
plt.xlabel('Categoria')
plt.title('Porcentagem de emails em cada categoria')
plt.show()
```

```
df['Tamanho_emails'] = df['Conteúdo'].str.len()
plt.figure(figsize=(12.8,6))
sns.distplot(df['Tamanho_emails'])
plt.ylabel('Distribuição')
plt.xlabel('Tamanho dos emails')
plt.title('Distribuição do tamanho dos emails')
plt.show()
```

```
with open('df.pickle', 'wb') as output:
    pickle.dump(df, output)
```

3.2. TRATAMENTO DE TEXTO

```
import pickle
import nltk
from nltk.stem import WordNetLemmatizer
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split
```

```
#Carregar a base de dados
path_df = 'df.pickle'
```

```
data = open(path_df, 'rb')
df = pickle.load(data)
data.close()
```

```
# Preparação e limpeza do texto - espaçamento
```

```

df['Conteúdo_Analisado_1'] = df['Conteúdo'].str.replace('\r', ' ')
df['Conteúdo_Analisado_1'] = df['Conteúdo_Analisado_1'].str.replace('\n', ' ')
df['Conteúdo_Analisado_1'] = df['Conteúdo_Analisado_1'].str.replace(' ', '')

# Transformação letras maiúsculas em minúsculas para uniformizar o texto
df['Conteúdo_Analisado_2'] = df['Conteúdo_Analisado_1'].str.lower()

# Sinais de pontuação
sinais_pontuação = list("?!.,;")
df['Conteúdo_Analisado_3'] = df['Conteúdo_Analisado_2']

for sinal_pontuação in sinais_pontuação:
    df['Conteúdo_Analisado_3'] = df['Conteúdo_Analisado_3'].str.replace(sinal_pontuação, "")

df['Conteúdo_Analisado_4'] = df['Conteúdo_Analisado_3'].str.replace("'", "")

# Lematização
# Descarregar punkt e wordnet do complemento NLTK
nltk.download('punkt')
print("-----")
nltk.download('wordnet')

# Guardar o lematizado sob a forma de objeto
wordnet_lematizador = WordNetLemmatizer()

n_linhas = len(df)
lista_texto_lematizado = []

for linha in range(0, n_linhas):

    # Criar uma lista vazia que contenha as palavras lemmatized
    lista_lematizada = []

    # Guardar o texto e respetivas palavras sob a forma de objeto
    texto = df.loc[linha]['Conteúdo_Analisado_4']
    palavras_texto = texto.split(" ")

    # Iterar por meio de cada palavra para lematizar
    for palavra in palavras_texto:
        lista_lematizada.append(wordnet_lematizador.lemmatize(palavra, pos="v"))

    # Juntar à lista
    texto_lematizado = " ".join(lista_lematizada)

    # Acrescentar à lista que contém os textos
    lista_texto_lematizado.append(texto_lematizado)

df['Conteúdo_Analisado_5'] = lista_texto_lematizado

# Descarregar a lista das stopwords
nltk.download('stopwords')

```

```

# Apresentar as stopwords em português
stop_words = nltk.corpus.stopwords.words('portuguese')
stop_words[0:10]
len(stop_words)
'a' in stop_words
print(stop_words)

df['Conteúdo_Analisado_6'] = df['Conteúdo_Analisado_5']

for stop_word in stop_words:
    exp_stopword = r"\b" + stop_word + r"\b"
    df['Conteúdo_Analisado_6'] = df['Conteúdo_Analisado_6'].str.replace(exp_stopword, "")

df.loc[5]['Conteúdo']
df.loc[5]['Conteúdo_Analisado_1']
df.loc[5]['Conteúdo_Analisado_2']
df.loc[5]['Conteúdo_Analisado_3']
df.loc[5]['Conteúdo_Analisado_4']
df.loc[5]['Conteúdo_Analisado_5']
df.loc[5]['Conteúdo_Analisado_6']
df.head(1)
lista_colunas = ["Nome_ficheiro", "Categoria", "Conteúdo", "Conteúdo_Analisado_6"]
df = df[lista_colunas]

df = df.rename(columns={'Conteúdo_Analisado_6': 'Conteúdo_Analisado'})

# Treino - teste separação
corpus_train, corpus_test, y_train, y_test = train_test_split(df['Conteúdo_Analisado'], df['Categoria'],
test_size=0.2, random_state=8, stratify=df['Categoria'])

# Representação texto
# Seleção parâmetros
ngram_range = (1,1)
min_df = 1
max_df = 1.
max_features = None

tfidf = TfidfVectorizer(encoding='utf-8', ngram_range=ngram_range, stop_words=None,
lowercase=False, max_df=max_df, min_df=min_df, max_features=max_features, norm='l2',
sublinear_tf=False)

X_train = tfidf.fit_transform(corpus_train).toarray()
print(X_train.shape)

X_test = tfidf.transform(corpus_test).toarray()
print(X_test.shape)

# X_train
with open('X_train.pickle', 'wb') as output:

```

```

    pickle.dump(X_train, output)

# X_test
with open('X_test.pickle', 'wb') as output:
    pickle.dump(X_test, output)

# y_train
with open('y_train.pickle', 'wb') as output:
    pickle.dump(y_train, output)

# y_test
with open('y_test.pickle', 'wb') as output:
    pickle.dump(y_test, output)

# TF-IDF object
with open('tfidf.pickle', 'wb') as output:
    pickle.dump(tfidf, output)

```

3.3. CLASSIFICADOR

```

import pickle
from sklearn.svm import LinearSVC
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
import matplotlib.pyplot as plt
import seaborn as sns

# Conjunto de dados (dataframe)
path_df = 'df.pickle'
with open(path_df, 'rb') as data:
    df = pickle.load(data)

# X_train
path_X_train = 'X_train.pickle'
with open(path_X_train, 'rb') as data:
    X_train = pickle.load(data)

# y_train
path_y_train = 'y_train.pickle'
with open(path_y_train, 'rb') as data:
    y_train = pickle.load(data)

# X_test
path_X_test = 'X_test.pickle'
with open(path_X_test, 'rb') as data:
    X_test = pickle.load(data)

# y_test
path_y_test = 'y_test.pickle'
with open(path_y_test, 'rb') as data:
    y_test = pickle.load(data)

```

```

print(X_train.shape)
print(X_test.shape)

clf = LinearSVC(penalty='l2', loss='squared_hinge', dual=True, tol=0.0001, C=1.0, multi_class='ovr',
fit_intercept=True, intercept_scaling=1, class_weight=None, verbose=0, random_state=None,
max_iter=1000)
clf.fit(X_train, y_train)

# PREVISÃO

clf_pred = clf.predict(X_test)

# Exatidão do treino
print("Exatidão do treino: ")
print(accuracy_score(y_train, clf.predict(X_train)))

# Exatidão do teste
print("Exatidão do teste: ")
print(accuracy_score(y_test, clf_pred))

# Relatório de classificação
print("Relatório de classificação")
print(classification_report(y_test, clf_pred))

# Matriz de confusão
labels = df['Categoria'].drop_duplicates().sort_values()
conf_matrix = confusion_matrix(y_test, clf_pred)
plt.figure(figsize=(12,8))
sns.heatmap(conf_matrix, annot=True, xticklabels=labels, yticklabels=labels, cmap="Blues")
plt.ylabel('Verdadeiro')
plt.xlabel('Previsto')
plt.title('Matriz de confusão')
plt.show()

with open('clf.pickle', 'wb') as output:
    pickle.dump(clf, output)

```

3.4. RESPOSTA AUTOMÁTICA

```

import pandas as pd
import pickle
import smtplib
from email.mime.multipart import MIMEMultipart
from email.mime.text import MIMEText
from getpass import getpass
from sys import argv
host = 'smtp.fe.up.pt'

```



```

def load_pickle(filename):
    pkl = open(filename, 'rb')
    data = pickle.load(pkl)
    pkl.close()
    return data

def predict(corpus):
    vectorizer = load_pickle('tfidf.pickle')
    clf = load_pickle('clf.pickle')
    X = vectorizer.transform(corpus)
    y_pred = clf.predict(X)
    return y_pred

def send_emails(fromaddr, password, y_pred, templates):
    subject = 'Teste de envio'
    toaddrs = ['ega12024@fe.up.pt']
    toaddrs_str = ', '.join(toaddrs)
    server = smtplib.SMTP(host=host, port=587)
    server.ehlo()
    server.starttls()
    server.ehlo()
    server.login(fromaddr, password)
    for pred in y_pred:
        msg = MIMEText()
        msg['Subject'] = subject
        msg['From'] = fromaddr
        msg['To'] = toaddrs_str
        body = templates[pred]
        msg.attach(MIMEText(body, 'plain'))
        text = msg.as_string()
        server.sendmail(fromaddr, toaddrs, text)
    server.quit()

def main():
    if len(argv) != 4:
        print("Usage: python3 %s <excel_file> <excel_column> <source_email>" % (argv[0]))
        return
    excel_file = argv[1]
    excel_column = argv[2]
    fromaddr = argv[3]
    templates = pd.read_excel('templates_das_categorias.xlsx')
    templates = dict([(templates['Categoria'][i], templates['Template'][i]) for i in range(len(templates))])
    print("Por favor, coloque a palavra-chave.")
    password = getpass()
    df = pd.read_excel(excel_file)
    corpus = df[excel_column]
    y_pred = predict(corpus)
    send_emails(fromaddr, password, y_pred, templates)

if __name__ == "__main__":
    main()

```

4. NÚMERO DE EMAILS E PERCENTAGEM DE EMAILS POR SUBCATEGORIA

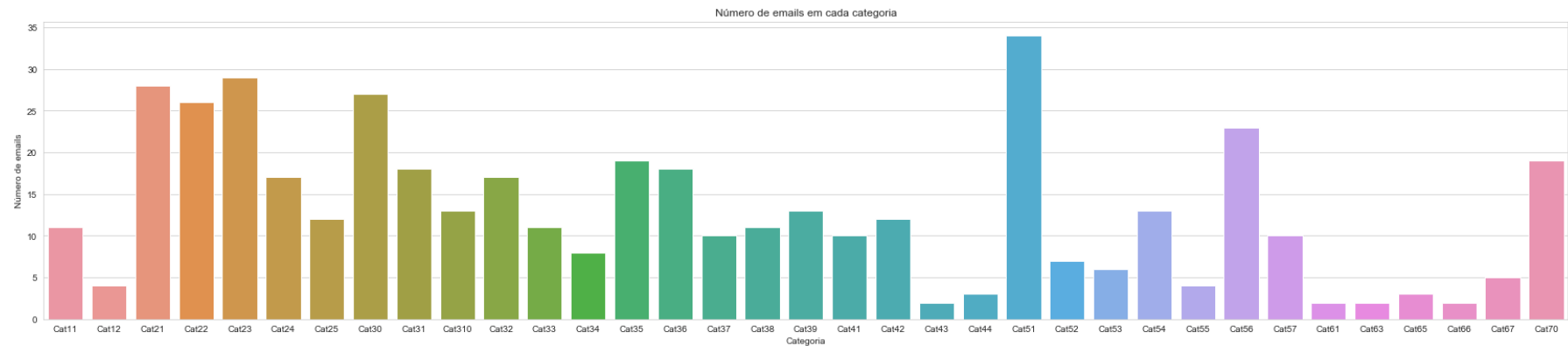


Fig. A4. 1 - Número de emails presentes em cada subcategoria

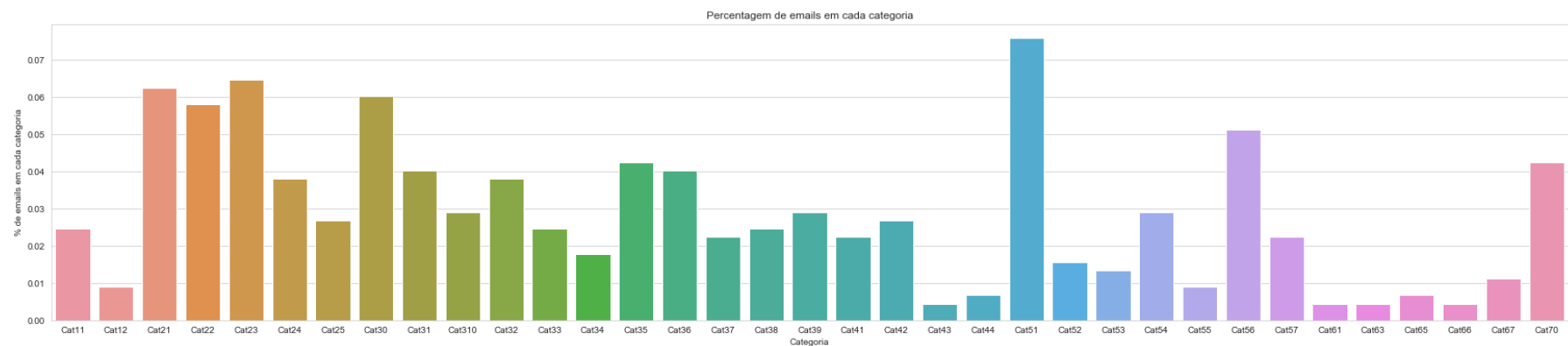


Fig. A4. 2 - Percentagem de emails presentes em cada subcategoria

5. RESPOSTAS TIPO PERSONALIZADAS PROJETADAS PARA O CÓDIGO

5.1. TEMPLATE PARA A CAT11

Caro Cliente,

Acusamos a receção do seu email que mereceu a nossa melhor atenção, e em resposta, informamos que iremos atender ao seu pedido de acesso ao Balcão Digital.

Para mais informações, consulte o link:

<https://www.aguasdoporto.pt/balcao-digital/c/como-registar-se>

Colocamo-nos ao seu dispor para qualquer esclarecimento adicional.

Com os melhores cumprimentos,

Serviço ao Cliente Águas do Porto



Para mais informações, por favor contacte-nos:

Email Geral/Avárias: geral@aguasdoporto.pt

Email Área Comercial: cliente@aguasdoporto.pt

Linha Verde Comercial: 800 206 235 (2.ª a 6.ª feira, das 9h00m às 17h).

5.2. TEMPLATE PARA A CAT21

Caro Cliente,

Na sequência da sua comunicação, informamos que o seu pedido para celebração do contrato de fornecimento de água e saneamento de águas residuais foi rececionado nas nossas instalações e encontra-se em processamento.

Para mais informações, consulte a seguinte hiperligação:

https://www.aguasdoporto.pt/cliente/novo-contrato#form_personal_data

Colocamo-nos ao seu dispor para qualquer esclarecimento adicional.

Com os melhores cumprimentos,

Serviço ao Cliente Águas do Porto



Contactos:

Email Geral/Avárias: geral@aguasdoporto.pt

Email Área Comercial: cliente@aguasdoporto.pt

Linha Verde Comercial: 800 206 235 (2.ª a 6.ª feira, das 9h00m às 17h).

5.3. TEMPLATE PARA A CAT22

Caro Cliente,

Na sequência da sua comunicação, informamos que o seu pedido para alteração de titularidade foi rececionado nas nossas instalações e encontra-se em processamento.

Para mais informações, consulte a seguinte hiperligação:

https://www.aguasdoporto.pt/cliente/novo-contrato#form_personal_data

Colocamo-nos ao seu dispor para qualquer esclarecimento adicional.

Com os melhores cumprimentos,

Serviço ao Cliente Águas do Porto



Contactos:

Email Geral/Avárias: geral@aguasdoporto.pt

Email Área Comercial: cliente@aguasdoporto.pt

Linha Verde Comercial: 800 206 235 (2.ª a 6.ª feira, das 9h00m às 17h).

5.4. TEMPLATE PARA A CAT23

Caro Cliente,

Na sequência da sua comunicação, informamos que o seu pedido para cancelamento do contrato de fornecimento de água e saneamento de águas residuais foi rececionado nas nossas instalações e encontra-se em processamento.

Para mais informações, consulte a seguinte hiperligação:

<https://www.aguasdoporto.pt/cliente/rescisao-contratual>

Colocamo-nos ao seu dispor para qualquer esclarecimento adicional.

Com os melhores cumprimentos,

Serviço ao Cliente Águas do Porto



Contactos:

Email Geral/Avárias: geral@aguasdoporto.pt

Email Área Comercial: cliente@aguasdoporto.pt

Linha Verde Comercial: 800 206 235 (2.ª a 6.ª feira, das 9h00m às 17h).

5.5. TEMPLATE PARA A CAT24

Caro Cliente,

Na sequência da sua comunicação, informamos que o seu pedido para alteração da morada de envio de faturação foi rececionado nas nossas instalações e encontra-se em processamento.

Colocamo-nos ao seu dispor para qualquer esclarecimento adicional.

Com os melhores cumprimentos,

Serviço ao Cliente Águas do Porto



Contactos:

Email Geral/Avárias: geral@aguasdoporto.pt

Email Área Comercial: cliente@aguasdoporto.pt

Linha Verde Comercial: 800 206 235 (2.ª a 6.ª feira, das 9h00m às 17h).

5.6. TEMPLATE PARA A CAT31

Caro Cliente,

Em resposta à sua solicitação no âmbito da adesão ao débito direto, confirmamos a sua receção e contactá-lo-emos com a maior brevidade possível. Receberá o documento "Autorização de Débito Direto SEPA", o qual deve ser assinado e devolvido à Águas do Porto para ativação do serviço.

Para mais informações, consulte a seguinte página do nosso website:
<https://www.aguasdoporto.pt/cliente/debito-direto-sepa>

Colocamo-nos ao seu dispor para qualquer esclarecimento adicional.

Com os melhores cumprimentos,

Serviço ao Cliente Águas do Porto



Contactos:

Email Geral/Avárias: geral@aguasdoporto.pt

Email Área Comercial: cliente@aguasdoporto.pt

5.7. TEMPLATE PARA A CAT32

Caro Cliente,

Acusamos a receção do seu email que mereceu a nossa melhor atenção, e em resposta, informamos que iremos atender ao seu pedido de alteração de IBAN/NIB associado à modalidade de débito direto.

Colocamo-nos ao seu dispor para qualquer esclarecimento adicional.

Com os melhores cumprimentos,

Serviço ao Cliente Águas do Porto



Para mais informações, por favor contacte-nos:

Email Geral/Avárias: geral@aguasdoporto.pt

Email Área Comercial: cliente@aguasdoporto.pt

Linha Verde Comercial: 800 206 235 (2.ª a 6.ª feira, das 9h00m às 17h).

5.8. TEMPLATE PARA A CAT33

Caro Cliente,

Em resposta à sua solicitação no âmbito da adesão à fatura eletrónica, confirmamos a sua receção e contactá-lo-emos com a maior brevidade possível.

Para mais informações, consulte a seguinte página do nosso website:
<https://www.aguasdoporto.pt/cliente/fatura-electronica>

Colocamo-nos ao seu dispor para qualquer esclarecimento adicional.

Com os melhores cumprimentos,

Serviço ao Cliente Águas do Porto



Contactos:

Email Geral/Avárias: geral@aguasdoporto.pt

Email Área Comercial: cliente@aguasdoporto.pt

Linha Verde Comercial: 800 206 235 (2.ª a 6.ª feira, das 9h00m às 17h).

5.9. TEMPLATE PARA A CAT34

Caro Cliente,

Acusamos a receção do seu email que mereceu a nossa melhor atenção, e em resposta, informamos que iremos atender ao seu pedido de alteração de email associado à modalidade de fatura eletrónica.

Colocamo-nos ao seu dispor para qualquer esclarecimento adicional.

Com os melhores cumprimentos,

Serviço ao Cliente Águas do Porto



Para mais informações, por favor contacte-nos:

Email Geral/Avarias: geral@aguasdoporto.pt

Email Área Comercial: cliente@aguasdoporto.pt

Linha Verde Comercial: 800 206 235 (2.ª a 6.ª feira, das 9h00m às 17h).

5.10. TEMPLATE PARA A CAT35

Caro Cliente,

Acusamos a receção do seu email que mereceu a nossa melhor atenção, e em resposta, informamos que iremos atender ao seu pedido de 2ª via da(s) fatura(s) em questão.

Colocamo-nos ao seu dispor para qualquer esclarecimento adicional.

Com os melhores cumprimentos,

Serviço ao Cliente Águas do Porto



Para mais informações, por favor contacte-nos:

Email Geral/Avarias: geral@aguasdoporto.pt

Email Área Comercial: cliente@aguasdoporto.pt

Linha Verde Comercial: 800 206 235 (2.ª a 6.ª feira, das 9h00m às 17h).

5.11. TEMPLATE PARA A CAT36

Caro Cliente,

Na sequência da sua comunicação, informamos que o seu pedido para análise/correção de fatura(s) foi rececionado nas nossas instalações e encontra-se em processamento.

Colocamo-nos ao seu dispor para qualquer esclarecimento adicional.

Com os melhores cumprimentos,

Serviço ao Cliente Águas do Porto



Contactos:

Email Geral/Avarias: geral@aguasdoporto.pt

Email Área Comercial: cliente@aguasdoporto.pt

Linha Verde Comercial: 800 206 235 (2.ª a 6.ª feira, das 9h00m às 17h).

5.12. TEMPLATE PARA A CAT37

Caro Cliente,

Na sequência da sua comunicação, informamos que a leitura transmitida será registada nos nossos serviços.

Para mais informações, consulte a seguinte página do nosso website:
<https://www.aguasdoporto.pt/cliente/leitura-contador>

Colocamo-nos ao seu dispor para qualquer esclarecimento adicional.

Com os melhores cumprimentos,

Serviço ao Cliente Águas do Porto



Contactos:

Email Geral/Avarias: geral@aguasdoporto.pt

Email Área Comercial: cliente@aguasdoporto.pt

Linha Verde Comercial: 800 206 235 (2.ª a 6.ª feira, das 9h00m às 17h).

5.13. TEMPLATE PARA A CAT38

Caro Cliente,

Acusamos a receção do seu email que mereceu a nossa melhor atenção, e em resposta, informamos que iremos atender ao seu pedido de extrato de conta corrente relativo ao período solicitado.

Colocamo-nos ao seu dispor para qualquer esclarecimento adicional.

Com os melhores cumprimentos,

Serviço ao Cliente Águas do Porto



Para mais informações, por favor contacte-nos:

Email Geral/Avárias: geral@aguasdoporto.pt

Email Área Comercial: cliente@aguasdoporto.pt

Linha Verde Comercial: 800 206 235 (2.ª a 6.ª feira, das 9h00m às 17h).

5.14. TEMPLATE PARA A CAT39

Caro Cliente,

Acusamos a receção do seu email que mereceu a nossa melhor atenção, e em resposta, informamos que iremos atender ao seu pedido de nota de crédito.

Colocamo-nos ao seu dispor para qualquer esclarecimento adicional.

Com os melhores cumprimentos,

Serviço ao Cliente Águas do Porto



Para mais informações, por favor contacte-nos:

Email Geral/Avárias: geral@aguasdoporto.pt

Email Área Comercial: cliente@aguasdoporto.pt

Linha Verde Comercial: 800 206 235 (2.ª a 6.ª feira, das 9h00m às 17h).

5.15. TEMPLATE PARA A CAT310

Caro Cliente,

Na sequência da sua comunicação, informamos que após a data limite de pagamento não é possível gerar nova referência MB. Poderá deslocar-se aos nossos balcões de atendimento (Sede, situada na Rua Barão de Nova Sintra, ou Gabinete do Município) ou aguardar pela fatura seguinte.

Colocamo-nos ao seu dispor para qualquer esclarecimento adicional.

Com os melhores cumprimentos,

Serviço ao Cliente Águas do Porto



Contactos:

Email Geral/Avárias: geral@aguasdoporto.pt

Email Área Comercial: cliente@aguasdoporto.pt

Linha Verde Comercial: 800 206 235 (2.ª a 6.ª feira, das 9h00m às 17h).

5.16. TEMPLATE PARA A CAT41

Caro Cliente,

No seguimento do seu email, informamos que o seu pedido de intervenção no âmbito da rede de saneamento está a ser analisado pelos nossos colaboradores. Contactá-lo-emos com a máxima brevidade possível, de modo a agendar uma visita a efetuar pela equipa técnica ao local em questão.

Para mais informações, consulte o link: <https://www.aguasdoporto.pt/paginas/pedido-de-servico>

Colocamo-nos ao seu dispor para qualquer esclarecimento adicional.

Com os melhores cumprimentos,

Serviço ao Cliente Águas do Porto



Contactos:

Email Geral/Avárias: geral@aguasdoporto.pt

Email Área Comercial: cliente@aguasdoporto.pt

Linha Verde Comercial: 800 206 235 (2.ª a 6.ª feira, das 9h00m às 17h).

5.17. TEMPLATE PARA A CAT42

Caro Cliente,

No seguimento do seu email, informamos que o seu pedido no âmbito do contador está a ser analisado pelos nossos colaboradores. Contactá-lo-emos com a máxima brevidade possível, de modo a agendar uma visita a efetuar por um técnico ao local em questão.

Colocamo-nos ao seu dispor para qualquer esclarecimento adicional.

Com os melhores cumprimentos,

Serviço ao Cliente Águas do Porto



Contactos:

Email Geral/Avárias: geral@aguasdoporto.pt

Email Área Comercial: cliente@aguasdoporto.pt

Linha Verde Comercial: 800 206 235 (2.ª a 6.ª feira, das 9h00m às 17h).

5.18. TEMPLATE PARA A CAT43

Caro Cliente,

No seguimento do seu email, informamos que o seu pedido de intervenção no âmbito da rede de abastecimento de água está a ser analisado pelos nossos colaboradores. Contactá-lo-emos com a máxima brevidade possível, de modo a agendar uma visita a efetuar pela equipa técnica ao local em questão.

Para seguir os pedidos de serviço em curso, basta aceder a:
<https://www.aguasdoporto.pt/paginas/pedido-de-servico>

Colocamo-nos ao seu dispor para qualquer esclarecimento adicional.

Com os melhores cumprimentos,

Serviço ao Cliente Águas do Porto



Contactos:

Email Geral/Avárias: geral@aguasdoporto.pt

Email Área Comercial: cliente@aguasdoporto.pt

Linha Verde Comercial: 800 206 235 (2.ª a 6.ª feira, das 9h00m às 17h).

5.19. TEMPLATE PARA A CAT44

Caro Cliente,

No seguimento do seu email, informamos que o seu pedido de intervenção no âmbito do serviço de águas pluviais está a ser analisado pelos nossos colaboradores. Contactá-lo-emos com a máxima brevidade possível, de modo a agendar uma visita a efetuar pela equipa técnica ao local em questão.

Para seguir os pedidos de serviço em curso, basta aceder a:
<https://www.aguasdoporto.pt/paginas/pedido-de-servico>

Colocamo-nos ao seu dispor para qualquer esclarecimento adicional.

Com os melhores cumprimentos,

Serviço ao Cliente Águas do Porto



Contactos:

Email Geral/Avárias: geral@aguasdoporto.pt

Email Área Comercial: cliente@aguasdoporto.pt

Linha Verde Comercial: 800 206 235 (2.ª a 6.ª feira, das 9h00m às 17h).

5.20. TEMPLATE PARA A CAT51

Caro Cliente,

Em resposta ao seu email, informamos que o seu pedido de condições de ligação (PT - Planta Topográfica) foi rececionado nos nossos serviços e prosseguirá para análise. Entraremos em contacto consigo brevemente.

Para mais informações, consulte os links:
<https://www.aguasdoporto.pt/licenciamento/licenciamento-digital>
<https://www.aguasdoporto.pt/licenciamento/-pedido-de-informacao-sobre-as-condicoes-de-ligacao-as-redes-publicas-pt-planta-topografica?token=SXzaHpzSFxxTvNrzd20pktNtiB227xUqBOOiltM>

Colocamo-nos ao seu dispor para qualquer esclarecimento adicional.

Com os melhores cumprimentos,

Serviço ao cliente Águas do Porto



Para mais informações, por favor contacte-nos:

Email Geral/Avarias: geral@aguasdoporto.pt

Email Área Comercial: cliente@aguasdoporto.pt

Linha Verde Comercial: 800 206 235 (2.ª a 6.ª feira, das 9h00m às 17h).

5.21. TEMPLATE PARA A CAT52

Caro Cliente,

No seguimento do seu email, informamos que o seu pedido de autorização de utilização das redes prediais (vistoria) está a ser analisado pelos nossos colaboradores. Contactá-lo-emos com a máxima brevidade possível, de modo a agendar uma visita a efetuar pela equipa técnica ao local em questão.

Para mais informações, consulte os links:

<https://www.aguasdoporto.pt/licenciamento/licenciamento-digital>

<https://www.aguasdoporto.pt/licenciamento/pedido-de-autorizacao-de-utilizacao-das-redes-prediais-vistoria-?token=KRIB8QQQtIDiUt2e5Hvmy8gIVLHROvLsPPmIQfW>

Colocamo-nos ao seu dispor para qualquer esclarecimento adicional.

Com os melhores cumprimentos,

Serviço ao Cliente Águas do Porto



Contactos:

Email Geral/Avarias: geral@aguasdoporto.pt

Email Área Comercial: cliente@aguasdoporto.pt

Linha Verde Comercial: 800 206 235 (2.ª a 6.ª feira, das 9h00m às 17h).

5.22. TEMPLATE PARA A CAT53

Caro Cliente,

Em resposta ao seu email, informamos que o seu pedido no âmbito de projetos de redes prediais para apreciação (projeto novo, projeto retificado, telas finais, aditamento) foi rececionado nos nossos serviços e prosseguirá para análise. Entraremos em contacto consigo brevemente.

Para mais informações, consulte os links:

<https://www.aguasdoporto.pt/licenciamento/licenciamento-digital>

<https://www.aguasdoporto.pt/licenciamento/projeto-das-redes-prediais-de-abastecimento-de-agua-drenagem-de-aguas-residuais-e-aguas-pluviais>

Colocamo-nos ao seu dispor para qualquer esclarecimento adicional.

Com os melhores cumprimentos,

Serviço ao cliente Águas do Porto



Para mais informações, por favor contacte-nos:

Email Geral/Avárias: geral@aguasdoporto.pt

Email Área Comercial: cliente@aguasdoporto.pt

Linha Verde Comercial: 800 206 235 (2.ª a 6.ª feira, das 9h00m às 17h).

5.23. TEMPLATE PARA A CAT54

Caro Cliente,

No seguimento do seu email, informamos que o seu pedido de execução de ligações às redes públicas (ramais) está a ser analisado pelos nossos colaboradores. Contactá-lo-emos com a máxima brevidade possível, de modo a agendar uma visita a efetuar pela equipa técnica ao local em questão.

Para mais informações, consulte os links:

<https://www.aguasdoporto.pt/licenciamento/licenciamento-digital>

<https://www.aguasdoporto.pt/licenciamento/ligacoes-as-redes-publicas?token=kfH98Zzj5LG4q1iMOvbaEjO2z6xvLGuuPtDgMNQI>

Colocamo-nos ao seu dispor para qualquer esclarecimento adicional.

Com os melhores cumprimentos,

Serviço ao Cliente Águas do Porto



Contactos:

Email Geral/Avárias: geral@aguasdoporto.pt

Email Área Comercial: cliente@aguasdoporto.pt

Linha Verde Comercial: 800 206 235 (2.ª a 6.ª feira, das 9h00m às 17h).

5.24. TEMPLATE PARA A CAT55

Caro Cliente,

No seguimento do seu email, informamos que a sua comunicação de início de obras está a ser analisado pelos nossos colaboradores.

Importa referir que a solicitação das ligações à rede pública deve ocorrer com a antecedência de 5 dias úteis através do Balcão Digital da Águas do Porto e caso seja necessário fornecimento de água para obras, deve requisitar a visita de um técnico para verificação das condições de instalação ou alteração do contador

Contactá-lo-emos com a máxima brevidade possível.

Para mais informações, consulte os links:

<https://www.aguasdoporto.pt/licenciamento/licenciamento-digital>

<https://www.aguasdoporto.pt/licenciamento/comunicacao-de-inicio-de-obra?token=eNOc7a1VfoU1aiBvRnrwStAiDgp9t2TimaubnPqB>

Colocamo-nos ao seu dispor para qualquer esclarecimento adicional.

Com os melhores cumprimentos,

Serviço ao Cliente Águas do Porto



Contactos:

Email Geral/Avárias: geral@aguasdoporto.pt

Email Área Comercial: cliente@aguasdoporto.pt

Linha Verde Comercial: 800 206 235 (2.ª a 6.ª feira, das 9h00m às 17h).

5.25. TEMPLATE PARA A CAT57

Caro Cliente,

No seguimento do seu email, informamos que o seu pedido de verificação de abastecimento de água/contador está a ser analisado pelos nossos colaboradores. Contactá-lo-emos com a máxima brevidade possível, de modo a agendar uma vistoria a efetuar por um técnico ao local em questão.

Colocamo-nos ao seu dispor para qualquer esclarecimento adicional.

Com os melhores cumprimentos,

Serviço ao Cliente Águas do Porto



Contactos:

Email Geral/Avárias: geral@aguasdoporto.pt

Email Área Comercial: cliente@aguasdoporto.pt

Linha Verde Comercial: 800 206 235 (2.ª a 6.ª feira, das 9h00m às 17h).

5.26. TEMPLATE PARA A CAT61

Caro Cliente,

No seguimento do seu email, informamos que o seu pedido de intervenção em tampa/grelha está a ser analisado pelos nossos colaboradores. Contactá-lo-emos com a máxima brevidade possível, de modo a acionar a deslocação da equipa técnica ao local em questão.

Colocamo-nos ao seu dispor para qualquer esclarecimento adicional.

Com os melhores cumprimentos,

Serviço ao Cliente Águas do Porto



Contactos:

Email Geral/Avárias: geral@aguasdoporto.pt

Email Área Comercial: cliente@aguasdoporto.pt

Linha Verde Comercial: 800 206 235 (2.ª a 6.ª feira, das 9h00m às 17h)

5.27. TEMPLATE PARA A CAT62

Caro Cliente,

No seguimento do seu email, informamos que o seu pedido de regularização de pavimento está a ser analisado pelos nossos colaboradores. Contactá-lo-emos com a máxima brevidade possível.

Colocamo-nos ao seu dispor para qualquer esclarecimento adicional.

Com os melhores cumprimentos,

Serviço ao Cliente Águas do Porto



Contactos:

Email Geral/Avárias: geral@aguasdoporto.pt

Email Área Comercial: cliente@aguasdoporto.pt

Linha Verde Comercial: 800 206 235 (2.ª a 6.ª feira, das 9h00m às 17h).

5.28. TEMPLATE PARA A CAT63

Caro Cliente,

No seguimento do seu email, informamos que o seu pedido de reposição das condições iniciais após obras efetuadas pela Águas do Porto está a ser analisado pelos nossos colaboradores. Contactá-lo-emos com a máxima brevidade possível, de modo a acionar a deslocação da equipa técnica ao local em questão.

Colocamo-nos ao seu dispor para qualquer esclarecimento adicional.

Com os melhores cumprimentos,

Serviço ao Cliente Águas do Porto



Contactos:

Email Geral/Avárias: geral@aguasdoporto.pt

Email Área Comercial: cliente@aguasdoporto.pt

Linha Verde Comercial: 800 206 235 (2.ª a 6.ª feira, das 9h00m às 17h).

5.29. TEMPLATE PARA A CAT64

Caro Cliente,

No seguimento do seu email, informamos que o seu pedido de religação do fornecimento de água na sequência de corte está a ser analisado pelos nossos colaboradores. Contactá-lo-emos com a máxima brevidade possível, de modo a acionar a deslocação da equipa técnica ao local em questão.

Colocamo-nos ao seu dispor para qualquer esclarecimento adicional.

Com os melhores cumprimentos,

Serviço ao Cliente Águas do Porto



Contactos:

Email Geral/Avárias: geral@aguasdoporto.pt

Email Área Comercial: cliente@aguasdoporto.pt

Linha Verde Comercial: 800 206 235 (2.ª a 6.ª feira, das 9h00m às 17h).

5.30. TEMPLATE PARA A CAT65

Caro Cliente,

No seguimento do seu email, informamos que o seu pedido de intervenção em fuga de água está a ser analisado pelos nossos colaboradores. Contactá-lo-emos com a máxima brevidade possível, de modo a acionar a deslocação da equipa técnica ao local em questão.

Colocamo-nos ao seu dispor para qualquer esclarecimento adicional.

Com os melhores cumprimentos,

Serviço ao Cliente Águas do Porto



Contactos:

Email Geral/Avárias: geral@aguasdoporto.pt

Email Área Comercial: cliente@aguasdoporto.pt

Linha Verde Comercial: 800 206 235 (2.ª a 6.ª feira, das 9h00m às 17h).

5.31. TEMPLATE PARA A CAT66

Caro Cliente,

Em resposta à sua solicitação, cumpre-nos informar que o seu pedido no âmbito de danos relacionados com inundações e/ou infiltrações foi rececionado nas nossas instalações e encontra-se em processamento.

Colocamo-nos ao seu dispor para qualquer esclarecimento adicional.

Com os melhores cumprimentos,

Serviço ao Cliente Águas do Porto



Contactos:

Email Geral/Avárias: geral@aguasdoporto.pt

Email Área Comercial: cliente@aguasdoporto.pt

Linha Verde Comercial: 800 206 235 (2.ª a 6.ª feira, das 9h00m às 17h).

6. CÓDIGO COM REDES NEURONAIS DESENVOLVIDO EM PYTHON

```
# Load Larger STM network and generate text
import sys
import numpy
from keras.models import Sequential
from keras.layers import Dense
from keras.layers import Dropout
from keras.layers import LSTM
from keras.callbacks import ModelCheckpoint
from keras.utils import np_utils
from pandas import read_excel
from sys import argv

# check arguments
n_args = len(argv)
mode = argv[1] if 1 < n_args else None
if not ((n_args == 2 and mode == "--train") or (n_args == 3 and mode == "--predict")):
    print("Usage: python3 %s --train" % (argv[0]))
    print("    python3 %s --predict <weights file>" % (argv[0]))
    quit()

# load text and convert to lowercase
df1 = read_excel('teste_emails_1.xlsx')
emails = [email for email in df1['Conteúdo']]
categories = [c for c in df1['Categoria']]
df2 = read_excel('templates_das_categorias.xlsx')
templates = dict((df2['Categoria'][i], df2['Template'][i]) for i in range(len(df2['Categoria'])))
raw_text = []
for email, category in zip(emails, categories):
    template = templates[category]
    raw_text.append(email)
    raw_text.append(template)
```

```

raw_text = '\0'.join(raw_text) # The '\0' character is expected to occur only when separating emails
from templates.
raw_text = raw_text.lower()
# create mapping of unique chars to integers, and a reverse mapping
chars = sorted(list(set(raw_text)))
char_to_int = dict((c, i) for i, c in enumerate(chars))
int_to_char = dict((i, c) for i, c in enumerate(chars))

# summarize the loaded data
n_chars = len(raw_text)
n_vocab = len(chars)
print("Total Characters: ", n_chars)
print("Total Vocab: ", n_vocab)
# prepare the dataset of input to output pairs encoded as integers
seq_length = 100
dataX = []
dataY = []
for i in range(0, n_chars - seq_length, 1):
    seq_in = raw_text[i:i + seq_length]
    seq_out = raw_text[i + seq_length]
    dataX.append([char_to_int[char] for char in seq_in])
    dataY.append(char_to_int[seq_out])
n_patterns = len(dataX)
print("Total Patterns: ", n_patterns)

# reshape X to be [samples, time steps, features]
X = numpy.reshape(dataX, (n_patterns, seq_length, 1))

# normalize
X = X / float(n_vocab)

# one hot encode the output variable
y = np_utils.to_categorical(dataY)

# define the LSTM model
model = Sequential()
model.add(LSTM(256, input_shape=(X.shape[1], X.shape[2]), return_sequences=True))
model.add(Dropout(0.2))
model.add(LSTM(256))
model.add(Dropout(0.2))
model.add(Dense(y.shape[1], activation='softmax'))

if mode == '--train':
    model.compile(loss='categorical_crossentropy', optimizer='adam')
    # define the checkpoint
    filepath="weights-improvement-{epoch:02d}-{loss:.4f}-bigger.hdf5"
    checkpoint = ModelCheckpoint(filepath, monitor='loss', verbose=1, save_best_only=True,
mode='min')
    callbacks_list = [checkpoint]
    # fit the model
    model.fit(X, y, epochs=50, batch_size=64, callbacks=callbacks_list)
elif mode == '--predict':
    # load the network weights

```

```

filename = argv[2]
model.load_weights(filename)
model.compile(loss='categorical_crossentropy', optimizer='adam')
# pick a random seed
start = numpy.random.randint(0, len(dataX)-1)
pattern = dataX[start]
print("Seed:")
print("\'", ".join([int_to_char[value] for value in pattern]), "\'")
#start = numpy.random.randint(0, len(emails))
#example = emails[start].lower()[0:100] # or [-100] or [len(example)//2-
50:len(example)//2+50]
#pattern = [char_to_int[char] for char in example]
#print("Seed:")
#print("\'", example, "\'")
# generate characters
#null_count = 0
#while True:
for i in range(1000):
    x = numpy.reshape(pattern, (1, len(pattern), 1))
    x = x / float(n_vocab)
    prediction = model.predict(x, verbose=0)
    index = numpy.argmax(prediction)
    result = int_to_char[index]
    seq_in = [int_to_char[value] for value in pattern]
    sys.stdout.write(result)
    sys.stdout.flush()
    pattern.append(index)
    pattern = pattern[1:len(pattern)]
    #if result == '\0':
    #    null_count = null_count + 1
    #    if null_count >= 2:
    #        break
print("\nDone.")
else:
    raise ValueError("Only the options '--train' and '--predict' are accepted.")

```