

ISTA - Escola de Tecnologias e Arquitectura

**eWOM para instituições públicas: aplicação ao caso  
do Exército Português**

**Joana de Azinhaes Horta**

Dissertação submetida como requisito parcial para obtenção do grau de **Mestre  
em Sistemas Integrados de Apoio à Decisão**

**Orientadores:**

Doutor Fernando Batista, Professor Associado,  
ISCTE-IUL – Instituto Universitário de Lisboa

Doutor João Ferreira, Professor Auxiliar (com Agregação),  
ISCTE-IUL – Instituto Universitário de Lisboa

11 de Julho de 2020



# Resumo

O crescimento das plataformas de *social media* veio facilitar o acesso à opinião pública acerca de diversos serviços e produtos. A recolha e análise dessa informação sobre as opiniões que os utilizadores partilham online (denominado *electronic Word-of-Mouth*), permite extrair conhecimento sobre a reputação das organizações. Embora esta reputação já seja monitorizada no setor privado, considera-se que a sua implementação no setor público poderá trazer benefícios no apoio à tomada de decisão sobre as políticas de governação institucionais, mas também na promoção de estratégias de marketing que beneficiem a sua imagem de utilidade pública e imagem de empregador.

Para responder à necessidade existente, o presente estudo teve como objetivo desenvolver uma metodologia que extraia informação relevante sobre o eWOM em *social media*, usando estratégias de *text mining* e processamento de língua natural. Esta metodologia foi aplicada no caso do Exército Português e revelou potencialidade para distinguir a polaridade de sentimentos em comentários, encontrar os principais tópicos emergentes e fornecer informação sobre a reputação institucional. De acordo com os resultados obtidos, foram propostas recomendações, potencialidades e limitações.

## Palavras chave

*eWOM*

*Social Media*

Reputação

Sector público



# ***Abstract***

easier access to public opinion on several services and products. The collection and analysis of this information regarding the opinions shared by users online (called electronic Word-of-Mouth), allows the extraction of knowledge about an organization's reputation. Although this reputation is already monitored in the private sector, its implementation in public sector institutions may also bring several benefits, especially in supporting decision-making on governance policies and in the development of marketing strategies that promote their image of public utility and their brand as an employer.

To offer a solution aimed at the needs of the public sector, the goal of this research was to develop a methodology that extracts relevant information about eWOM in social media, using text mining and natural language processing techniques. This methodology was applied to the context of the Portuguese Army, in which it revealed potential to distinguish feelings in comments, extract topics and provide indicators of institutional reputation. According to the results obtained, a set of recommendations, strengths and limitations was proposed.

## **Keywords**

eWOM

Social Media

Reputation

Public Sector



# ***Agradecimentos***

## ***Acknowledgements***

Aos professores Fernando Batista e João Ferreira, por me terem orientado e apoiado com visão, espírito crítico e disponibilidade constante. As suas sugestões, o rigor exigido e a confiança demonstrada, foram um contributo inestimável para impulsionar e transformar uma ideia numa solução eficaz.

Ao Artur, que apoiou o desenvolvimento deste trabalho através da incansável discussão de ideias, atitude positiva nos momentos difíceis e partilha nos momentos de sucesso. Mas também, por ser uma inesgotável fonte de incentivo, paciência e disponibilidade, capaz de me encorajar a ultrapassar todos os desafios ao longo do último ano.

Aos meus pais, pela confiança inabalável e amor incondicional. Não existem palavras capazes de expressar a gratidão pelo papel que sempre desempenharam no meu percurso pessoal, profissional e académico.

A todos os camaradas e amigos que pacientemente toleraram a ausência e cansaço durante o mestrado, sem prescindirem de uma atitude compreensiva e de apoio incondicional.

Lisboa, junho de 2020

Joana Azinhaes



# Conteúdo

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Motivação . . . . .	2
1.2	Objetivo e questões de investigação . . . . .	2
1.3	Metodologia de investigação . . . . .	3
1.4	Estrutura do documento . . . . .	4
<b>2</b>	<b>Conceitos introdutórios e trabalho relacionado</b>	<b>5</b>
2.1	Reputação no setor público . . . . .	5
2.2	eWOM . . . . .	6
2.3	<i>Social media</i> . . . . .	7
2.4	Trabalho relacionado . . . . .	8
<b>3</b>	<b>Metodologia para monitorização do eWOM</b>	<b>13</b>
3.1	Características e necessidades da instituição . . . . .	14
3.2	Identificação de potenciais fontes de informação . . . . .	15
3.3	Recolha da informação . . . . .	15
3.4	Preparação da informação . . . . .	17
3.5	Extração do conhecimento . . . . .	18
3.6	Avaliação . . . . .	19
3.7	Recomendações . . . . .	20
<b>4</b>	<b>Caso de estudo: eWOM no Exército Português</b>	<b>23</b>
4.1	Características e necessidades da instituição . . . . .	23
4.2	Identificação de potenciais fontes de informação . . . . .	26
4.2.1	Público-alvo . . . . .	27
4.2.2	Termos de pesquisa . . . . .	28
4.2.3	Seleção das fontes de informação . . . . .	28
4.3	Recolha da informação . . . . .	30
4.3.1	Extração da informação . . . . .	30
4.3.2	Exploração da informação . . . . .	34
4.3.3	Descrição dos resultados . . . . .	36
4.4	Preparação da informação . . . . .	46
4.5	Extração do conhecimento . . . . .	48
4.5.1	Análise de sentimento manual . . . . .	49

4.5.2	Análise de sentimento automática . . . . .	60
4.5.3	Análise de tópicos . . . . .	61
4.6	Avaliação . . . . .	62
4.6.1	Análise de sentimento manual . . . . .	62
4.6.2	Análise de sentimento automática . . . . .	62
4.6.3	Análise de tópicos . . . . .	63
4.7	Recomendações . . . . .	63
4.8	Sumário . . . . .	66
<b>5</b>	<b>Conclusões</b>	<b>69</b>

# Lista de Figuras

1.1	Processo DSRM . . . . .	4
2.1	Análise temporal do número de artigos do SCOPUS . . . . .	10
3.1	CRISP-DM . . . . .	14
4.1	Organização do Ministério da Defesa Nacional . . . . .	24
4.2	Análise temporal do efetivo militar . . . . .	25
4.3	Exemplo de atributos extraídos de uma publicação . . . . .	31
4.4	Exemplo de atributos extraídos de um comentário . . . . .	32
4.5	Exemplo de comentários com poucos caracteres . . . . .	35
4.6	Exemplo de comentários na rede social Instagram . . . . .	36
4.7	Exemplo de comentários com identificação de utilizadores . . . . .	36
4.8	Exemplo de comentário de estimulação à interação . . . . .	37
4.9	Exemplo de comentários com elevada especificidade . . . . .	37
4.10	Número de publicações por mês em 2019 . . . . .	40
4.11	Média das reações por publicação em cada fonte de informação . . . . .	42
4.12	Média de comentários por publicação em cada fonte . . . . .	43
4.13	Número médio de utilizadores por publicação para cada fonte de informação . . . . .	45
4.14	Média anual de comentários por publicação e tipo de fonte . . . . .	46
4.15	Nuvem de palavras - Termos nos comentários por tipo de fonte . . . . .	47
4.16	Número de comentários por sentimento e organismo público . . . . .	51
4.17	Comentários anuais por sentimento - Exército . . . . .	53
4.18	Nuvem de palavras - Exército . . . . .	54
4.19	Nuvem de palavras (TF-IDF) - Exército . . . . .	55
4.20	Comentários anuais por sentimento - Marinha . . . . .	56
4.21	Nuvem de palavras - Marinha . . . . .	57
4.22	Comentários anuais por sentimento - Força Aérea . . . . .	58
4.23	Nuvem de palavras - Força Aérea . . . . .	58
4.24	Comentários anuais por sentimento - Forças Armadas . . . . .	59
4.25	Nuvem de palavras - Forças Armadas . . . . .	60

# Lista de Tabelas

2.1	Número de publicações obtidas em pesquisas bibliográficas . . . . .	9
4.1	Atributos extraídos dos comentários . . . . .	32
4.2	Atributos extraídos das publicações . . . . .	32
4.3	Descritivas da informação extraída por fonte, comentários e publicações . . . . .	38
4.4	Comentários extraídos por fonte e ano . . . . .	39
4.5	Descrição das publicações com maior número de reações por tema . . . . .	41
4.6	Média de comentários por publicação dos principais utilizadores . . . . .	44
4.7	Número de utilizadores por fonte de informação . . . . .	44
4.8	Descritivas do Reddit por organismo público . . . . .	50
4.9	Percentagem de comentários por sentimento e organismo público . . . . .	51
4.10	Média de caracteres por comentário e sentimento . . . . .	52
4.11	Tópicos emergentes . . . . .	62
4.12	Resultados do treino de modelos de classificação automática . . . . .	63

# ***Lista de Acrónimos***

**FA** - Forças Armadas

**FAP** - Força Aérea Portuguesa

**MDN** - Ministério da Defesa Nacional

**EMGFA** - Estado-Maior-General das Forças Armadas

**DARH** - Direção de Administração de Recursos Humanos

**OE** - Objetivo Estratégico

**CEME** - Chefe do Estado-Maior do Exército

**ESE** - Escola de Sargentos do Exército

**NATO** - North Atlantic Treaty Organization

**eWOM** - electronic Word-of-mouth

**WOM** - Word-of-mouth

**CRISP-DM** - Cross Industry Standard Process for Data Mining

**DSRM** - Design Science Research Methodology

**FB** - Facebook

**SVM** - Support Vector Machine

**IDF** - Inverse Document Frequency

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

**LSA** - Latent Semantic Analysis

**LDA** - Latent Dirichlet Allocation

**NLTK** - Natural Language Toolkit



# Introdução



Com o uso generalizado da internet, os comportamentos e relações inerentes ao Homem foram reproduzidos no mundo digital. As aplicações de *social media* constituem um exemplo deste fenómeno, constituindo um ambiente desenhado para estimular comportamentos de partilha e interação online que são intrínsecos aos seus utilizadores. A crescente utilização destas aplicações (blogs, redes sociais ou comunidades online) fornece um conjunto de benefícios para os cibernautas, dos quais se destaca a possibilidade de obter informação sobre organizações, serviços e produtos (Hennig-Thurau et al., 2004).

A procura de informação online dá acesso ao *electronic Word-Of-Mouth* (eWOM) definido como “qualquer afirmação negativa ou positiva, realizada por um potencial, atual ou antigo cliente sobre um produto ou organização e que se encontra disponível para diversas pessoas e instituições através da Internet” (Hennig-Thurau et al., 2004). Esta informação, disponível publicamente, é prioritária no processo de tomada de decisão dos utilizadores (Ash, 1951; Banerjee, 1992), tornando-se crucial para o desempenho das organizações. Assim, o elevado volume de dados gerado em *social media*, está a tornar-se numa fonte de informação com potencial para, através de *business intelligence*, apoiar a gestão da reputação e aumentar a competitividade dos negócios (He et al., 2015).

Embora seja reconhecida a relevância do eWOM no setor privado, a sua monitorização no setor público é rara (Gurau, 2008) e recorre principalmente a métodos pouco personalizados, insuficientemente abrangentes e incapazes de fornecer informação atualizada (He et al., 2015). Este problema é agravado pela cultura organizacional vigente, pelas restrições orçamentais e baixas condições tecnológicas, que dificultam o desenvolvimento de soluções capazes de responder às necessidades do setor (Gurau, 2008).

Para ultrapassar essas limitações e contribuir para solução do problema, o presente estudo tem como objetivo desenvolver uma metodologia para monitorização do eWOM que recorra à extração automatizada de informação publicada em *social media* e à descoberta de tópicos e opiniões dos utilizadores através de técnicas de *text mining*. Este método visa capacitar as instituições com conhecimento sobre a perceção dos seus clientes e com recomendações sobre a gestão da sua imagem pública (Hennig-Thurau et al., 2004).

Como forma de testar a metodologia proposta e avaliar a sua aplicabilidade em organismos públicos, procedeu-se ao desenvolvimento e aplicação para o Exército Português.

Esta organização enfrenta dificuldades de atratividade e de competitividade no mercado de trabalho, com conseqüente diminuição do número de candidaturas para o serviço militar voluntário, em particular na categoria de Praças (Silva et al., 2019). A partir dos resultados obtidos, ficará demonstrado o potencial da metodologia para apoiar a gestão da reputação online do Exército, mas também a sua replicabilidade a outras instituições do âmbito da Defesa. Simultaneamente, serão abordadas as vantagens de monitorizar o eWOM como forma de avaliar a reputação institucional, e assim, melhorar as políticas de governação ou promover estratégias de marketing que beneficiem a sua imagem de utilidade pública e de empregador.

## 1.1 Motivação

As instituições do setor público português possuem um papel fundamental no funcionamento do país e representam uma parte significativa do orçamento de estado. A sua existência depende da gestão de uma reputação positiva junto da população portuguesa, a qual contribui para o seu funcionamento, financiamento e capital humano. Apesar disso, os organismos públicos possuem um investimento limitado para a sua divulgação e uma necessidade de métodos de monitorização sistemáticos, automatizados e eficazes, que permitam conhecer a perceção dos cidadãos face aos produtos ou serviços que disponibilizam. O presente estudo vai ao encontro desta necessidade através de uma solução para a análise da reputação institucional, em particular, no caso do Exército Português.

O Exército Português e os restantes ramos das Forças Armadas (FA) enfrentam graves problemas de imagem, atratividade e retenção de militares. Motivada por esta necessidade real e sabendo que nenhum obstáculo é intransponível, a dissertação aqui descrita, visa desenvolver uma solução acessível, inovadora e capaz de contribuir para a missão da instituição.

## 1.2 Objetivo e questões de investigação

O presente estudo tem como objetivo desenvolver uma metodologia que extraia informação relevante sobre o eWOM em *social media*, usando estratégias de *text mining* e processamento de língua natural. As questões em estudo são:

1. Qual o conjunto de fontes de informação com potencial para conter informação sobre a reputação de um organismo público?
2. Qual o subconjunto de fontes de informação que se aplica a um determinado organismo em estudo?
3. Quais os principais tópicos emergentes nos comentários sobre o organismo público?

4. A análise da polaridade dos comentários permite obter informação relevante sobre o organismo público?

### 1.3 Metodologia de investigação

Tendo como objetivo obter uma metodologia capaz de monitorizar o eWOM e responder às questões de investigação, a solução proposta será aplicada ao caso do Exército Português. Durante o desenvolvimento desta metodologia, pretende-se garantir que segue um conjunto de processos e fases rigorosos, sistemáticos e generalizáveis que permitam relacionar o comportamento humano (referente à perceção dos utilizadores) aos sistemas de informação (como meio de obtenção dos dados).

Abordar esta interação entre sistemas, organizações e pessoas, requer a criação de soluções práticas com explicação teórica e replicabilidade. Esta necessidade enquadra-se no conceito de *Design Science Research Methodology* (DSRM) (Peffer et al., 2007) que combina os paradigmas provenientes das ciências sociais e da engenharia para, através do método científico, produzir um conjunto de diretrizes capazes de conduzir, avaliar e apresentar soluções (artefactos) em problemas organizacionais identificados. Esta metodologia possui um conjunto de componentes e atividades que se podem ver na Figura 1.1 e que se referem aos seguintes passos de acordo com Peffer et al. (2007):

1. **Identificação do problema e motivação:** reconhecer o problema a abordar e a importância de desenvolver uma solução.
2. **Definição dos objetivos para a solução:** definir objetivos realistas para a solução desejada de acordo com o estado da arte.
3. **Desenhar e desenvolver:** criar um artefacto, o que pode referir-se a construtos, modelos, métodos, instâncias ou novas propriedades para recursos existentes. Este artefacto baseia-se nos objetivos e conhecimento científico relevante.
4. **Demonstrar:** realizar a prova de conceito através de experimentação, simulação ou estudo de caso.
5. **Avaliar:** medir a qualidade do artefacto através de métricas e técnicas de análise (e.g. medidas de desempenho) e comparar com os objetivos delineados. O tipo de avaliação realizada depende das necessidades do contexto de aplicação e objetivos da organização. Esta fase pode implicar melhorias da solução encontrada ou a aceitação para a atividade seguinte.
6. **Comunicar:** descrever as características e resultados dos passos anteriores, com rigor e de acordo com estruturas aceites pela comunidade científica.

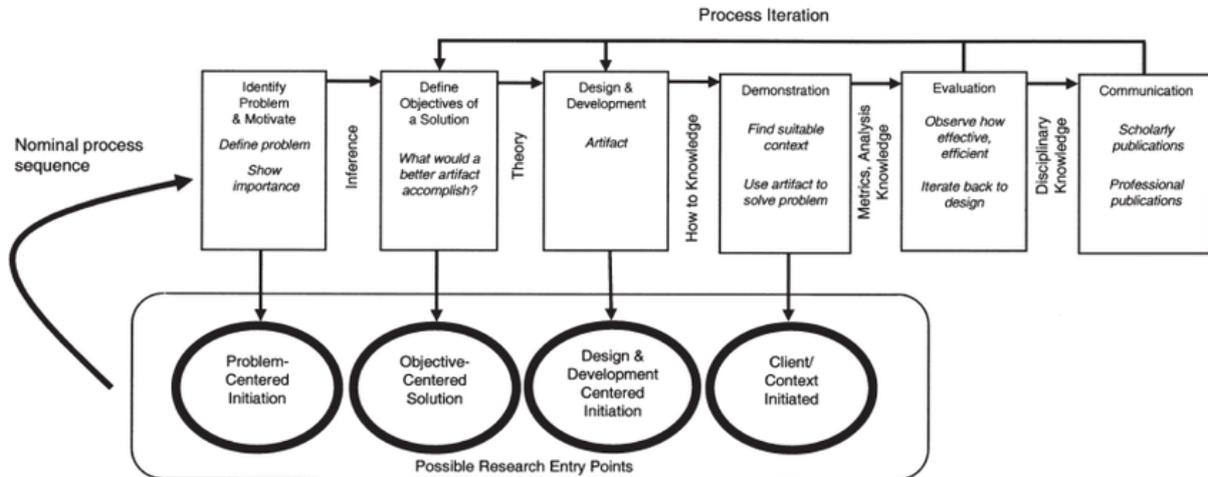


Figura 1.1: Processo DSRM, extraído de Peffers et al. (2007)

Considerando que o presente estudo pretende resolver um problema concreto mas generalizável a outros contextos, as diretrizes do DSRM foram utilizadas para desenhar e apresentar o estudo em conformidade com a comunidade científica. Nesse sentido, o processo inicial incidirá na descrição do problema em análise, nas soluções existentes e nos objetivos a alcançar com o desenvolvimento de uma nova solução. Relativamente ao artefacto construído, o desenho da solução foi baseado nas tarefas do Cross Industry Standard Process for Data Mining (CRISP-DM) (Wirth and Hipp, 2000). Por último, a sua demonstração e avaliação foi realizada no contexto do Exército e, como previsto em Hevner et al. (2004), a solução final consiste numa metodologia para extração de informação sobre o eWOM em *social media* com aplicabilidade demonstrada e com capacidade de generalização a diferentes contextos institucionais do setor público.

## 1.4 Estrutura do documento

O presente documento é composto por 5 capítulos que pretendem contextualizar, descrever e demonstrar a metodologia desenvolvida. Após o Capítulo introdutório, o Capítulo 2 descreve os conceitos que alicerçam a metodologia proposta e analisa as soluções existentes em problemas relacionados. Seguidamente, o Capítulo 3 descreve a metodologia desenvolvida, as respetivas fases, tarefas e processos. Para testar a solução proposta, o Capítulo 4 demonstra a aplicação da metodologia para monitorização do eWOM no caso do Exército Português e o seu potencial de aplicação a outros organismos públicos. Por último, o trabalho culmina no Capítulo 5 onde são resumidas as principais conclusões, limitações e sugestões para estudos futuros.

# Conceitos introdutórios e trabalho relacionado

# 2

Este capítulo pretende introduzir um conjunto de conceitos relacionados com o problema em estudo, analisar as soluções encontradas em literatura relacionada, e ainda, justificar as opções tomadas no desenvolvimento da metodologia.

A primeira secção define o conceito de reputação e analisa a relevância da sua gestão e monitorização no setor público. Seguidamente, é descrito o conceito de *electronic Word-Of-Mouth*, as suas características e relevância para a reputação e desempenho organizacional. A secção seguinte introduz os *social media* como meios de comunicação e partilha de informação pública online. Finalmente, o capítulo encerra com a análise de trabalhos relacionados onde se conclui que a monitorização do eWOM a partir da informação extraída em *social media* e analisada com técnicas de *text mining* constitui uma solução inovadora e com potencial para avaliar a reputação no setor público.

## 2.1 Reputação no setor público

A reputação pode definir-se como um conjunto de opiniões formadas a partir das características financeiras, sociais e ambientais atribuídas a uma organização ao longo do tempo. Este conceito é frequentemente associado aos termos «imagem» ou «marca», mas apesar da intersecção de conceitos, estes últimos distinguem-se principalmente pela ênfase na capacidade da empresa se destacar face ao mercado e de possuir símbolos diferenciadores (Barnett et al., 2006).

A gestão destas opiniões e perceções é uma prática generalizada no contexto empresarial, que investe na reputação para satisfazer os seus *stakeholders* e aumentar o nível de competitividade face ao mercado. Esta gestão tem sido igualmente sugerida para o setor público, o qual, também possui *stakeholders*, ou seja, indivíduos que constituem parte interessada nas organizações. Neste caso, a reputação refere-se à perceção formada pelos contribuintes, votantes, utilizadores dos serviços ou consumidores dos produtos fornecidos por entidades públicas (Luoma-aho, 2007).

Nos últimos anos, as instituições têm feito esforços no desenvolvimento de *slogans*, valores e missões (Christopher et al., 2002; Kirby and Marsden, 2006), porém a administração pública continua a desvalorizar a gestão da reputação devido à falta de competitividade no

mercado, restrições orçamentais e tecnológicas e à cultura organizacional (Gurau, 2008; Serrat, 2017). Contrariamente a esta tendência, tem sido sugerido que a utilização do erário público em ferramentas capazes de analisar e modificar a percepção dos utilizadores poderá constituir um passo fundamental para apoiar a tomada de decisão nas políticas de governação, para aumentar qualidade das respostas às necessidades das comunidades, para melhorar o recrutamento de trabalhadores para a função pública, e para promover o sentimento de partilha de responsabilidade e de participação dos cidadãos (Serrat, 2017).

De acordo com a análise realizada nesta secção, é possível deduzir que o setor público português partilha um conjunto de limitações comuns ao contexto descrito na literatura, pelo que também poderá beneficiar de uma gestão da opinião pública acerca das suas instituições.

## 2.2 eWOM

O ser humano é confrontado diariamente com inúmeras situações que requerem a sua tomada de decisão. Embora diversos mecanismos psicológicos intervenham neste processo, há várias décadas que as disciplinas de comportamento económico e social demonstram uma tendência para decidir em conformidade com grupos de pessoas (Ash, 1951; Banerjee, 1992). Esta influência social cria e modifica percepções através de distorções cognitivas associadas à falta de confiança, dificuldade em divergir dos outros ou sobrevalorização da capacidade do grupo para tomar boas decisões (Ash, 1951). Independentemente do enviesamento cognitivo que lhe dá origem, este padrão comportamental é adaptativo e facilitador da aprendizagem social. Assim, perante novas situações, os indivíduos recorrem à experiência dos outros como forma de obter informação, aprender quais as melhores opções e evitar erros por experiência própria (Bandura, 1971).

Quando esta influência interpessoal é aplicada à partilha de informação sobre produtos, serviços ou marcas com impacto nas decisões do cliente, passa a ser denominada por *Word-Of-Mouth* (WOM) (Hennig-Thurau et al., 2004; Zavattaro, 2018). Um caso particular desse conceito que cresce com a expansão da internet é o *electronic Word-of-Mouth*: “qualquer afirmação negativa ou positiva, realizada por um potencial, atual ou antigo cliente, acerca de um produto ou organização e que se encontra disponível para diversas pessoas e instituições através da Internet” (Hennig-Thurau et al., 2004). Ambos os conceitos de WOM e eWOM demonstram um efeito significativo na decisão dos potenciais interessados nos serviços, produtos ou oportunidades disponibilizadas pelas organizações. No entanto, a vertente eletrónica caracteriza-se pela rápida disseminação da mensagem, pela dificuldade em controlar os seus efeitos e pelo custo elevado na reversão que a informação negativa exerce sobre a imagem da empresa ou organismo (Kirby and Marsden, 2006). Esta comunicação entre utilizadores pretende fornecer informações ou opiniões e tem origem em fatores como a novidade de um produto/serviço; o nível de satisfação; a lealdade ou tempo

de relação com a organização; a discussão gerada por publicações ou anúncios; a obtenção de gratificação pessoal; e ainda, o apoio ou resolução de problemas (Hennig-Thurau et al., 2004; Zavattaro, 2018). Tipicamente, este eWOM ocorre através de um conjunto diversificado de meios, dos quais se destacam fóruns, redes sociais, microblogs, grupos de discussão, sites noticiosos e plataformas de *reviews* (Chu and Kim, 2011; Cheung and Thadani, 2012).

Independentemente das motivações geradoras de conversa, os potenciais clientes tendem a confiar mais no eWOM face aos meios publicitários tradicionais (Cheung and Thadani, 2012) e a basear-se nessa comunicação para modelar as suas expectativas, perceções, escolhas e aquisições (Jalilvand et al., 2011). Assim, conhecer e monitorizar este fenómeno online, adquire uma importância vital para a gestão da reputação, imagem e sucesso das organizações (Kirby and Marsden, 2006), com impacto no mercado de clientes e de trabalhadores, no relacionamento com fornecedores e na relação com influenciadores (Christopher et al., 2002).

Face ao exposto, considera-se que a informação partilhada online permite conhecer a perceção pública sobre as organizações, podendo influenciar o seu sucesso. A gestão desta informação é fundamental para controlar e beneficiar dos seus efeitos sociais e deve incidir nas plataformas de *social media*, onde se inserem os principais meios de desenvolvimento e propagação do eWOM.

## 2.3 Social media

Nos dias de hoje, o acesso público à informação é indissociável da internet. Foi a sua expansão que ofereceu aos utilizadores a capacidade de comunicar e interagir de forma acessível, global e rápida. Esta facilitação do relacionamento interpessoal tem sido considerada a característica impulsionadora da internet, satisfazendo a necessidade humana de obter gratificação (Ko et al., 2005). Com o surgimento da *web 2.0*, essa motivação foi potenciada pelo surgimento de *social media*: um conjunto de aplicações que permitem criar, partilhar e consultar conteúdos gerados pelos utilizadores.

Através de *social media*, os indivíduos conseguem facilmente – e sem custos - comunicar, interagir, entreter, relaxar, passar o tempo ou procurar e trocar informação (Whiting and Williams, 2013). Essas atividades, tornaram estas aplicações numa fonte de gratificação frequentemente utilizada, com conseqüente aumento da diversidade e funcionalidade das suas aplicações.

Atualmente, o conceito de *social media* inclui uma diversidade de aplicações, o que torna necessário e, simultaneamente, difícil categorizá-las. A título de exemplo, ambas as aplicações Youtube e Facebook permitem criar e partilhar conteúdos, possuindo no entanto, características bastante distintas. Apesar disso, têm sido propostos agrupamentos baseados no tipo de presença digital, distinguindo blogs, projetos colaborativos, redes sociais,

comunidades online, jogos e mundos virtuais (Kaplan and Haenlein, 2010). Ainda assim, torna-se difícil delimitar estes grupos e emergem categorias referentes a *microblogues*, *social bookmarks*, *media sharing*, entre outros (Sterne, 2010).

Com a diversidade de aplicações e a transferência da comunicação presencial para o meio digital, as organizações também acompanharam esta tendência. Surgiu então, a utilização de *social media* como meio de divulgação, diálogo, gestão do conhecimento e socialização junto de clientes e trabalhadores (Schlagwein and Hu, 2017). Este esforço de ligação entre indivíduos e organizações tem-se generalizado, como se pode verificar num estudo com 809 entrevistas em Portugal, onde a grande maioria dos inquiridos afirmou utilizar o Facebook, e em que mais de metade assumiu seguir empresas ou marcas na sua rede social de eleição (Marktest, 2019). Entre os principais benefícios desta interação entre os utilizadores e as marcas, salienta-se a capacidade para gerar um elevado fluxo de dados cuja monitorização permite aceder ao eWOM, e assim, obter informação sobre o mercado e a imagem organizacional (Constantinides, 2014), o que é relevante para as estratégias de aumento do lucro, diminuição dos custos e desenvolvimento da relação com o cliente (Sterne, 2010).

Globalmente, conclui-se que as aplicações de *social media* são meios de comunicação frequentemente utilizados e capazes de gerar um elevado volume de informação. As opiniões partilhadas e o acesso público sugerem que estas aplicações podem constituir fontes de informação úteis à gestão da reputação organizacional.

## 2.4 Trabalho relacionado

Para responder às necessidades do setor público e desenvolver uma metodologia capaz de extrair e analisar o seu eWOM, importa rever a literatura existente sobre problemas e soluções relacionadas. De acordo com a reflexão realizadas nas secções 2.1, 2.2 e 2.3, considerou-se que a informação partilhada online poderá refletir a opinião dos cidadãos e fornecer indicadores sobre a reputação institucional. Para analisar os trabalhos desenvolvidos sob esta premissa, procedeu-se à análise bibliográfica onde não foram encontrados estudos concretos que avaliem a reputação do setor público através do eWOM em *social media*.

Na Tabela 2.1 encontra-se o resumo dos resultados obtidos na pesquisa de artigos académicos em língua inglesa e cujos resumos continham os termos em estudo. Como se pode observar, não foram encontrados artigos que incorporassem os temas em análise, tendo sido necessária uma generalização para encontrar trabalhos que incidiam sobre estas áreas ou simplesmente as mencionassem por motivos diversos.

Apesar do baixo número de artigos publicados sobre a interseção dos conceitos em estudo, foi realizada uma pesquisa global destes termos na base de dados SCOPUS. Na Figura 2.1, é possível visualizar a análise temporal referente ao número crescente (escala

<b>Termos de pesquisa</b>	<b>B-On</b>	<b>Scopus</b>
«eWOM»	1051	551
«social media»	78 281	28 401
«public sector» OR «public entity» OR «government» OR «non-profit organizations»	1 084 494	372 400
«reputation»	55 675	23 159
«reputation» AND («public sector» OR «public entity» OR «government» OR «non-profit organizations»)	3368	1435
«reputation» AND «social media» AND («public sector» OR «public entity» OR «government» OR «non-profit organizations»)	121	20
«eWOM» AND «social media»	235	128
«eWOM» AND «reputation»	22	23
«eWOM» AND «reputation» AND «social media»	12	5
«eWOM» AND («public sector» OR «public entity» OR «government» OR «non-profit organizations»)	4	5
«eWOM» AND «social media» AND («public sector» OR «public entity» OR «government» OR «non-profit organizations»)	0	0
«eWOM» AND «reputation» AND («public sector» OR «public entity» OR «government» OR «non-profit organizations»)	0	0

Tabela 2.1: Número de publicações obtidas em pesquisas bibliográficas em Fevereiro de 2020

logarítmica) de trabalhos publicados sobre eWOM, *social media* e setor público, sendo este último aquele que mantém uma presença mais consistente na literatura. Esta figura constitui um indicador que, embora limitado, sugere a relevância e atualidade dos temas em estudo, em particular no âmbito de *social media* e eWOM, que desde 2004 surgiram e cresceram consideravelmente.

Na pesquisa bibliográfica realizada, o estudo da reputação em instituições no setor público revelou pouca evidência científica e, quando documentado, estava circunscrito a situações esporádicas e limitadas no tempo (por exemplo, análise ao efeito de campanhas realizadas) (Hoye and Lievens, 2007; Marktest, 2019). Adicionalmente, a literatura referia como principal método de recolha de informação, a aplicação de questionários desenvolvidos somente para o efeito (Luoma-aho, 2007).

Embora não tenham sido encontrados trabalhos correspondentes ao tema em análise, os estudos relacionados confirmaram que a avaliação do eWOM como forma de monitorizar a reputação de uma organização ou serviço é uma prática moderna e cada vez mais utilizada, com casos documentados nos setores bancário (Augusto and Torres, 2018), hoteleiro (Anagnostopoulou et al., 2020), comercial (Kousheshi et al., 2019), académico (Le et al., 2019), da saúde (Purcarea et al., 2015), entre outros.

As soluções capazes de avaliar esta opinião dos utilizadores distinguem duas componentes fundamentais: os métodos de recolha e os métodos de análise. No que se refere à obtenção dos dados, o estudo do WOM recorre tradicionalmente a questionários, *focus groups* e estudos de mercado (Constantinides, 2014), no entanto, observa-se que a sua vertente ele-

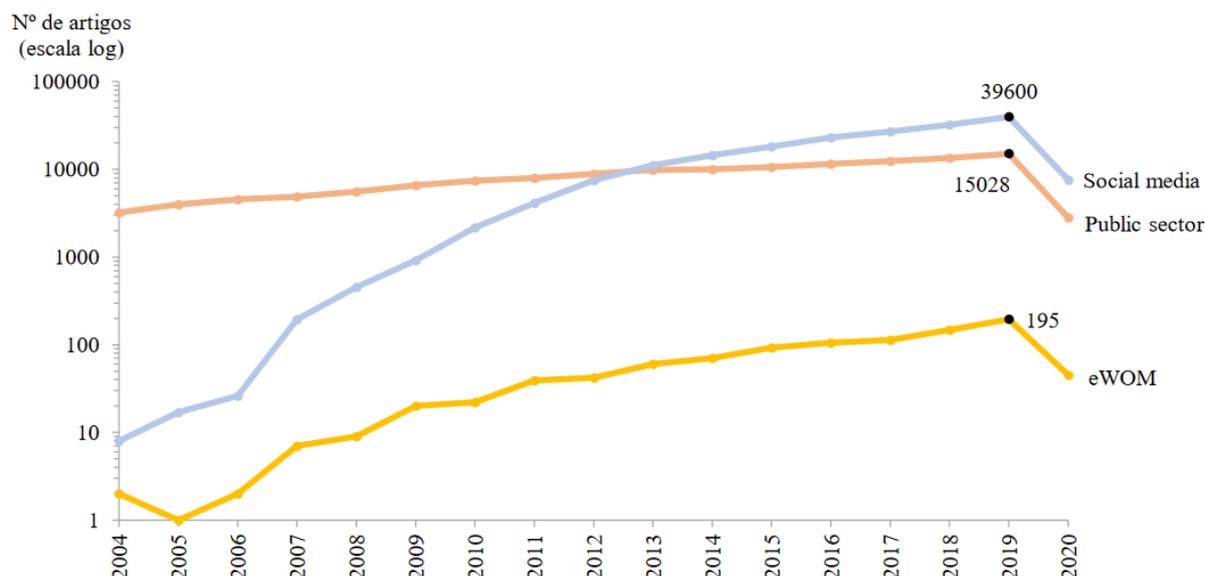


Figura 2.1: Análise temporal do número de artigos sobre «social media», «eWOM» e «public sector»

trónica inclui também fontes convencionais como entrevistas (Royo-Vela and Casamassina, 2011), questionários aplicados presencialmente (Kirby and Marsden, 2006; Christopher et al., 2002; Hoye and Lievens, 2007; Vázquez-Casielles et al., 2013), por telefone (Anderson, 1998) ou online (Lee and Youn, 2015; Poulis et al., 2019), e ainda alternativas, tais como métricas das redes sociais (Constantinides, 2014) e comentários/ reviews online (Le et al., 2019). Embora em número restrito, a recolha automatizada de dados para o eWOM também se encontra referida na literatura, nomeadamente através da utilização de *Application Program Interface* (API) em aplicações de *social media* como o «Twitter» (Jansen et al., 2009), o Facebook (Lo and Lin, 2017) ou o «TripAdvisor» (Saura et al., 2018). Estes métodos produzem benefícios significativos face aos meios comumente usados, destacando-se a extração rápida de uma maior quantidade de dados (devido à automatização do processo), com maior precisão (não depende da memória do indivíduo) e sem restrições (não se limita a questões concretas ou respostas socialmente aceites) (Vargo et al., 2019). No entanto, tal como os métodos tradicionais, estes meios automatizados também podem limitar-se a públicos e interações específicos da aplicação de *social media* utilizada, carecendo assim, da diversificação das fontes de informação para aumentar a generalização dos resultados (Muneta, 2017).

Relativamente aos métodos de análise utilizados, os modelos de eWOM descritos na literatura recorreram principalmente a estatística convencional com análise confirmatória e/ou exploratória. Estes métodos são frequentemente utilizados quando existe um processo de recolha pouco automatizado, um baixo volume de dados e necessidade de responder a perguntas concretas (Hand, 1999). Alternativamente, começam a ser aplicadas técnicas de *data mining* para dados textuais (*text mining*) que recorrem a algoritmos capazes de

analisar grandes conjuntos de dados não-estruturados e descobrir os padrões existentes (Aggarwal and Zhai, 2012). A análise do eWOM em *social media* com recurso a técnicas de *text mining* tem demonstrado potencial para descobrir os termos mais utilizados, as opiniões, as características das mensagens e a sua valência relacionada com os produtos ou serviços das marcas (Jansen et al., 2009; Sterne, 2010). Independentemente das técnicas aplicadas, o estudo da comunicação entre cibernautas acerca de uma organização deve incluir a análise de visitantes únicos, visitantes mais frequentes, comentários, “gostos”, sentimentos e diversos *Key Performance Indicators* (KPI) (Paine, 2011; Sterne, 2010).

Globalmente, o presente capítulo permitiu concluir que apesar da dificuldade em encontrar bibliografia correspondente aos conceitos e técnicas em estudo, a análise de trabalhos relacionados sugere que o setor público poderá beneficiar de uma solução inovadora e adaptada ao seu contexto, com recurso a métodos de recolha e de análise automáticos capazes de obter a informação sobre o eWOM presente em *social media*.



# Metodologia para monitorização do eWOM no setor público

# 3

Avaliar e gerir a reputação de uma instituição é fundamental para garantir o apoio dos cidadãos que pretende servir. Apesar dessa necessidade, o setor público não possui uma capacitação tecnológica e orientação para a competitividade semelhante ao setor privado. Assim, nas instituições portuguesas, o problema é agravado por dificuldades como o baixo volume de informação disponível online, estratégias de marketing pouco definidas, requisitos bastante específicos do seu contexto, organização tendencialmente hierárquica e pouco flexível e ainda, dificuldade em obter recursos humanos e monetários para investir na gestão da sua imagem interna ou externa. Para responder aos requisitos e condicionamentos do setor, este capítulo propõe uma metodologia composta por um conjunto de fases, tarefas e processos capazes de extrair uma amostra abrangente de opiniões em *social media* que, através de técnicas de *text mining* e processamento natural da língua, permitem obter uma ideia aproximada dos temas e sentimentos associados à reputação institucional. Pretende-se também expor a solução apresentada de forma personalizável, sistematizada e replicável, tornando-a acessível para diferentes organismos públicos.

Com esse objetivo, o desenvolvimento da solução foi baseado na metodologia *Cross Industry Standard Process for Data Mining* (CRISP-DM) comumente aplicada a projetos de *data mining* e que pretende obter processos replicáveis, fidedignos e rápidos, de forma a diminuir a tendência para desenvolver soluções demasiado específicas (Wirth and Hipp, 2000). Este processo de criação possui 6 fases que se interligam, e cuja sequência depende das características do projeto. Essa flexibilidade do CRISP-DM permitiu a adaptação dos seus processos e respetiva nomenclatura na metodologia desenvolvida, da qual resultou a equivalência exposta na Figura 3.1. As principais adaptações decorrem do foco em instituições públicas, ausência de dados disponíveis, dificuldade em implementar um sistema totalmente automatizado, e ainda, da importância de desenvolver recomendações concretas.

Baseando-se nas fases da metodologia proposta, o capítulo inicia-se com as tarefas requeridas para obter as características e necessidades da organização. A partir deste conhecimento contextual, segue-se o processo de seleção das fontes de informação que visa escolher os meios presentes em *social media* capazes de cumprir os critérios de análise e responder aos requisitos institucionais. Esta fase é sucedida da recolha de informação, com diretrizes relativas aos procedimentos de extração automatizada e análise explora-

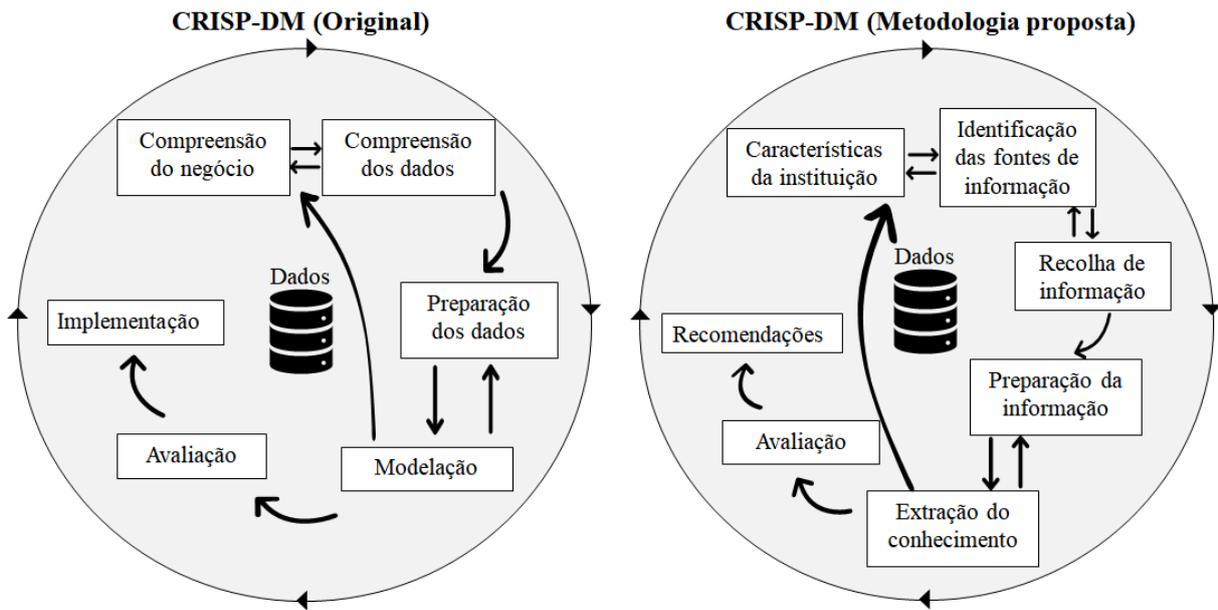


Figura 3.1: CRISP-DM (lado esquerdo) e metodologia proposta (lado direito) (adaptado de Wirth and Hipp (2000))

tória da informação extraída. Após estes passos, sugere-se um conjunto de técnicas para preparação da informação, seguida da fase de extração do conhecimento onde são descritos procedimentos manuais e de aprendizagem automática para a análise de sentimento. Por último, a metodologia encerra com a transformação do conhecimento em recomendações concretas de apoio à decisão.

Globalmente, este capítulo descreve uma metodologia desenvolvida com base no CRISP-DM capaz de responder às necessidades concretas do setor público e constituída por componentes automáticas e manuais, das quais resulta uma solução acessível e adaptável.

### 3.1 Características e necessidades da instituição

Tal como sugerido na metodologia CRISP-DM, é fundamental explorar as necessidades e características da entidade onde vai ocorrer a implementação da metodologia (Wirth and Hipp, 2000). Para isso, esta fase compreende a análise das características institucionais, tais como, a sua organização interna, terminologia específica, estratégia de marketing delineada, serviços e/ou produtos disponibilizados e a sua população-alvo. Esta avaliação do contexto culmina com o levantamento de necessidades ou limitações relevantes para a implementação, a identificação dos objetivos a alcançar e a delimitação das áreas de reputação avaliadas (a título de exemplo, reputação global, de empregador, de comunidade local, entre outros). A conclusão desta fase, permite orientar todas as fases posteriores da metodologia, em particular, na seleção, tratamento e interpretação da informação para o projeto de *data mining* implementado.

### 3.2 Identificação de potenciais fontes de informação

A partir do conhecimento obtido na Secção 3.1, esta fase inclui um conjunto de tarefas que visam identificar as fontes de informação relevantes para a instituição e respetiva reputação. Com esse intuito, sugerem-se os seguintes passos:

1. Identificar o público-alvo que gera ou consultar informação partilhada online acerca da instituição e que seja relevante para a prossecução dos objetivos institucionais;
2. Definir os termos identificativos da instituição e que podem ser utilizados na pesquisa de fontes de informação;
3. Selecionar as fontes de informação:
  - Análise dos principais canais de comunicação online entre a instituição e os seus utilizadores;
  - Pesquisa das fontes independentes de informação online onde se realizam discussões acerca da instituição e de acordo com a população-alvo definida;
  - Definição dos critérios de filtragem e exclusão das fontes de informação de acordo com os objetivos e requisitos delineados na fase antecedente;
  - Obter um conjunto de fontes suficientemente diversificado para garantir a representatividade da opinião institucional.

Esta etapa poderá depender de um processo iterativo entre a pesquisa das fontes de informação e a informação institucional fundamental para a sua seleção. No final, espera-se obter um conjunto alargado de fontes de informação online com potencial para obter informação sobre a opinião dos cibernautas, uma quantidade de conteúdo não estruturado adequado às técnicas de *text mining*, e ainda uma diversidade de comentários e utilizadores suficientemente elevada para garantir uma imagem generalizada da imagem institucional.

### 3.3 Recolha da informação

A terceira fase da metodologia, refere-se à recolha da informação contida nas fontes de informação selecionadas e inclui os processos de extração, exploração e descrição da informação. Importa salientar que os resultados obtidos podem implicar o regresso ao passo anterior (processo iterativo ilustrado na Figura 3.1) devido à recolha e exploração da informação: caso a informação recolhida não seja suficiente em quantidade e qualidade, será necessário procurar fontes alternativas.

As tarefas incluídas nesta fase permitem a obtenção e familiarização com a informação, mas também a análise da qualidade necessária para as técnicas de análise posteriores (Wirth and Hipp, 2000). Com esse objetivo, sugerem-se os seguintes passos e cuidados na extração de informação em *social media*:

### 1. Extração da informação:

- Seleção dos atributos: refere-se à definição dos atributos a extrair em cada fonte de informação (comentários, reações, publicações, entre outros), de acordo com a sua relevância para o modelo de *data mining*, os objetivos da instituição e do projeto delineado.
- Garantia da integridade dos registos: refere-se à criação de atributos para associação entre comentários e respetivas publicações (e.g. ID para publicação) e/ou entre comentários principais (resposta a publicação) e secundários (resposta a comentários). Este procedimento promove a integração dos dados e fontes de informação.
- Datas para recolha a informação: definir a janela temporal da informação que se pretendem obter, nomeadamente, selecionar a data do registo mais antigo e do registo mais recente. Em entidades do setor público nacional, aconselha-se um intervalo de vários anos, de forma a recolher um volume de informação suficiente para os algoritmos a utilizar.
- Definição de critérios de filtragem e exclusão das publicações, comentários ou *reviews*: de acordo com a informação extraída e a exploração realizada, definir critérios de inclusão ou exclusão do conteúdo extraído para melhorar o método de extração com conseqüente aumento da qualidade e pertinência da informação face aos objetivos propostos.
- Método de extração: refere-se à escolha, desenvolvimento e aplicação do método de recolha da informação. Sugere-se um processo de extração através de API (*application programming interface*) ou *web scrapers*. A primeira opção possui limitações que devem ser consideradas, nomeadamente, estar indisponível em algumas fontes de informação selecionadas (e.g. Fóruns específicos), conter restrições na quantidade de dados extraídos (e.g. Facebook) ou exigir um conjunto de requisitos acesso específicos. No que se refere à segunda opção, esta implica um esforço no desenvolvimento de *scripts* adequados a cada fonte de informação e na procura de alternativas às páginas que não permitem este tipo de extração, porém, depois de programado este processo de *web scraping* decorre de forma automática, rápida, com poucos custos e sem limite na quantidade de dados recolhidos. Após a escolha e desenvolvimento do processo, deverá realizar-se a extração da informação com a presença de um técnico capaz de garantir a execução do processo automatizado e que os dados estão a ser extraídos corretamente.

### 2. Exploração da informação:

- Análise às características da informação extraída para identificar fatores que poderão influenciar a qualidade dos resultados obtidos e dos quais se destacam: textos com baixo número de caracteres; a inclusão de *emojis*; a identificação de

utilizadores; a existência de atributos semelhantes em formatos diferentes; a obtenção de texto com ligações e/ou citações; a existência de publicações promotoras de respostas repetitivas ou ações específicas dos utilizadores; as publicações sem respostas; e ainda, textos com temáticas não ligadas à instituição estudada ou respostas a concursos online. Algumas aplicações online tendem a apresentar um maior número de limitações para a análise da comunicação entre utilizadores, por exemplo, as redes sociais «Facebook» e «Instagram» tendem a apresentar textos curtos e com necessidade de limpeza, enquanto os fóruns e comunidades online revelam comentários mais elaborados. Globalmente, esta exploração dos dados pode exigir o regresso à definição dos dados a extrair, seleção de novas fontes de dados e novos processos de extração.

### 3. Descrição dos resultados:

- Refere-se à caracterização e descrição global dos dados extraídos, tendo como objetivo verificar a sua viabilidade e colocar as primeiras hipóteses acerca do padrão encontrado.

Globalmente, esta secção descreve processos fundamentais para obter rapidamente um conjunto de informação alargada, atualizada, com qualidade e apta para diferentes níveis de análise. A recolha de informação pode resumir-se na escolha de um método de extração automatizado, programado para obter atributos alinhados com os objetivos do estudo, numa janela temporal definida e capaz de garantir a máxima qualidade e relação entre registos e atributos. Considera-se que esta fase está concluída após uma descrição dos resultados capaz de fornecer indicadores sobre comportamentos online e confirmada a presença de comentários/publicações com uma quantidade e diversidade significativa ao nível dos temas e dos sentimentos expressados.

## 3.4 Preparação da informação

A fase de preparação da informação visa transformar a informação recolhida em dados limpos, estruturados e integrados. Para isso, compreende um conjunto de procedimentos que devem incluir:

1. **Uniformização dos formatos:** quando existem atributos similares com formatos diferentes (e.g. data), definir e transformar do formato original para um formato comum a todos os ficheiros de extração;
2. **Junção da informação e criação de atributos derivados:** criar um *dataset* com dados extraídos das diversas fontes. Sugere-se a criação de novos atributos, em particular, um atributo que identifique a “fonte de informação” de origem, um “ID” para cada registo no *dataset* global e outros atributos calculados;

3. **Eliminação dos registos em branco:** algumas fontes de dados possuem um código de conduta específico e procedem à eliminação de texto proibido, deixando somente o nome do utilizador. Estes casos são extraídos sem conteúdo e podem ser excluídos do *dataset*.
4. **Pré-processamento da informação em *social media*:**
  - Transformação de palavras: uniformização das palavras escritas de forma diferente, mas com significado igual, por exemplo, os termos “IEFP” e “Instituto de Emprego e Formação Profissional” referem-se à mesma instituição. O conhecimento obtido na compreensão da instituição é fundamental para detetar estes termos, criar dicionários e uniformizar num termo comum.
  - Eliminação de identificações: Identificação e eliminação de utilizadores identificados no conteúdo dos comentários/ *reviews* extraídos. Dependendo da fonte de informação, a deteção e eliminação deste conteúdo poderá recorrer a Expressões Regulares (por exemplo, eliminação associada a caracteres específicos como o «@») ou listas de nomes (tais como listas de nomes e apelidos nacionais).
  - Normalização dos dados: inclui a aplicação de letras minúsculas a todas as palavras; a eliminação de pontuação; a eliminação de *stopwords* portuguesas, a tokenização e *stemming* (quando aplicável).

Esta secção de preparação da informação encerra o conjunto de fases metodológicas para obtenção do material necessário à análise. Assim, considera-se concluída quando a informação extraída fica transformada em dados prontos para computação e modelação.

### 3.5 Extração do conhecimento

A fase da Extração do Conhecimento pretende conhecer a opinião pública acerca da organização e os temas emergentes na comunicação entre cibernautas. Para atingir esse objetivo, a metodologia proposta inclui os seguintes passos:

- **Análise de sentimento:** a análise de sentimento (ou *opinion mining*) permite obter as opiniões, atitudes ou emoções dos utilizadores acerca de entidades ou atributos (Aggarwal and Zhai, 2012). Para as instituições portuguesas, a utilização de modelos treinados e disponibilizados em bibliotecas de fontes abertas é condicionada pela língua, características do contexto e terminologia específica, com consequente diminuição da precisão dos resultados. Para ultrapassar essas limitações, sugere-se:
  1. Classificação manual do sentimento em cada comentário: esta tarefa deve ser realizada por um avaliador especialista na área de estudo e requer disponibilidade para obter uma coleção alargada de comentários classificados.

2. Avaliação do nível de subjetividade da classificação manual: para garantir que é possível classificar sentimentos de forma sistemática, deverá proceder-se a uma segunda classificação dos comentários através de um segundo especialista.
  3. Análise de sentimento manual: a partir dos comentários classificados, podem ser desenvolvidos um conjunto de indicadores (a título de exemplo, sentimento público avaliado a partir do cálculo do número de comentários por sentimento), criados os dicionários de termos para cada sentimento (através da frequência de palavras) e/ou a visualização dos termos por sentimento (propõem-se nuvens de palavras).
  4. Análise de sentimento automática: sugere-se também, a utilização de técnicas de Aprendizagem Supervisionada capazes de detetar padrões no repositório de comentários classificado, e assim, treinar modelos de classificação automática capazes de avaliar o sentimento em novos comentários. Encontrar o melhor modelo para os dados extraídos deverá contemplar o ajustamento de parâmetros, a comparação entre diferentes modelos de classificação (a título de exemplo: *Naive Bayes*, Regressão Logística, Árvores de Decisão, *Support-vector machines* e Redes Neurais) e ainda, a aplicação de diferentes técnicas de normalização e extração de dimensões.
  5. Análise dos resultados: ao longo dos passos anteriores, sugere-se a elaboração de uma análise dos resultados e a inferência sobre o eWOM e reputação institucional, de acordo com o conhecimento da instituição e a literatura relacionada.
- **Análise de tópicos:** sugerem-se as técnicas de Aprendizagem Não-Supervisionada de *Latent Dirichlet Allocation* (LDA) ou *Latent Semantic Analysis* (LSA) que, através de processos de agrupamento e redução dimensional, permitem detetar tópicos latentes nos comentários extraídos (Aggarwal and Zhai, 2012). Para encontrar o melhor modelo de tópicos, sugere-se o ajustamento de parâmetros e a utilização de diferentes métodos de extração de dimensões.

A fase de Extração do Conhecimento considera-se concluída quando os dados são transformados em conhecimento sobre o eWOM institucional. Os resultados obtidos deverão fornecer informação sobre a imagem pública da instituição, a sua reputação e os principais temas que a envolvem.

### 3.6 Avaliação

A fase de Avaliação tem como objetivo avaliar e selecionar os modelos obtidos na fase anterior (Wirth and Hipp, 2000). Quando revela desempenhos satisfatórios, garante o grau de certeza necessário para gerar conclusões e generalizações a partir dos resultados obtidos. Para isso, devem ser considerados os seguintes passos ou valores de referência:

1. **Análise de sentimento manual:** avaliar a polaridade de sentimento aumenta o nível de subjetividade dos resultados e requer a comparação com um segundo avaliador. Esta consistência inter-avaliador pode ser avaliada através do *Kappa* de Cohen e, para que a variabilidade dos resultados seja aceitável, deverá apresentar valores de  $k \geq .60$ .
2. **Análise de sentimento automática:** Cada modelo de classificação automática de sentimento deve ser precedido de uma divisão treino-teste (valor de referência: 80%-20%) que permita a análise do seu desempenho. Para avaliar e comparar os modelos desenvolvidos, a respetiva matriz de confusão permite calcular as medidas de desempenho, como sugerido por Bengfort et al. (2018):

- **Accuracy:** Reflete o comportamento global do modelo a partir dos seus acertos, no entanto, não informa sobre a capacidade de classificar cada classe.

$$Accuracy = \frac{TP+TN}{P+N} \quad (1)$$

- **Precision:** Refere-se à capacidade do modelo prever uma classe de acordo com a frequência global dessa classe.

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

- **Recall:** Consiste numa medida de sensibilidade que demonstra a capacidade para detetar corretamente as classes relevantes.

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

- **F1:** Consiste na média harmónica das medidas *Precision* e *Recall*, sendo considerada uma medida mais informativa que a *Accuracy* do modelo.

$$F1 = 2 \cdot \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

3. **Análise de tópicos:** o modelo de tópicos obtido pelas técnicas LDA ou LSA, deve ser precedido da seleção de comentários para treino e teste. Para cada modelo obtido, sugere-se a comparação entre os tópicos atribuídos automaticamente à amostra de teste e o conteúdo do comentário.

### 3.7 Recomendações

Após a deteção e análise de padrões referentes ao eWOM em *social media*, esta fase tem como objetivo elaborar um conjunto de recomendações capazes de responder às necessidades definidas inicialmente e apoiar o processo de tomada de decisão sobre a reputação

institucional. Para isso, as recomendações devem basear-se no conhecimento institucional para transformar o conhecimento extraído em sugestões concretas de melhoria na gestão da imagem, gestão do eWOM, gestão institucional dos *social media* ou outras medidas relevantes para a estratégia de comunicação e marketing interno/externo.



# ***Caso de estudo: Monitorização do eWOM no Exército Português***

# 4

Para validar a metodologia descrita no Capítulo 3, é necessário testar o seu desempenho na extração de informação relevante sobre o eWOM em *social media*, com recurso a estratégias de *text mining* e processamento de língua natural. Tendo como objetivo realizar essa prova de conceito, foi utilizado o caso do Exército Português: uma instituição com dificuldades de atração e retenção de recursos humanos e que apresenta especificidades e dificuldades comuns a outros organismos públicos e que formam um importante empregador na sociedade onde se inserem. Adicionalmente, esta capacidade de generalização da metodologia foi testada com recurso a outras instituições no âmbito da Defesa.

Considerando a necessidade real e prática da utilização deste tipo de metodologia na instituição militar, o presente capítulo visa demonstrar a aplicação da metodologia proposta, testar a sua replicabilidade e simultaneamente contribuir para a avaliação e intervenção na reputação do Exército. Para isso, encontra-se organizado segundo as fases da metodologia proposta, iniciando-se com a Secção referente às características e necessidades do Exército e das Forças Armadas. Com base nesse conhecimento, a Secção seguinte descreve detalhadamente a identificação das fontes de informação, as quais incluem fontes do Exército, mas também da Marinha, Força Aérea e Forças Armadas para replicação da análise nessas instituições e na organização global. A partir das fontes selecionadas, o capítulo prossegue para a Secção de recolha da informação, onde se expõe o processo de extração desenvolvido e a exploração inicial da informação extraída. Posteriormente, as Secções de extração do conhecimento e de avaliação contêm a análise de sentimento para o caso de estudo e casos de validação, mas também o estudo da potencialidade para treinar modelos de classificação automática e de tópicos. Por último, o capítulo encerra com a Secção de recomendações, onde se descrevem sugestões práticas para a gestão e monitorização do eWOM no Exército e nas Forças Armadas.

## **4.1 Características e necessidades da instituição**

Enquanto instituição pública, o Exército insere-se no setor da Defesa. Em Portugal, a defesa nacional visa « (...) garantir a soberania do Estado, a independência nacional e a integridade territorial de Portugal, bem como assegurar a liberdade e a segurança das populações e a proteção dos valores fundamentais da ordem constitucional (...)» e ainda,

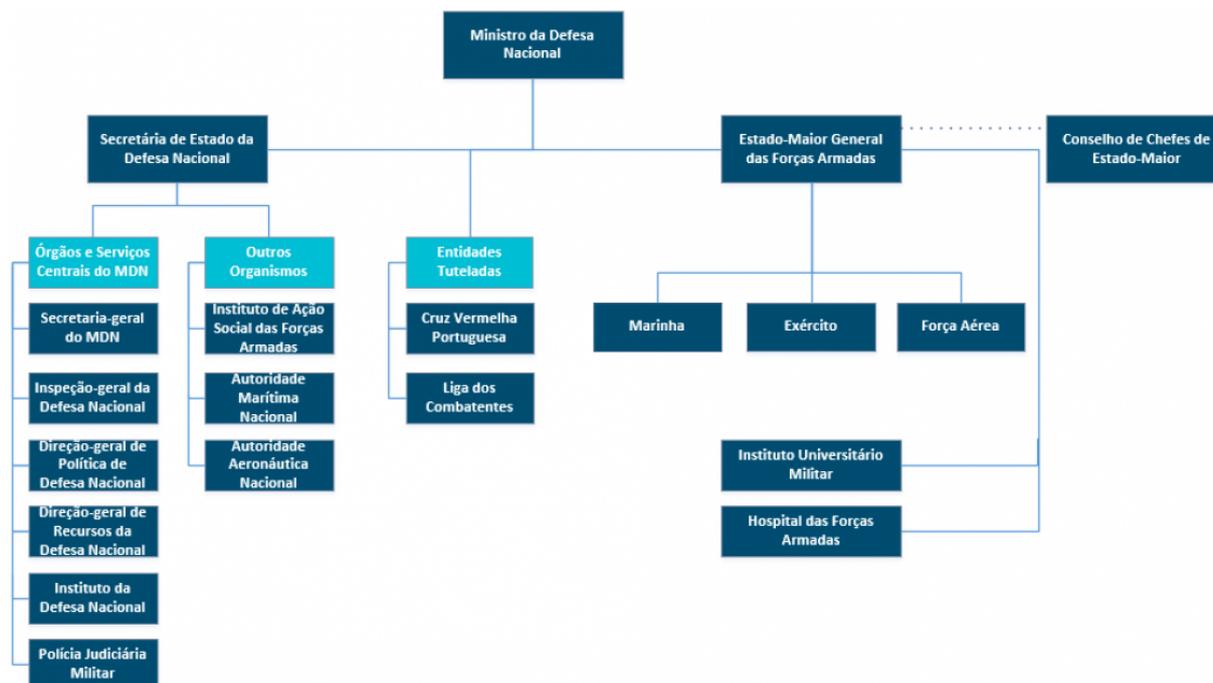


Figura 4.1: Organização do Ministério da Defesa Nacional

cumprir os «(...) compromissos internacionais do Estado no domínio militar (...)» (Lei orgânica n.º5/2014). Esta capacidade é constituída por órgãos do Estado responsáveis pela Defesa e pelas Forças Armadas como demonstrado na Figura 4.1<sup>1</sup>.

As Forças Armadas são responsáveis pelo aprontamento e emprego operacional de forças destinadas à realização de missões no âmbito da defesa nacional. Para isso, a sua estrutura inclui 3 ramos – Marinha, Exército e Força Aérea – que possuem legislação e missões próprias, mas que se integram no sistema de forças da defesa militar da República (Lei Orgânica n.º 6/2014).

Desde 2004, as Forças Armadas abandonaram o serviço militar obrigatório e passaram a depender somente de militares voluntários. Com esta mudança, os seus ramos passaram a competir com outros empregadores, criando a necessidade de um esforço institucional para se diferenciarem face ao mercado de trabalho e conseguirem atrair os recursos humanos capazes de manter a sua capacidade operacional (Reis et al., 2019). Para responder à necessidade criada, o Ministério da Defesa Nacional (MDN) e os ramos têm implementado diversas medidas que visam informar, atrair e recrutar jovens da sociedade civil. No mesmo sentido, orientaram esforços na divulgação através redes sociais e outros canais de comunicação. Apesar disso, e embora o funcionamento das Forças Armadas dependa grandemente do efetivo militar em regime de voluntariado e contrato<sup>2</sup>, tem-se registado uma diminuição

<sup>1</sup>Extraído em janeiro de 2020 de <https://recrutamentomilitar.bud.gov.pt/missao-e-organizacao-do-ministerio-da-defesa-nacional>

<sup>2</sup>Os regimes de contrato e voluntariado referem-se à prestação de serviço militar por um período de tempo limitado. Outros regimes incluem o serviço efectivo nos quadros permanentes e o serviço efectivo decorrente de convocação/mobilização (Lei n.º 174/99 - Lei do Serviço Militar).

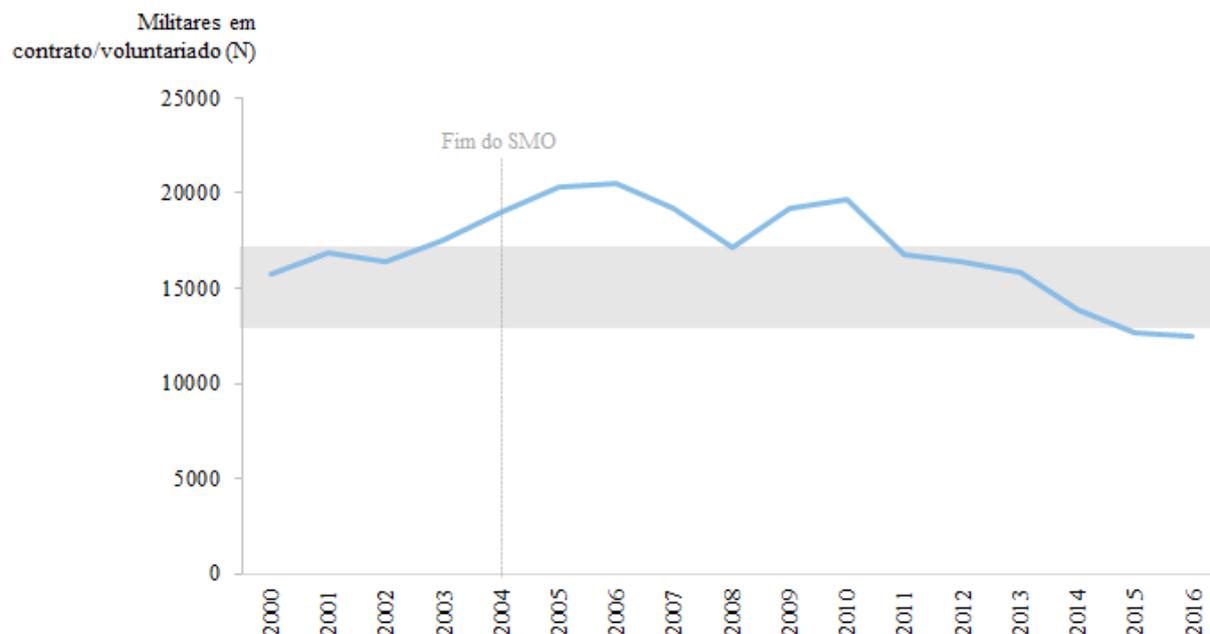


Figura 4.2: Análise temporal do efetivo em regime de contrato/ voluntariado

anual destes profissionais nos três ramos (Reis et al., 2019). A título de exemplo, a Figura 4.2 consiste numa adaptação dos dados do Ministério da Defesa Nacional (citados pelo jornal Diário de Notícias <sup>3</sup>) e permite visualizar uma tendência decrescente do efetivo militar existente, em particular desde 2015, quando o número de militares atingiu níveis abaixo do desvio-padrão.

Com este agravamento das dificuldades de atração e retenção, as instituições das Forças Armadas não conseguem preencher o número de vagas autorizadas e apresentam uma escassez de recursos humanos que se traduz na falta de pessoal necessário para o cumprimento adequado das missões<sup>4</sup>. Para inverter essa tendência, têm sido desenvolvidas medidas de divulgação e de recrutamento militar ao nível do Ministério da Defesa Nacional e dos respetivos ramos. No que se refere ao primeiro, destacam-se o Dia da Defesa Nacional, a plataforma de recrutamento da Direção-Geral de Recursos da Defesa Nacional, a promoção de um programa de profissionalização e a aprovação de um programa de incentivos ao serviço militar. Relativamente aos ramos das Forças Armadas, têm-se verificado atualizações nos sistemas de divulgação e recrutamento, alargamento do público-alvo da divulgação e a utilização de canais de comunicação diversos (publicidade física e digital, *website* institucional, redes sociais, centros de recrutamento e eventos).

Embora a fraca atratividade do serviço militar seja transversal aos três ramos, o Exército Português é aquele que requer uma maior porção de jovens militares e tem revelado

<sup>3</sup>Extraído em dezembro de 2019 de <https://www.dn.pt/portugal/forcas-armadasperderam-25-dos-efetivos-numa-decada-9024946.html>

<sup>4</sup>Extraído em dezembro de 2019 de <https://tvi24.iol.pt/sociedade/governo/saida-de-militares-nas-forcas-armadas-esta-a-provocar-sobrecargas-de-esforco-aos-que-ficam>

fraca capacidade de recrutamento, em particular, para militares da categoria de Praças<sup>5</sup>. A situação apresenta uma tendência negativa que atingiu os valores mais baixos em 2018, com menos de um candidato por cada vaga aberta nesse ano (Silva et al., 2019). No caso concreto do Exército, os esforços de recrutamento, retenção e gestão de recursos humanos são responsabilidade da Direção de Administração de Recursos Humanos (DARH) que acumula com uma diversidade e quantidade bastante significativa de necessidades institucionais.

Para avaliar as fontes de recrutamento mais eficazes, o Exército Português questionou os jovens incorporados acerca do meio a partir do qual obtiveram informação de ingresso. Os resultados indicaram que os “familiares ou amigos”, “familiares ou amigos militares” e as “redes sociais” são – por esta ordem – os principais meios que atraíram os novos militares (Silva et al., 2019). Estes dados são similares às principais fontes de confiança encontradas em consumidores, uma vez que as fontes externas à organização são percecionadas como parte desinteressada no processo de decisão e podem refletir opiniões mais transparentes (Reis et al., 2019). No mesmo sentido, são evidências semelhantes ao descrito na literatura sobre recrutamento militar, onde se tem detetado uma influência social sobre os processos cognitivos de perceção e tomada de decisão acerca do empregador escolhido (Hoye and Lievens, 2007).

A partir da análise das necessidades e características do Exército, conclui-se que a reputação da instituição junto dos cidadãos e dos militares pode ser fundamental para preencher as fileiras e garantir o cumprimento da sua missão. Este problema assume uma importância fundamental e encontra-se presente em dois dos seus objetivos estratégicos (OE): «OE1: Assegurar Recursos Estáveis e Suficientes» e «OE4: Dinamizar a Comunicação e a Imagem Interna e Externa»<sup>6</sup>. Para atingir estes objetivos, têm sido realizados esforços na divulgação digital, porém, não existe um sistema de avaliação da opinião pública (ou da eficácia desses esforços) ou diretrizes institucionais específicas para a gestão dessa reputação online. Assim, de acordo com as necessidades existentes e a importância da imagem neste tipo de problema, considera-se indicada a aplicação da metodologia proposta, com ênfase nos requisitos de gestão reputacional centrada numa imagem de utilidade pública e na marca de empregador positiva.

## 4.2 Identificação de potenciais fontes de informação

Após a recolha de informação institucional e de acordo com a metodologia proposta, esta fase teve como objetivo identificar fontes de informação em *social media*, com conteúdo suficiente para a aplicação de técnicas de *text mining*, mas capazes de refletir a opinião alar-

<sup>5</sup>A categoria de Praças encontra-se na base da hierarquia militar e refere-se a uma população-alvo com idade entre 18 e 24 anos, com 12<sup>o</sup> ano de escolaridade, aptidão física e psíquica adequada e vontade de prestar serviço militar entre 2 a 6 anos.

<sup>6</sup>Extraído em maio de 2020 de <https://www.exercito.pt/>

gada dos cibernautas acerca do Exército. Para isso, é descrito o primeiro passo referente à definição da população-alvo e que permitiu aceder a um conjunto alargado de utilizadores com consequente abrangência das aplicações disponíveis. Seguidamente, expõe-se o processo de definição dos termos de pesquisa, onde se optaram por termos diretos e genéricos (e.g. Exército) pela eficácia no acesso a páginas e aplicações com temas abrangentes e não sectorizados. Por último, foi possível obter um grupo limitado, mas diversificado, de fontes «candidatas» cuja potencialidade para o estudo dependeu de um processo de seleção com critérios de exclusão e inclusão. Embora subjetivos, estes critérios permitiram distinguir a pertinência e adequação das várias aplicações e páginas, tornando possível escolher um conjunto de fontes capaz de aceder a diferentes tipos de utilizadores, temas variados e com qualidade necessária às técnicas de *text mining*.

Conclui-se que existem fontes de informação com conteúdo relevante para um estudo deste tipo e considera-se que o objetivo foi alcançado com um conjunto de 19 fontes de informação pertencentes a diferentes categorias de *social media*, nomeadamente, redes sociais, comunidades ou fóruns. Estas fontes podem ser agrupadas em fontes oficiais (geridas pelas instituições) e fontes independentes (sem o controlo institucional).

#### 4.2.1 Público-alvo

A definição do público-alvo pretendia analisar se o eWOM pode contribuir para o comportamento da população deficitária na instituição militar e delinear características sócio-demográficas dos utilizadores com interesse para a reputação institucional online.

Relativamente ao pressuposto de contribuição do eWOM, embora não tenha sido encontrada literatura relevante no Exército Português, considera-se que a sua importância no contexto organizacional, aliada à tendência para jovens e adultos confiarem na sua rede de amigos e em opiniões publicadas online (Reis et al., 2019; Silva et al., 2019), sugere que a população-alvo recrutável e os seus referenciadores podem ser influenciados pelas opiniões dos cibernautas. Estima-se que esta relação entre o eWOM e a tomada de decisão seja particularmente forte devido às características das plataformas: a título de exemplo, as páginas das redes sociais assemelham-se às redes sociais reais, incluindo membros de confiança dos utilizadores que podem fortalecer a confiança nas mensagens partilhadas e favorecer a influência exercida (Chu and Kim, 2011).

No que se refere às características que permitam filtrar a opinião dos utilizadores mais relevantes, a análise das características institucionais indicou que além da população jovem, a opinião de referenciadores (obtidas através de amigos, pais e redes sociais) abrange uma população com características diversificadas e difíceis de delinear. Assim, optou-se pela análise alargada de comentários e sem restrições nos utilizadores objeto de análise.

### 4.2.2 Termos de pesquisa

O processo de definição dos termos de pesquisa a utilizar foi indissociável do conhecimento institucional. A partir desse conhecimento, foi desenvolvido um conjunto de termos para uma pesquisa exaustiva das potenciais fontes de informação disponíveis *online*. A lista de termos inicial permitiu recolher um conjunto alargado expressões específicas da instituição para pesquisa em diversas plataformas (a título de exemplo: «tropa», «portuguese army», «treino operacional», «serviço militar», entre outros). Após a análise inicial, foram eliminados termos de pesquisa que não devolviam resultados de acordo com os critérios definidos (ver Secção 4.2.3) e escolhidos os termos que se caracterizavam pelo elevado número de resultados, relação com o caso em estudo e tema suficientemente genérico para obter dados relacionados com a reputação das Forças Armadas.

De acordo com a terminologia e organização institucional avaliados na Secção 4.1, a lista de termos que melhor permitiram acesso às fontes de informação é a seguinte:

- Exército;
- Exército Português;
- Marinha;
- Marinha Portuguesa;
- Armada;
- Força Aérea;
- Força Aérea Portuguesa;
- Forças Armadas;
- Forças Armadas Portuguesas;
- Ministério da Defesa.

Importa referir que, apesar do caso em estudo ser referente ao Exército, a recolha de um número alargado de temas permitiu obter um maior volume de informação útil para as técnicas posteriores e efetuar análises comparativas entre instituições da Defesa.

### 4.2.3 Seleção das fontes de informação

A seleção das fontes de informação teve como objetivo escolher as fontes com maior quantidade de informação não-estruturada e conteúdo diversificado, mas relevante para o tema em análise. Para o caso de estudo, observou-se um conjunto restrito de potenciais fontes de informação. No entanto, definir os parâmetros que distinguem a sua relevância revelou

ser um processo difícil, devido à diversidade de aplicações de *social media*, da dispersão de páginas/canais/contas existentes e das suas características específicas.

Para adotar um processo sistemático de seleção, foram definidos um conjunto de critérios de exclusão e inclusão. Relativamente aos fatores de exclusão que garantiram os pressupostos da metodologia, foram excluídas as fontes que apresentavam:

- Baixa interação dos utilizadores (e.g. página “recrutamentomilitar.bud.gov.pt”);
- Secção de comentários desativada (e.g. o blog “Operacional”);
- Elevada especificidade da temática ou população-alvo (e.g. página de Facebook “BrigadaMecanizada”).

No que se refere aos critérios de inclusão, a representatividade e o volume de informação requeriam a inclusão de fontes que cumpriam os requisitos:

- Número de publicações igual ou superior a 3 para, pelo menos, um dos termos em estudo;
- Número total de comentários igual ou superior a 50.

A seleção de fontes e a definição de critérios constituiu um processo iterativo entre os termos de pesquisa e a exploração das fontes encontradas. Inicialmente, foi obtido um conjunto alargado aplicações de *social media* frequentemente utilizadas pela população portuguesa e as respetivas páginas oficiais das Forças Armadas. No entanto, a filtragem baseada nos critérios definidos resultou no seguinte conjunto de fontes institucionais:

1. **Exército:** @ExercitoRecrutamento (*Facebook*); @ExercitoPortuguesPRT (*Facebook*); exercitoportugues\_oficial (*Instagram*);
2. **Marinha:** @MarinhaPortuguesa (*Facebook*);
3. **Força Aérea:** @PortugueseAirForce (*Facebook*);
4. **Ministério da Defesa Nacional:** @defesa.pt (*Facebook*);
5. **Estado-Maior-General das Forças Armadas:** @emgfa (*Facebook*).

Destas fontes selecionadas, o Instagram demonstrou publicações e comentários maioritariamente gráficos, tendo sido considerada a fonte menos pertinente face às restantes. No entanto, após a fase de recolha de informação (Secção 4.3), que incluiu a extração e análise preliminar dos dados, foram detetadas um conjunto de limitações significativas (ver Secção 4.3.2) que motivaram o refinamento do projeto até cumprir os requisitos necessários para as fases seguintes. Essas limitações foram minimizadas através do alargamento da pesquisa a fontes de informação não-institucional, nomeadamente, fóruns, microblogues e

comunidades virtuais que revelaram menos restrições no tipo e valência dos comentários, mas também, um maior número de caracteres por texto. Além destas vantagens, a diversidade de fontes possibilitou o acesso a diferentes públicos, sem influência ou relação com a instituição. Após a filtragem realizada, as fontes não-institucionais selecionadas foram:

1. **Fórum defesa:** encontrado em [www.forumdefesa.com](http://www.forumdefesa.com), é definido como um espaço para discussão de temas sobre a Defesa portuguesa e, segundo as estatísticas da página, possui um tópico “Ingresso no Exército” no Top 10 em número de visualizações e respostas;
2. **Páginas de Facebook dos canais de notícias:** @ DiariodeNoticias.pt, @Publico, @jornalexpresso, @rtpnoticias, @cmjornal, @jornalnegocios, @tvi24, @tsfradio, @sicnoticias, @ObservadorOnTime;
3. **Reddit:** acedido em <https://www.reddit.com/r/portugal/>, é uma plataforma que reúne comunidades de utilizadores organizadas por áreas de interesse.

O processo de seleção deu-se por concluído quando obteve 19 potenciais fontes de informação com suficiente diversidade nos comentários, publicações, utilizadores e polaridade de sentimentos. Globalmente, estas fontes continham também um número significativo de opiniões partilhadas publicamente e que sugeria adequação às técnicas de análise posteriores.

### 4.3 Recolha da informação

Esta secção teve como objetivo desenhar um método de extração capaz de recolher o conteúdo não-estruturado das fontes de informação e confirmar a adequabilidade do material extraído. Para extrair o eWOM das fontes selecionadas e avaliar a sua qualidade, realizaram-se os passos de seleção dos atributos, desenvolvimento do processo de *web scraping*, definição de critérios para filtragem do conteúdo relevante e descrição da exploração inicial do conteúdo extraído. No final desta fase foi possível obter um sistema semi-automatizado que extraiu 13 358 comentários e 708 publicações, as quais revelaram limitações e potencialidades para o estudo. Globalmente, conclui-se que as fontes não-institucionais são o palco da maioria das informações partilhadas publicamente, registando também mais interação entre utilizadores e um conteúdo mais elaborado, o que sugere, uma maior adequação para as técnicas de *text mining* e potencial para obter informação abrangente.

#### 4.3.1 Extração da informação

Relativamente à extração de informação, optou-se pelo método de *web scraping*, que permite o fácil acesso aos dados e a adaptação a diferentes plataformas. Este processo para

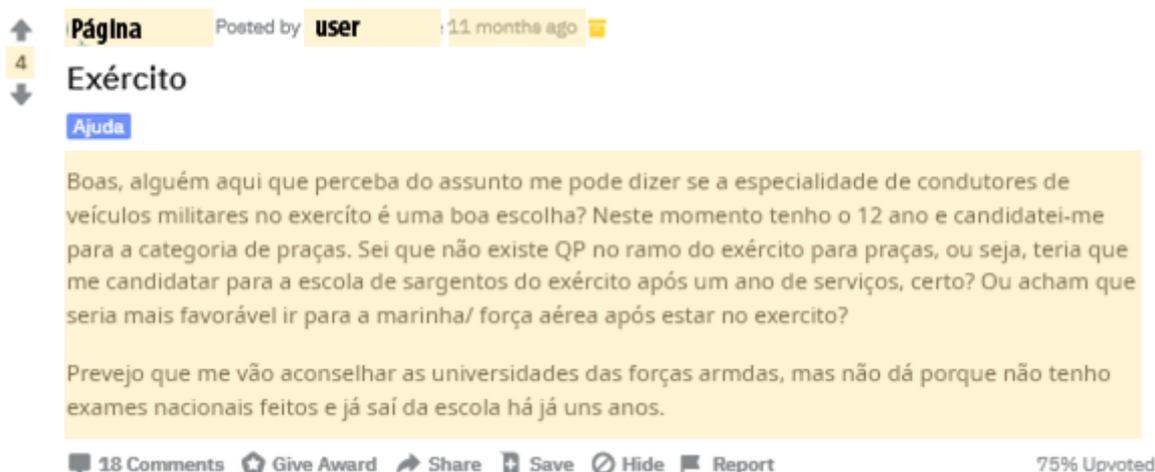


Figura 4.3: Exemplo de atributos extraídos de uma publicação

extração de informação presente em páginas *web* foi realizado através de um automatizador para *web browser*, programado em *Python* e personalizado para cada fonte de informação selecionada. Relativamente ao desenvolvimento desta extração automatizada, importa salientar as principais especificidades das fontes de informação e que exigiram um conjunto de decisões relacionadas, que se encontram descritas de acordo com a ordem de desenvolvimento:

1. Seleção dos atributos a recolher de acordo com sugestões encontradas na literatura e conhecimento da instituição:
  - A informação recolhida distingue-se em atributos referentes às Publicações e referentes aos Comentários, como exemplificado respetivamente nas Figuras 4.3 e 4.4 <sup>7</sup> (os atributos recolhidos encontram-se a sombreado; as identificações de página e de utilizador foram anonimizadas).
  - Para possibilitar a integração da informação extraída de diferentes fontes, foram selecionados e nomeados atributos semelhantes. Assim, as Tabelas 4.1 e 4.2, demonstram os atributos e nomenclatura usada durante o processo de extração da informação. Globalmente, foi extraído o conteúdo do texto que permite analisar as opiniões e sentimentos, as datas para análises temporais, a identificação do utilizador para a sua contabilização e ainda, reações ou pontuações obtidas.
2. Garantia da integridade dos registos:
  - Nome do atributo: Cada fonte de informação atribui diferentes nomes aos atributos que classificam a informação. Durante o processo de extração, foi atribuída uma nomenclatura idêntica para os atributos das diferentes fontes de informação. A título de exemplo, a informação referente a «Comments» (Reddit),

<sup>7</sup>Extraídas em janeiro de 2020 de <https://www.reddit.com/r/portugal/>

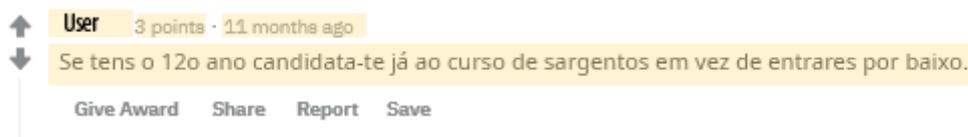


Figura 4.4: Exemplo de atributos extraídos de um comentário

Nome	Descrição	Exemplo
Texto	Texto referente ao comentário.	«O que tens feito ultimamente? algo com ação e adrenalina?»
User	Nome de utilizador referente ao autor do comentário.	Quillergs
Data	Data referente ao comentário (quando aplicável)	2 years ago
ID	Número identificador da publicação que deu origem ao comentário.	10
Fonte	Categoria identificadora da fonte de dados e temática correspondente	Reddit - Exército
Nível	Categoria referente ao nível do comentário (quando aplicável)	1
Pontuação	Número correspondente à votação que os utilizadores fazem do comentário (quando aplicável).	3

Tabela 4.1: Atributos extraídos dos comentários

Nome	Descrição	Exemplo
ID	Número inteiro identificador da publicação.	10
Texto	Texto referente à publicação, notícia ou tópico.	«Boas camaradas! Neste dia 10 de Junho, dia de Portugal, nada melhor para um recém Soldado do Exército Português, do que estar de Serviço 24h»
Pontuação	Número de reações ou votações que os utilizadores fazem da publicação.	46
Data	Data de publicação (quando aplicável)	2 years ago
Fonte	Categoria identificadora da fonte de dados e temática correspondente	Reddit - Exército
User	Nome de utilizador referente ao autor da publicação	u/born2roll

Tabela 4.2: Atributos extraídos das publicações

«Respostas» (Fórum Defesa) e «Comentário» (Facebook) pertenceram ao atributo «Texto». A criação de atributos semelhantes, permitiu integrar facilmente os dados para tratamento e análise posteriores.

- Publicação de origem: Para cada publicação foi atribuído um número de identificação. Durante a extração, cada comentário foi associado ao número da respetiva publicação. Este procedimento garantiu a possibilidade de selecionar comentários a publicações específicas ou analisar diferenças entre publicações.
- Hierarquia de comentários: Algumas fontes de informação permitem comentários principais (resposta a publicações) e secundários (resposta a comentários). Considerando a diferença temática entre categorias de comentários da mesma publicação, foi criado um atributo binário (0-principal e 1-secundário) que possibilita a filtragem em análise posterior.
- Associação entre tipos de informação: Durante o desenvolvimento do processo de extração, foi garantida a associação entre a informação e as respetivas fontes de informação e termos de pesquisa. Nesse sentido, durante a extração, foi associado a cada registo o atributo «Fonte» e descrita sua fonte de informação e respetivo termos (ver exemplo na Tabela 4.1).

### 3. Datas para recolha da informação:

- O processo de extração foi aplicado a uma janela temporal definida. Inicialmente, foi extraída informação partilhada de janeiro e a novembro 2019. No entanto, a quantidade de dados obtida revelou ser insuficiente para a análise posterior. Esta dificuldade foi particularmente problemática nas fontes de informação não institucionais, nas quais as publicações e comentários referentes aos termos de pesquisa são menos frequentes. No que se refere às fontes de informação institucional, foi aplicado um processo de extração em 2018 mas que revelou pouco conteúdo não-estruturado e um padrão muito semelhante ao obtido em 2019, tendo sido o único realizado para esse ano. Após a observação dos dados, optou-se por uma janela temporal entre 2014 e 2019 para fontes não-institucionais e de 2018 a 2019 para as fontes institucionais.

### 4. Definição de critérios de filtragem das publicações e respectivos comentários, de acordo com os requisitos delineados e para garantia da extração de conteúdo relevante (diminuição do erro). No caso em estudo, foram excluídas:

- Publicações que desencadeiam comentários repetitivos (e.g. “Parabéns”);
- Publicações que desencadeiam ações dos utilizadores (e.g. identificar amigos);
- Publicações referentes a concursos;
- Publicações sem comentários associados;
- Publicações geradoras de comentários essencialmente de político-partidários.

#### 5. Método de extração:

- Após a seleção das fontes de informação e dos atributos a recolher descritos nas secções anteriores, foi desenvolvida e aplicada a extração da informação com recurso à técnica de *web scraping*. Esta técnica de extração de informação disponível em *websites* foi realizada através de um automatizador gratuito, programado para iniciar, clicar, procurar e recolher o conteúdo HTML selecionado. O processo de *web scraping* foi realizado para cada fonte de informação (*social media*) e por página selecionada (e.g. página *@ExercitoRecrutamento* no Facebook), com posterior armazenamento da informação em ficheiro *.xls*.

A extração da informação ficou concluída quando os atributos foram filtrados e extraídos de acordo com os cuidados, critérios e período de tempo definido anteriormente. Importa referir que a diversidade de fontes, tipo de conteúdo e critérios de inclusão/exclusão tornam o processo complexo. Assim, a automatização total não foi alcançada, uma vez que requereu a intervenção humana para validação.

#### 4.3.2 Exploração da informação

A fase de exploração da informação é descrita enquanto processo iterativo e dependente da extração realizada. A intervenção humana referida anteriormente ocorre principalmente nesta fase, na qual é realizada a exploração do conteúdo extraído para conhecer as características do material e avaliar as decisões tomadas na fase de recolha da informação (Secção 4.3). A exploração inicial da informação obtida a partir de fontes institucionais revelou que as publicações e comentários existentes apresentavam um conjunto de limitações relevantes para a análise de dados posterior, especificamente:

1. O conteúdo apresentado nas fontes de informação referentes a páginas oficiais da instituição era gerido pelos administradores e podia não ser representativo da reputação pública (a título de exemplo, foi detetada a remoção de comentários negativos);
2. A rede social Facebook possuía, na generalidade, comentários com poucos caracteres, tornando este conteúdo insuficiente para a aplicação dos algoritmos de *text mining* (ver exemplo na Figura 4.5<sup>8</sup>);
3. Como exemplificado na Figura 4.6<sup>9</sup>, o Instagram possuía publicações essencialmente gráficas e comentários constituídos essencialmente por *emojis* e identificação de utilizadores;
4. O conteúdo dos comentários nas redes sociais incluía frequentemente a identificação de utilizadores, como se pode observar na Figura 4.7<sup>10</sup>;

<sup>8</sup>Extraído em janeiro de 2020 de <https://www.facebook.com/pg/ExercitoRecrutamento/>

<sup>9</sup>Extraído em janeiro de 2020 de [https://www.instagram.com/exercitoportugues\\_oficial/](https://www.instagram.com/exercitoportugues_oficial/)

<sup>10</sup>Extraído em janeiro de 2020 de <https://www.facebook.com/pg/ExercitoRecrutamento/>

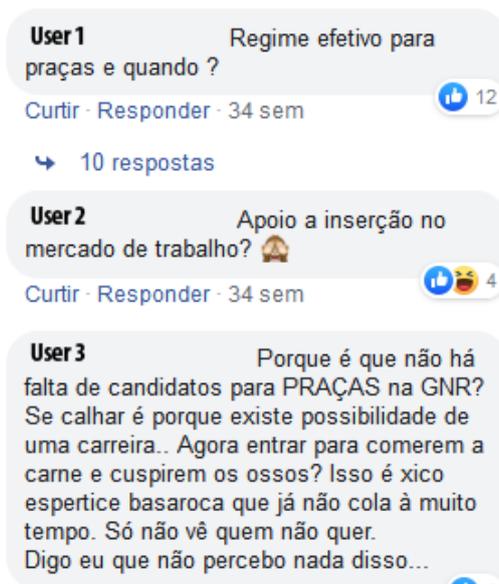


Figura 4.5: Exemplo de comentários com poucos caracteres na rede social Facebook

5. A aplicação de estratégias para estimular a interação dos utilizadores com a página, revelou um elevado número de comentários específicos e pouco relevantes para o estudo da imagem institucional. A Figura 4.8<sup>11</sup> exemplifica a resposta a uma publicação que apelava à identificação de utilizadores para realizar exercício físico;
6. Observou-se que algumas páginas apresentavam conteúdo com elevada especificidade militar, com consequentes comentários situacionais e comentadas somente por um conjunto de utilizadores restrito (maioritariamente militares e familiares). A Figura 4.9<sup>12</sup>, exemplifica um conjunto de comentários referentes às missões internacionais realizados por utilizadores com ligação à instituição;
7. Para analisar a distribuição da polaridade de sentimento nos comentários recolhidos, foi realizada a tradução e análise de polaridade com recurso à biblioteca TextBlob. Esta exploração identificou sentimentos com valência negativa na maioria dos comentários extraídos e sem número suficiente de comentários positivos e neutros para a análise de sentimento posterior.

Para ultrapassar as limitações das redes sociais Facebook e Instagram, foi necessário alargar os termos de pesquisa e diversificar o conjunto de fontes de informação definidos nas Secções 4.2.2 e 4.2.3. Após as modificações necessárias, foram adicionadas páginas não institucionais e fontes de informação de diferentes plataformas, permitindo garantir os pressupostos de qualidade da informação necessários para o processo de Extração do Conhecimento.

<sup>11</sup>Extraído em janeiro de 2020 de <https://www.facebook.com/pg/ExercitoRecrutamento/>

<sup>12</sup>Extraído em janeiro de 2020 de <https://www.facebook.com/emgfa/>



Figura 4.6: Exemplo de comentários na rede social Instagram



Figura 4.7: Exemplo de comentários com identificação de utilizadores na rede social Facebook

Globalmente, as características da informação analisada nesta Secção sugerem que as fontes de informação institucionais, em particular nas redes sociais Facebook e Instagram, poderão apresentar uma quantidade e diversidade de conteúdo insuficientes para a análise da reputação institucional.

### 4.3.3 Descrição dos resultados

Esta secção tem como objetivo de caracterizar a informação extraída, analisar indicadores relevantes para a gestão do eWOM e avaliar o cumprimento dos pressupostos necessários para a análise subsequente. Para isso, começa por descrever e caracterizar os principais resultados referentes a comentários e publicações para cada fonte de informação. Seguidamente, são expostas as diferenças observadas por categoria da fonte de informação (institucional vs não institucional) e apresentadas tendências temporais na utilização das plataformas de *social media* utilizadas.

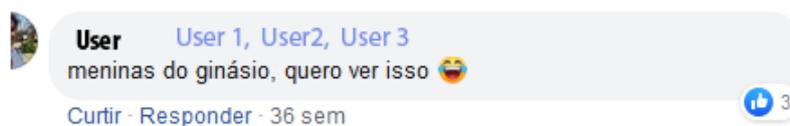


Figura 4.8: Exemplo de comentário de estimulação à interação na rede social Facebook: «Marca os teus amigos: o último a responder-te tem de fazer 50 flexões»

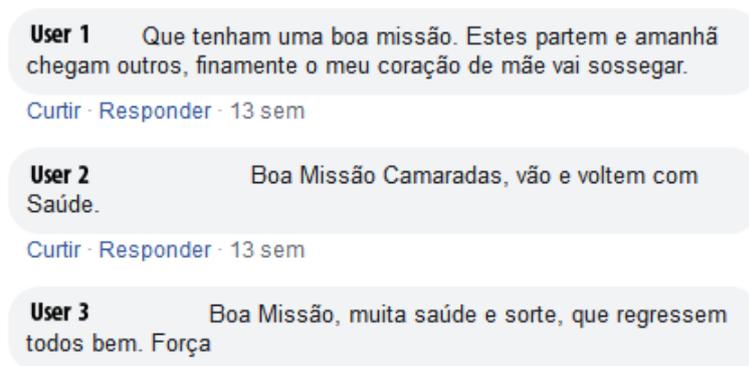


Figura 4.9: Exemplo de comentários com elevada especificidade na rede social Facebook

A análise que se segue distingue Comentários e Publicações para cada fonte de informação<sup>13</sup> e inicia-se com a descrição global dos resultados na Tabela 4.3. Como se pode observar, foi possível obter um total de 13 358 comentários com uma prevalência de comentários obtidos através de fontes oficiais (N=7850) face às fontes independentes (N=5508). A interação social acerca das instituições militares avaliada segundo estes valores revelou um número elevado de comentários nas páginas institucionais dos ramos, porém, as fontes independentes Reddit e Facebook dos jornais Público e Correio da Manhã também revelaram um número significativo de interações, confirmando a existência de opiniões relevantes em fontes de informação diversas.

Além dos comentários realizados pelos utilizadores, os dados associados à publicações que lhes dão origem são considerados um importante indicador acerca da interação entre a página e os seus seguidores (Sterne, 2010). Nesse sentido, o processo de *web scraping* permitiu extrair 706 publicações de todas as fontes de informação, com exceção do Fórum Defesa que não possui esta modalidade. Na Tabela 4.3, os dados discriminados por fonte, demonstram um padrão bastante semelhante ao detetado para os comentários. Porém, nas fontes independentes, os utilizadores do Reddit foram aqueles que mais publicações geraram acerca dos temas em análise (N=48), denotando o interesse dessa comunidade.

Seguidamente, procedeu-se ao cálculo do número médio de palavras para estudar a adequação da informação para análise do eWOM. Na Tabela 4.3, verifica-se que apesar do número de comentários extraídos, as fontes provenientes das redes sociais Facebook

<sup>13</sup>Para efeitos de visualização, as fontes de informação relativas a páginas do Facebook denominam-se «FB»

<sup>14</sup>A fonte de informação Fórum Defesa não possui publicações.

Fonte	Nº de Publicações	Palavras por publicação		Nº de Comentários	Palavras por comentário	
		M	DP		M	DP
<u>Fontes institucionais</u>						
FB Exército Português	37	<b>97</b>	153	652	15	17
FB Recrutamento Exército	<b>298</b>	48	46	<b>4571</b>	14	19
FB Marinha Portuguesa	27	79	67	666	13	15
FB Força Aérea	35	26	18	714	13	15
FB Ministério da Defesa	35	<b>144</b>	187	226	<b>23</b>	<b>27</b>
FB do EMGFA	43	<b>97</b>	51	485	17	20
Instagram Exército	<b>65</b>	48	39	536	10	11
<u>Fontes não institucionais</u>						
Forum Defesa	NA <sup>14</sup>	NA	NA	716	<b>53</b>	<b>60</b>
FB Correio da manhã	19	15	5	660	19	24
FB Diário Notícias	26	15	6	564	19	25
FB Expresso	11	32	24	307	20	23
FB Jornal de negócios	3	10	2	51	15	13
FB Observador	12	13	3	427	16	18
FB Público	20	20	7	<b>995</b>	19	24
FB Rtp Noticias	8	18	15	142	17	18
FB Sic Notícias	8	11	7	346	21	23
FB TSF	5	11	1	165	15	14
FB Tvi24	6	18	6	232	21	19
Reddit	<b>48</b>	42	47	<b>903</b>	<b>37</b>	<b>51</b>

Tabela 4.3: Descritivas da informação extraída por fonte, comentários e publicações

e Instagram, tendem a apresentar comentários curtos e com léxico insuficiente para análise. No que se refere aos comentários, as fontes independentes com uma tipologia de fórum/comunidade foram aquelas que revelaram um conteúdo textual mais rico, com valores de média (M) e desvio-padrão (DP) bastante superiores aos restantes. Relativamente às publicações, o conteúdo escrito é superior nas fontes oficiais das instituições, estando normalmente associado a conteúdo informativo, por oposição às páginas noticiosas que publicam a hiperligação do artigo relacionado.

Globalmente, a caracterização inicial dos dados confirma a dificuldade em obter uma grande quantidade de informação no setor público, apresentando um número satisfatório de comentários, mas limitado pelo número de palavras geralmente baixo e que pode dificultar o desempenho das técnicas de extração do conhecimento. Apesar disso, salienta-se que a comunidade virtual Reddit foi a fonte que apresentou os melhores resultados (a negrito) para a maioria das áreas avaliadas e aquela que mais se destacou nas fontes independentes.

Relativamente à janela temporal dos dados extraídos, a Tabela 4.4 revelou que a maioria dos comentários extraídos se referem ao ano 2019 (de janeiro a novembro). Os resultados deste último ano decorrem da prioridade atribuída à extração de informação recente, em particular nas redes sociais institucionais. Apesar disso, 2018 foi o ano que registou um

<b>Fonte</b>	<b>2014</b>	<b>2015</b>	<b>2016</b>	<b>2017</b>	<b>2018</b>	<b>2019</b>
<u>Fontes institucionais</u>						
FB Exército Português						652
FB Recrutamento Exército					883	3688
FB Marinha Portuguesa						666
FB Força Aérea						714
FB Ministério da Defesa						226
FB do EMGFA						485
Instagram Exército						536
<b>Total</b>					<b>883</b>	<b>6967</b>
<u>Fontes não-institucionais</u>						
Forum Defesa	321	102	89	149	32	23
FB Correio da manhã		30	43	50	492	75
FB Diário Notícias			55	22	257	200
FB Expresso			16	67	160	64
FB Jornal de negócios				15		36
FB Observador				48	102	277
FB Público			61	157	522	255
FB Rtp Notícias				21	25	96
FB Sic Notícias				84	95	167
FB TSF					119	46
FB Tvi24				82	127	23
Reddit	35		56	331	303	178
<b>Total</b>	<b>356</b>	<b>132</b>	<b>320</b>	<b>1026</b>	<b>2234</b>	<b>1440</b>
<b>Total global</b>	<b>356</b>	<b>132</b>	<b>320</b>	<b>1026</b>	<b>3117</b>	<b>8407</b>

Tabela 4.4: Comentários extraídos por fonte e ano

maior número de opiniões partilhadas nas fontes independentes, totalizando 2234 registos. Globalmente, a análise temporal sugere a importância de extrair vários anos de informação publicada, uma vez que as aplicações de *social media* não apresentaram partilhas anuais suficientes para monitorizar o eWOM global.

Além da resposta dos utilizadores, importa analisar as publicações que lhes dão origem e que refletem a gestão das redes sociais. Na Figura 4.10<sup>15</sup> observou-se que as fontes institucionais tendem a exibir um número elevado de publicações mensais quando comparadas com as restantes. Estes dados são congruentes com a importância atribuída à divulgação e gestão direcionada para a instituição, mas podem denotar uma tendência para a estimulação de tráfego na página, sem conteúdo gerador de diálogo e envolvimento dos utilizadores.

A extração de dados pertencentes às publicações permitiu também a obtenção de reações ou votações associadas. No que se refere a esta contabilização, a comparação na Figura 4.11 demonstrou que as fontes institucionais, em particular a conta de Instagram do Exército, são aquelas que contabilizam um maior número de reações. Atualmente, este indicador não é considerado fundamental para o marketing eletrónico (Peters et al., 2013),

<sup>15</sup>Esta análise não inclui a fonte de dados Fórum Defesa

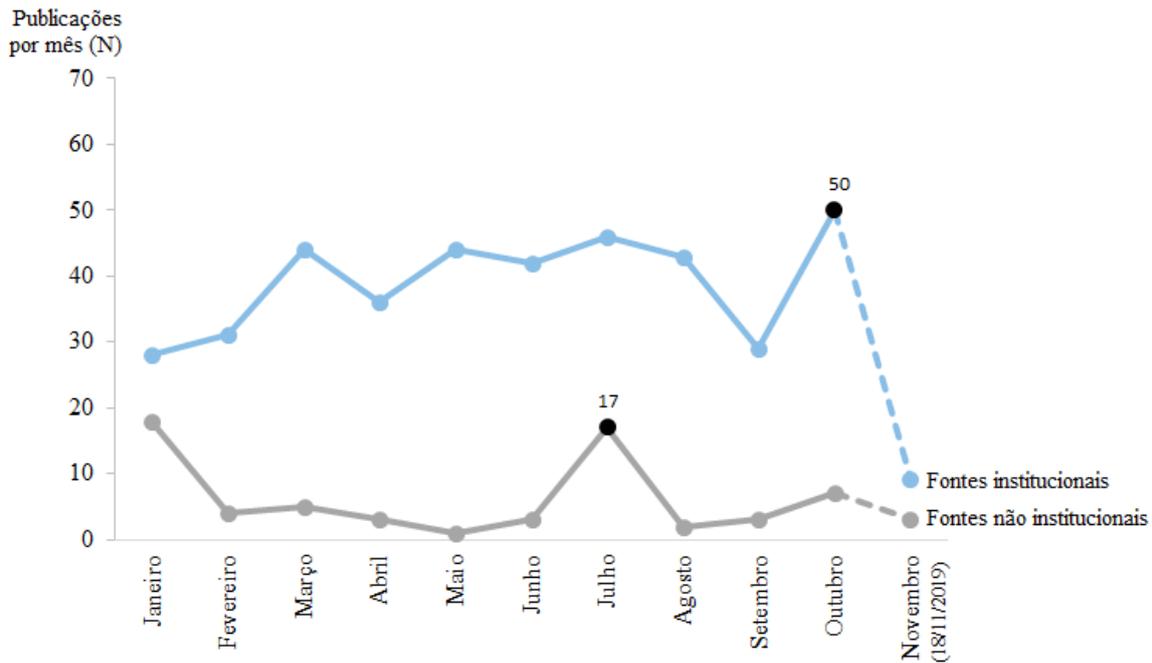


Figura 4.10: Número de publicações por mês em 2019

uma vez que a elevação no número de reações pode derivar de estratégias como a descrita para a Figura 4.11 e não informa sobre a valência das mesmas.

Apesar do número de reações ser pouco informativo, permite conhecer o tipo de publicações que estimulou um maior número de reações dos utilizadores. Na Tabela 4.5, encontram-se expostos os assuntos mais populares e cuja análise sugere que a partilha de publicações relacionadas com a atividade operacional e de apoio à sociedade civil são aquelas que reúnem um maior número de reações.

Além dos quantitativos extraídos, foram calculadas métricas capazes de indicar o comportamento existente em cada fonte de informação. Primeiramente, foi estudado o alcance de cada publicação através da sua relação com os comentários com resultados expostos na Figura 4.12. Nesta Figura, referente à média de comentários por publicação, é possível observar que apesar do elevado número de comentários extraídos das fontes institucionais, são as fontes independentes quem gera, em média, uma maior interação por cada publicação partilhada. A título de exemplo, as fontes de dados que denotam maior alcance nas suas publicações foram o jornal Público (M=124), o Correio da Manhã (M=94) e a plataforma Reddit (M=64). No sentido oposto, a interação por publicação foi mais baixa nas páginas institucionais do Ministério da defesa - “Defesa Nacional” (M=6), do EMGFA - “emgfa” (M=11) e do Exército: “ExercitoRecrutamento” (M=15) e “ExercitoPortuguesPRT” (M=18). Estes resultados sugerem que as fontes de informação independente, embora não sejam dirigidas somente à instituição, geram publicações com elevada capacidade para estimular o eWOM dos seus utilizadores e constituem canais de influência interpessoal relevantes que devem ser monitorizados.

<b>Tema</b>	<b>Publicação</b>	<b>Reações</b>	<b>Fonte</b>	<b>Ano</b>
<u>Fontes institucionais</u>				
Forças Armadas	«O apoio torna-nos mais fortes! Esta foi a vez da atleta Rosa Mota ir deixar uma mensagem de apoio, força e determinação ao Soldado Aliu Camará. Com constante acompanhamento, o militar do Exército português prossegue com uma recuperação bastante favorável no Hospital das Forças Armadas.»	2800	FB do EMGFA	2019
Exército	Achas que tens o que é necessário?	7734	Instagram Exército	2019
Marinha	«Depois da tempestade... a bonança. A fragata D. Francisco de Almeida atravessou os efeitos do furacão Dorian, este sábado, enquanto navegava na baía de Fundy, a oeste da Nova Escócia no Canadá. O navio largou de Halifax para poder afastar-se da tempestade. Tudo correu como esperado e a guarnição e o navio encontram-se bem, tendo mesmo recebido uma "escolta" dos navegadores mais destemidos do oceano Atlântico (...)»	2200	FB Marinha	2019
Força Aérea	«A Esquadra 751 resgatou, esta tarde, três pescadores que se encontravam à deriva dentro de uma balsa salva-vidas, após a embarcação onde navegavam ter naufragado, a 102 quilómetros a sudoeste do Montijo.»	1800	FB Força Aérea	2019
<u>Fontes não-institucionais</u>				
Forças Armadas	«Governo quer creches em quartéis para ter mais mulheres militares.»	942	Público	2019
Exército	«Exército português leva comida e medicamentos a aldeias isoladas de Moçambique.»	2800	Observador	2019
Marinha	«Marinha, Polícia Marítima e Unidade de Controlo de Costa da GNR operam no sul de Espanha e de Itália e em várias ilhas gregas.»	664	Público	2019
Força Aérea	«Altas patentes da Força Aérea detidas por corrupção.»	1700	Correio da Manhã	2016

Tabela 4.5: Descrição das publicações com maior número de reações por tema

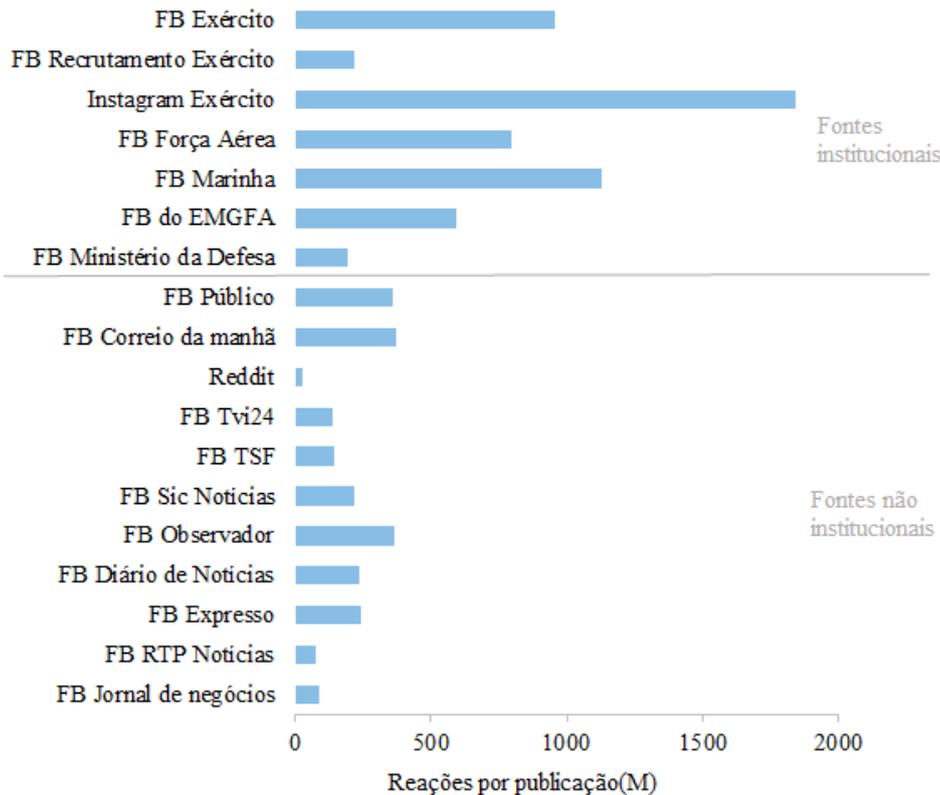


Figura 4.11: Média das reações por publicação em cada fonte de informação

Além do alcance, foi estudado o envolvimento dos utilizadores através do número de comentários que realizam. Na Tabela 4.6, encontra-se o número médio de comentários por publicação dos 3 utilizadores mais ativos para cada fonte de informação. Esta análise indicadora de uma comunidade participativa, destacou positivamente a plataforma Reddit, o jornal Público e as páginas de Marinha e Recrutamento Exército. Salienta-se também que as fontes institucionais apresentaram uma média superior face às fontes noticiosas, sendo este resultado congruente com a tipologia das fontes, os seus objetivos e a população abrangida.

No que se refere aos utilizadores, foi estudado se os comentários obtidos representam a opinião de um público restrito ou alargado. Para isso, foi calculado o número de utilizadores (distintos) em cada fonte de informação. Como se pode verificar na Tabela 4.7, existiu uma diversidade significativa de pessoas a comentar notícias e publicações, em particular nos jornais Público, Correio da Manhã e nas fontes institucionais. A partir da diversidade observada, conclui-se que há potencial de generalização dos resultados do eWOM para uma população alargada.

Além da análise global de público abrangido, importa também medir o alcance (número de utilizadores que comentaram cada publicação) por fonte de informação. O alcance médio, calculado para cada fonte na Figura 4.13, demonstrou que, apesar da diversidade de utilizadores existentes nas fontes institucionais, a sua interação é pouco frequente quando

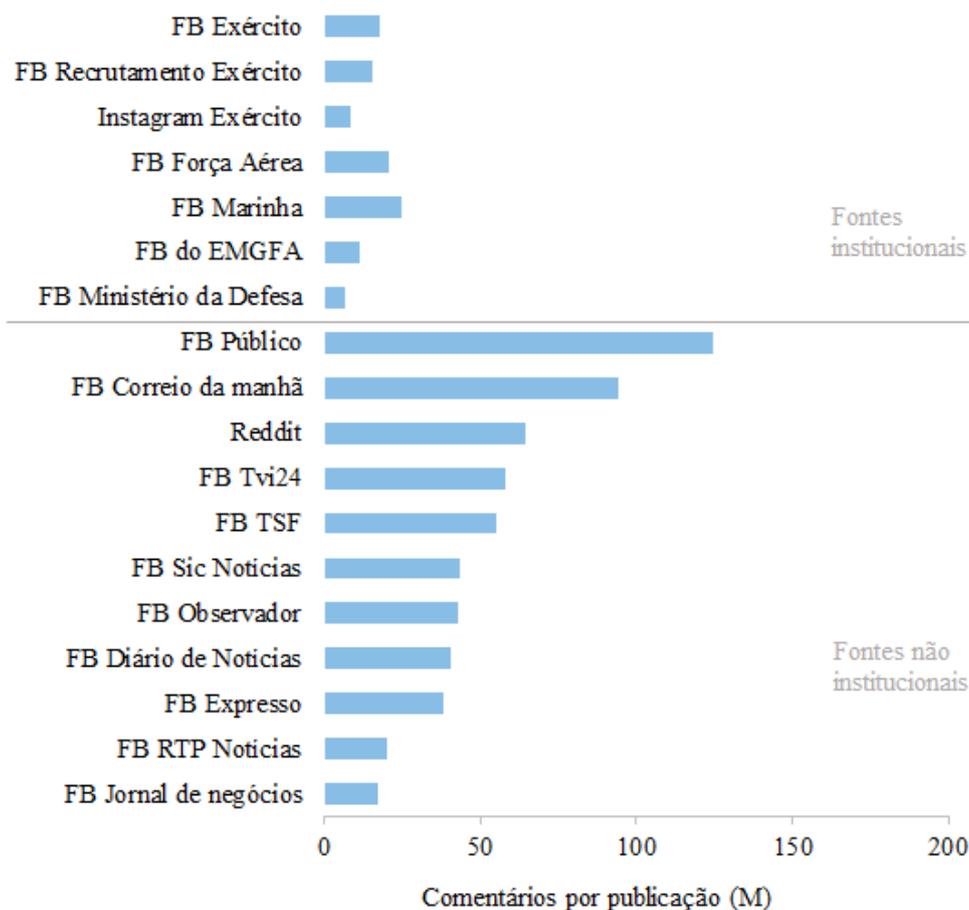


Figura 4.12: Média de comentários por publicação em cada fonte

comparada com os utilizadores das fontes independentes. Globalmente, foram as fontes não institucionais que alcançaram, em média, um maior número de utilizadores em cada publicação, destacando-se os jornais Público ( $M=117$ ) e Correio da Manhã ( $M=87$ ) que apresentaram valores que denotam uma melhor relação com o público. No sentido oposto, as páginas pertencentes ao MDN, EMGFA e Exército apresentaram resultados menos positivos. Esta análise sugere que, apesar das páginas oficiais possuírem um elevado número de utilizadores, não estão a conseguir envolvê-los consistentemente na sua atividade *online*.

Seguidamente, foi analisada a dinâmica dos utilizadores face aos *social media*, sendo esse comportamento um indicador com conteúdo mais informativo que os números estáticos (Peters et al., 2013). Na Figura 4.14, observa-se a evolução temporal da relação entre publicações e comentários por categoria da fonte. Os resultados foram congruentes com a descrição anterior, observando-se que, para as fontes não oficiais, foi em 2018 que se registou a maior interação entre utilizadores e publicações. Esta avaliação temporal, permitiu aferir que – para as fontes de informação utilizadas – a conversa acerca dos temas em análise encontrava-se estabilizada entre 2014 e 2016, tendo apresentado uma subida até 2018. Importa salientar que, apesar da conversa existente, este indicador não permite interpretar a valência positiva ou negativa dos comentários. Assim, a Figura 4.14 permite

Fontes institucionais	Top Users			Fontes não institucionais	Top Users		
	U1	U2	U3		U1	U2	U3
FB Exército Português	5	5	5	FB Correio da manhã	4	3	3
FB Recrutamento Exército	<b>9</b>	<b>8</b>	<b>7</b>	FB Diário Notícias	3	2	2
FB Marinha Portuguesa	<b>10</b>	<b>5</b>	<b>4</b>	FB Expresso	3	2	2
FB Força Aérea	4	4	4	FB Jornal de negócios	2	1	1
FB Ministério da Defesa	3	2		FB Observador	2	2	2
FB do EMGFA	5	5		FB Público	6	3	3
Instagram Exército	4	3		FB Rtp Notícias	2	2	2
				FB Sic Notícias	3	2	2
				FB TSF	2	2	2
				FB Tvi24	2	2	2
				Reddit	<b>14</b>	<b>12</b>	<b>10</b>

Tabela 4.6: Média de comentários por publicação dos principais utilizadores

Fontes institucionais	Utilizadores	Fontes não institucionais	Utilizadores
FB Exército Português	485	Forum Defesa	67
FB Recrutamento Exército	2018	Correio da manhã	563
FB Marinha Portuguesa	456	Diário Notícias	521
FB Força Aérea	517	Expresso	277
FB Ministério da Defesa	174	Jornal de negócios	50
FB do EMGFA	312	Observador	375
Instagram Exército	399	Público	848
		Rtp Notícias	132
		Sic Notícias	320
		TSF	157
		Tvi24	218
		Reddit	374

Tabela 4.7: Número de utilizadores por fonte de informação

concluir que o fluxo de informação possui variações e que apresenta uma diferença elevada entre os dois tipos de fonte. A título de exemplo, em 2019, as fontes institucionais registaram uma média de 30 comentários por publicação enquanto as fontes não-institucionais atingiram uma média de 122 comentários. Globalmente, esta análise reforça a importância de monitorizar os diversos canais de comunicação, pois embora as entidades da Defesa controlem as suas páginas oficiais, verifica-se que o eWOM ocorre também – e em grande medida – através de fontes diversas.

O último passo teve como objetivo identificar os principais termos representativos dos comentários obtidos nas fontes institucionais e não institucionais para o setor da Defesa. Para isso, foi realizado o tratamento da informação (ver 4.4) e utilizada a técnica «nuvem de palavras» exposta na Figura 4.15 e que reflete os termos e respetiva frequência de utilização. Na Figura, os termos a cinzento provêm maioritariamente de fontes de informação não institucional e destacam comentários sobre as instituições da Defesa («militares»,

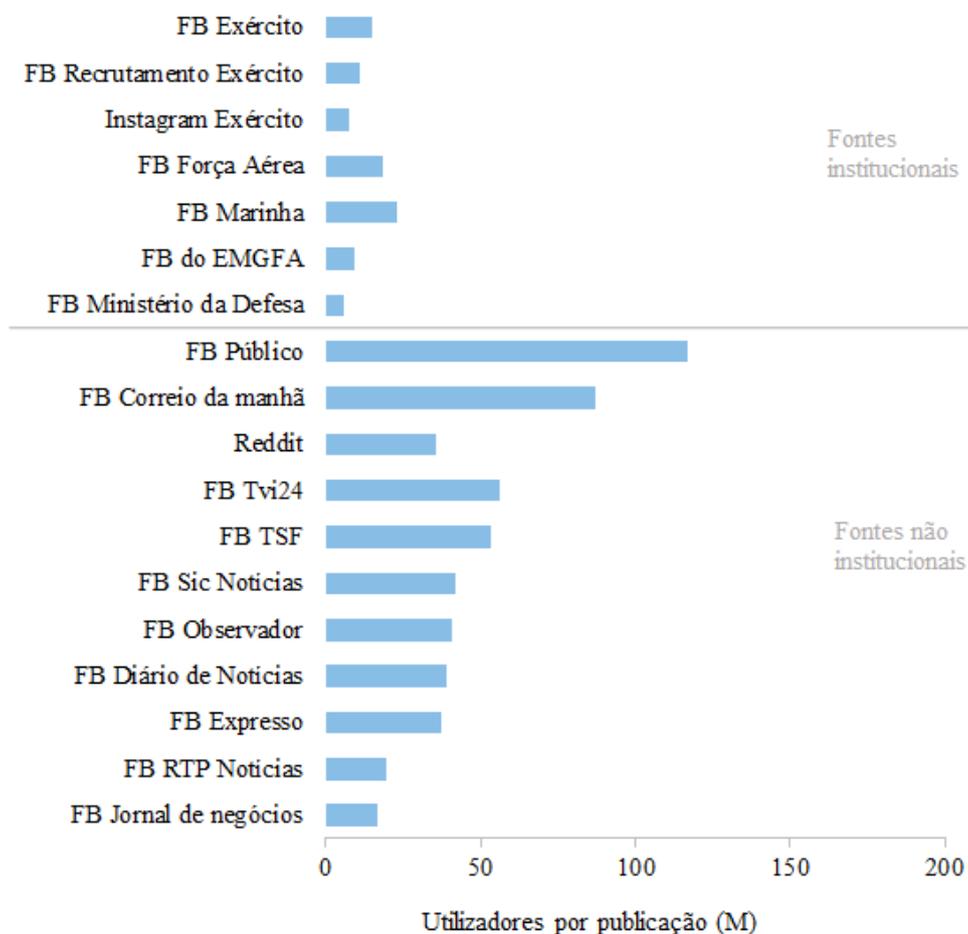


Figura 4.13: Número médio de utilizadores por publicação para cada fonte de informação

«exército», «Forças Armadas»); a sua associação à «guerra»; a discussão sobre o serviço militar obrigatório («militar obrigatório», «serviço militar»), sua utilidade para o país («Portugal», «País») e recrutamento militar («ser», entre outros). No que se refere aos termos que distinguem a categoria das fontes institucionais - a azul - estes denotam conteúdos mais positivos e que poderão derivar de utilizadores com uma pré-disposição para o tema destas páginas, mas também o efeito da gestão de comentários realizada pela organização (eliminação de opiniões negativas). Por último, a preto, encontram-se termos que se distribuem de forma equilibrada pelos grupos de fontes de informação e que incluem palavras pouco relevantes para o estudo («todo») mas também, termos representativos das instituições («marinha», «força aérea», «tropa», «soldado»). Esta análise sugere a aplicabilidade deste tipo de técnica, apresentando porém apresentou conteúdos pouco relevantes para a obtenção das opiniões acerca das Forças Armadas.

Globalmente, a análise descritiva da informação permitiu obter um conjunto de indicadores úteis à análise das fontes de informação, refletiu uma diferença significativa entre fontes, mas também entre gestão institucional vs não institucional. A avaliação realizada foi congruente com a hipótese colocada na Secção 4.3.2, observando-se uma menor qualidade

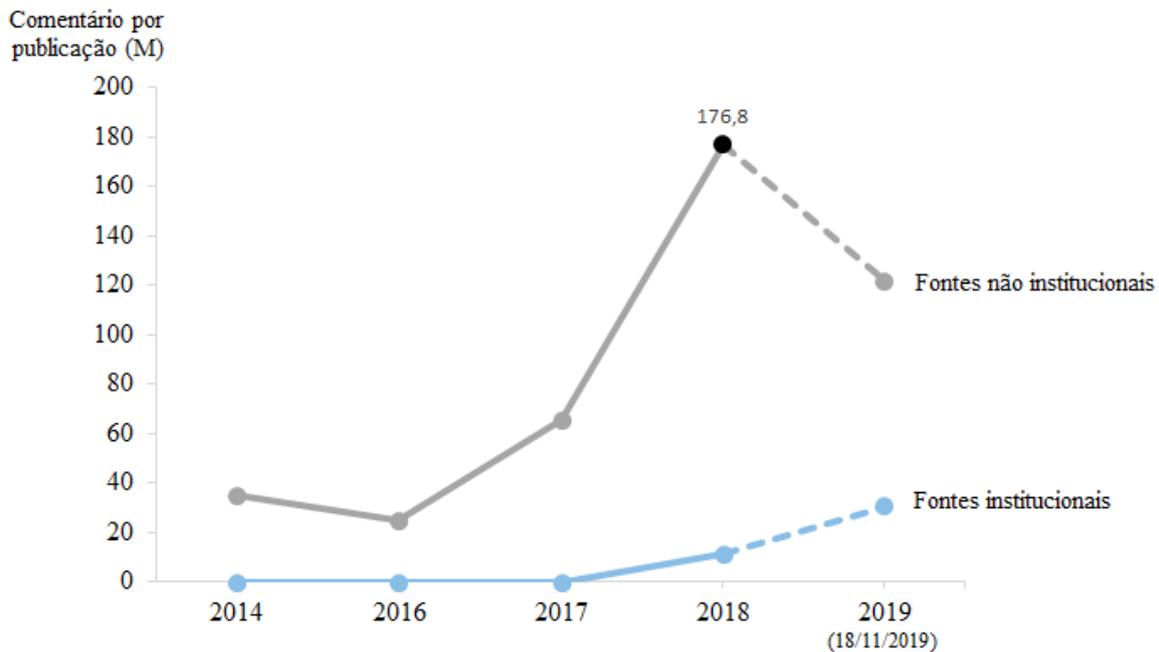


Figura 4.14: Média anual de comentários por publicação e tipo de fonte

e diversidade da informação em fontes institucionais caracterizada pela inferior interação dos utilizadores, frequência de termos e média de caracteres nos comentários. Portanto, conclui-se que as fontes de informação independentes possuem maior potencialidade para a análise do eWOM.

#### 4.4 Preparação da informação

A Secção de Preparação da informação visa garantir o melhor desempenho das técnicas de *text mining* e processamento da língua natural através de um conjunto de processos de limpeza, uniformização e transformação da informação. Assim, nesta fase, destacam-se as seguintes tarefas:

- **Uniformização dos formatos:** o atributo «Data» foi extraído com diferentes formatos de acordo com sua fonte de origem. Assim, o formato original foi transformado para formato *dd/mm/hh* em todos os registos para os quais era aplicável.
- **Junção da informação obtida em diversas fontes:** Este processo consistiu na criação de um *dataset* global com todos os registos extraídos nas diversas fontes de informação. Neste *dataset* foram criados atributos a partir dos existentes, nomeadamente: atributos «mês» e «ano» que permitiram análises temporais entre todas as fontes de informação (a título de exemplo, a comunidade virtual Reddit possui apenas o ano da publicação); atributo «Sentimento» constituído pelas categorias Positivo, Neutro e Negativo correspondentes à valência de cada comentário.



1. Transformação de palavras: Foi observado que o texto recolhido a partir de *social media* apresentou um conjunto de termos que apesar de diferentes, são referentes à mesma semântica. Assim, para melhorar o desempenho das técnicas de contabilização de palavras, foi criado um dicionário para identificação e substituição de termos pelo seu significado comum: a título de exemplo, o termo «SMO» é transformado em «Serviço Militar Obrigatório».
2. Eliminação de utilizadores no texto: No conteúdo recolhido em *social media* foram identificados nomes de utilizadores inseridos em comentários e que poderiam interferir na análise dos resultados. Estes nomes, não possuíam um símbolo identificativo ou forma de distinção ao longo do texto. Para proceder à sua eliminação, foram apagados os nomes existentes em duas listas com os principais nomes e apelidos disponibilizados pelo Instituto de Registo e Notariado.<sup>16</sup>
3. Normalização dos dados: No que se refere à normalização do conteúdo extraído, foram aplicados um conjunto de procedimentos que visam tratar e normalizar os dados para aumentar a qualidade da análise posterior. Estes procedimentos recorreram ao *Natural Language Toolkit* (NLTK) para *Python* e consistiram em:
  - Aplicação de letras minúsculas a todas as palavras;
  - Eliminação de pontuação;
  - Eliminação de *stopwords* portuguesas (quando aplicável);
  - Tokenização (quando aplicável);
  - Stemming (RSLPStemmer) (quando aplicável).

A experiência decorrente dos processos nesta fase de Preparação da informação, sugere a importância da limpeza e uniformização na utilização do conteúdo não estruturado em *social media*. Os passos descritos, consideram-se concluídos quando é atingido um melhor desempenho nas fases de análise posteriores.

## 4.5 Extração do conhecimento

De acordo com a metodologia proposta, a fase de Extração do Conhecimento pretende obter os temas, opiniões e polaridades provenientes dos comentários públicos extraídos e preparados anteriormente.

Para alcançar o objetivo proposto, esta secção inicia-se com a necessidade de obter a polaridade do sentimento expresso em cada comentário. Face ao número elevado de comentários extraídos, foi testado o desempenho de um modelo de classificação automática previamente treinado para classificar o conjunto de dados. Porém, este sistema não

---

<sup>16</sup>Extraído em novembro de 2019 de [https://www.irn.mj.pt/sections/irn/a\\_registral/registos-centrais/docs-da-nacionalidade/vocabulos-admitidos-e/downloadFile/file/Lista\\_de\\_nomes\\_2017\\_11.pdf](https://www.irn.mj.pt/sections/irn/a_registral/registos-centrais/docs-da-nacionalidade/vocabulos-admitidos-e/downloadFile/file/Lista_de_nomes_2017_11.pdf)

apresentou um desempenho eficaz, tornando necessária uma classificação manual de sentimento que se previa morosa para a totalidade de registos. Assim, a partir da análise na descrição dos resultados (Secção 4.3.3) foi selecionada a fonte de informação Reddit para classificação e demonstração da análise ao eWOM institucional.

A Secção 4.5.1 caracteriza a análise de sentimento manual realizada a partir dos comentários do Reddit e descreve os resultados obtidos para o Exército, replicando-os para os demais ramos e Forças Armadas. Globalmente, a interpretação das opiniões positivas e negativas obtidas permitiu reconhecer o eWOM como processo dinâmico, capaz de providenciar informação relativa à imagem organizacional e congruente com indicadores institucionais (a título de exemplo, instituições com maiores dificuldades de recrutamento, apresentaram uma reputação mais negativa).

Seguidamente, a Secção 4.5.2 descreve uma análise de sentimento automática através da aplicação de técnicas de aprendizagem automática (supervisionada) para obtenção de um modelo capaz de classificar automaticamente a polaridade de sentimento em novos comentários. Esta modelação baseou-se na totalidade de comentários do Reddit para treinar e testar um modelo eficaz na classificação de novos dados sem intervenção humana.

Por último, a Secção 4.5.3 demonstra a aplicação de técnicas não supervisionadas através da análise de tópicos e revela a deteção de cinco tópicos emergentes nos comentários no sector da Defesa.

A partir da Extração do Conhecimento realizada, ficou demonstrado que a informação partilhada em aplicações de *social media* permite conhecer a imagem da instituição a partir do eWOM, identificar os principais temas de discussão e monitorizar essa informação ao longo do tempo.

#### **4.5.1 Análise de sentimento manual**

Esta secção refere-se à análise de sentimento para exploração das características, termos e dinâmica das opiniões em comentários extraídos, tendo como objetivo desenvolver indicadores úteis à gestão da reputação institucional.

A escolha deste tipo de análise decorreu da dificuldade em obter um bom desempenho das classificações automáticas com recurso à tradução e utilização de modelos previamente treinados e disponibilizados pela biblioteca TextBlob. Os resultados obtidos através desta biblioteca, apresentaram dificuldade em discriminar os sentimentos (maioritariamente polaridades próximas de 0) que poderão justificar-se pela especificidade do conteúdo, caracterizando-se pelo recurso à ironia, termos específicos do contexto militar, contextualização face ao contexto português e erros ortográficos que dificultam a tradução e classificação automática.

Perante as dificuldades de automatização na classificação de sentimentos, procedeu-se à categorização manual de cada comentário. Contrariamente às diretrizes delineadas, não

Organismo público	Publicações		Comentários		Comentários por publicação	
	N	%	N	%	M	DP
Marinha	14	29%	276	31%	19.64	22.87
Exército	12	25%	238	26%	19.83	22.59
Força Aérea	11	23%	106	12%	9.64	8.38
Forças Armadas	11	23%	283	31%	25.73	41.19

Tabela 4.8: Descritivas do Reddit por organismo público

foi possível classificar manualmente todos os comentários, optando-se por um subconjunto de fontes de informação adequadas para a prova de conceito. Neste caso, foi escolhida a comunidade Reddit, cuja escolha foi baseada nos critérios de independência face às instituições militares e características do seu conteúdo, nomeadamente, um elevado número de palavras, comentários e publicações. Na Tabela 4.8 pode observar-se que o conjunto de dados proveniente desta fonte contém publicações e comentários sobre os temas Exército, Marinha, Forças Armadas e Força Aérea em quantidades semelhantes (com exceção da última). Esta visão comparativa que se segue, poderá ser uma mais-valia na compreensão das semelhanças e diferenças entre o Exército e os restantes.

Apesar da distribuição equilibrada de comentários por tema, foi estudado o alcance que as publicações obtiveram junto dos *redditors*<sup>17</sup> através da média de comentários por publicação. A comparação desta dinâmica entre instituições encontra-se na Tabela 4.8 e demonstra como o setor da Defesa - avaliado através das «Forças Armadas» - obteve um maior alcance junto dos utilizadores. Ou seja, comparativamente aos ramos, as suas publicações geram, em média, um número superior de comentários e com elevada discrepância entre as publicações avaliadas (M=25.73; DP=41.19). Relativamente à distinção entre ramos, observou-se que o Exército e a Marinha apresentaram um padrão semelhante, com uma interação por publicação bastante superior à Força Aérea. Estes resultados sugerem que a informação partilhada sobre as Forças Armadas, a Marinha e a Exército atinge um elevado número de utilizadores e estimula uma interação superior à Força Aérea.

Confirmada a existência de eWOM nesta informação, importa compreender o seu conteúdo e valência. No que se refere à valência dos comentários extraídos, a Figura 4.16 revelou uma prevalência dos sentimentos negativos face aos positivos, sugerindo um eWOM desfavorável sobre a instituição militar. Esta opinião negativa ocorreu com maior frequência nas Forças Armadas em geral (N=199), mas também para os respetivos ramos, com maior incidência nos comentários sobre a Marinha (N=172) e o Exército (N=144). Estes dados são congruentes com a maior dificuldade de recrutamento nestes ramos e com os estudos que revelam maior satisfação organizacional dos militares na Força Aérea (Coelho et al., 2019). Importa também salientar que, de acordo com a literatura, a informação negativa online tende a exercer um efeito superior à informação positiva. Nesse sentido, esta quantidade de opiniões negativas poderá exercer um impacto bastante significativo sobre

<sup>17</sup>Participante da comunidade virtual Reddit

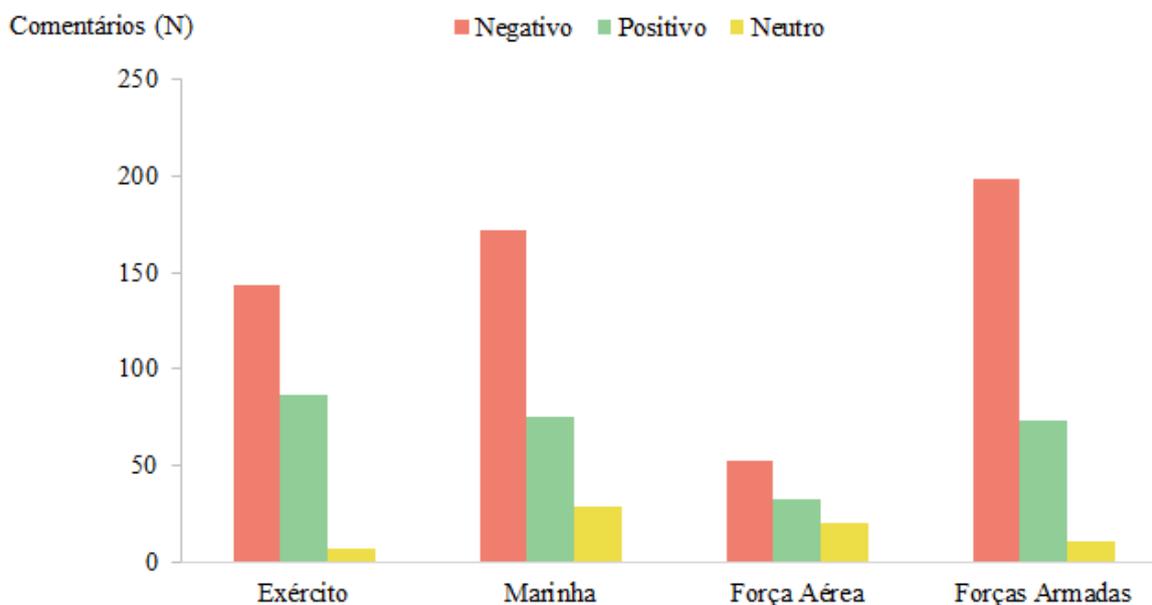


Figura 4.16: Número de comentários por sentimento e organismo público

Organismo público	Sentimento		
	Negativo	Positivo	Neutro
Marinha	62%	27%	11%
Exército	61%	37%	3%
Força Aérea	50%	31%	19%
Forças Armadas	70%	26%	4%

Tabela 4.9: Percentagem de comentários por sentimento e organismo público

a reputação dos ramos, com consequências na sua atratividade junto da população-alvo e respectivos referenciadores.

Além de analisar a opinião pública através da frequência da polaridade de sentimento, importa também estudar a sua proporção face ao total de comentários em cada organismo da Defesa. De acordo com os dados na Tabela 4.9 pode observar-se que a informação partilhada publicamente é maioritariamente negativa para as instituições avaliadas. Nesta análise, o Exército apresentou uma imagem aparentemente negativa junto da população avaliada, revelando 61% de comentários negativos, com uma situação comparável à Marinha e às Forças Armadas em geral. Como referido anteriormente, o padrão difere para a Força Aérea, onde se observam opiniões equilibradas, com apenas metade dos comentários a apresentar valência negativa. Assim, a polaridade dos comentários remete para uma diferença na imagem dos utilizadores acerca dos ramos e sugere uma capacidade discriminativa deste tipo de análise comparativa.

Além da prevalência de sentimentos, foi também avaliada a extensão do conteúdo textual para cada categoria. Na Tabela 4.10 observa-se que as opiniões neutras foram as mais sucintas, seguindo-se as positivas que foram menos elaboradas face às negativas. A

Organismo público	Sentimento		
	Negativo	Positivo	Neutro
Marinha	228	96	40
Exército	216	185	45
Força Aérea	357	223	142
Forças Armadas	435	262	38

Tabela 4.10: Média de caracteres por comentário e sentimento

comparação da média de caracteres destacou que a informação partilhada sobre as Forças Armadas é mais extensa, que a Marinha possuiu comentários significativamente mais curtos e que o Exército revelou a quantidade mais equilibrada entre número de elementos textuais nos comentários positivos e negativos. Globalmente, estes resultados sugerem que além de existir um menor número de comentários positivos, estes possuem menos termos disponíveis para análise e reforçam a incapacidade para equilibrar o efeito do eWOM negativo.

#### 4.5.1.1 Análise de sentimento aplicada ao Exército

No que se refere à análise de sentimento no Exército Português, esta secção pretende conhecer as opiniões partilhadas publicamente e inferir sobre a sua imagem institucional.

Primeiramente, na Figura 4.17 foi analisada a variabilidade temporal dos comentários e testada a hipótese de existirem eventos associados ao aumento ou diminuição dos comentários e sua valência. Como esperado, existiram variações anuais no número de comentários, salientando-se um aumento da interação em 2017, uma discrepância superior entre comentários positivos e negativos nos anos 2017 e 2018 e a diminuição dos comentários em 2019<sup>18</sup>.

A análise dos acontecimentos noticiados e as publicações do Reddit nesse espaço temporal permite associar o aumento do fluxo de informação com o roubo de munições em Tancos e os incêndios desse verão (Pedrógão e Mação), mas também observar uma maior proporção de comentários negativos quando foram divulgadas notícias sobre o défice de recursos humanos causado por dificuldades de retenção e também a diminuição dos requisitos de acesso aos quadros permanentes da Escola de Sargentos do Exército (ESE).

Além desta análise, procedeu-se à análise de frequência de termos utilizados pelos utilizadores e por polaridade de sentimento. Numa perspetiva visual, a frequência de palavras foi analisada através da nuvem de palavras na Figura 4.18. Aqui, o vermelho representa palavras encontradas principalmente em comentários negativos, o verde indica termos nos comentários positivos e o preto conteúdo presente em ambas as categorias.

<sup>18</sup>Em 2019, a extração da informação incluiu os meses de janeiro a novembro e não permite excluir a possibilidade de variação após essa data.

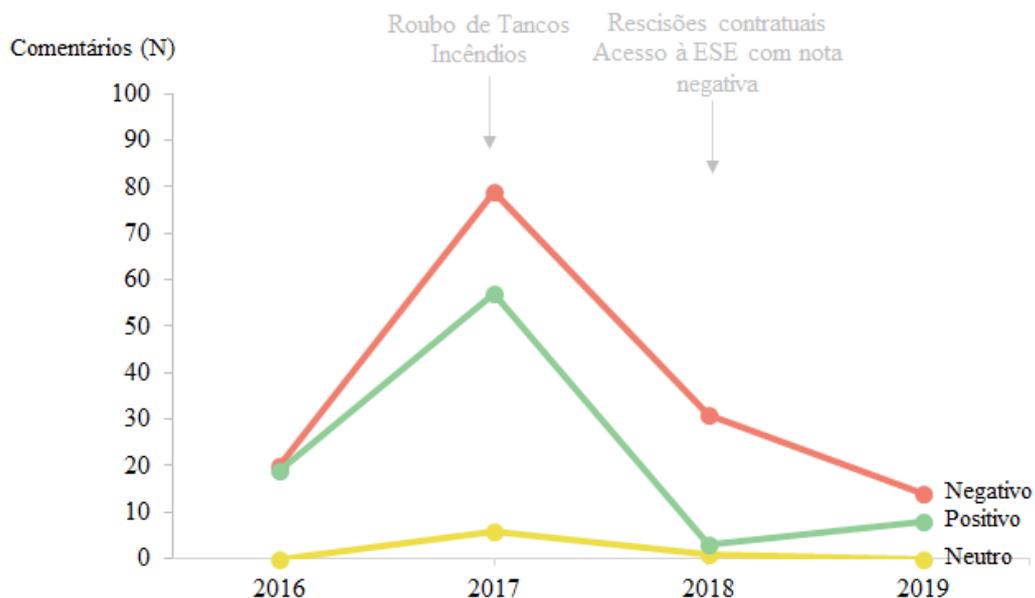


Figura 4.17: Comentários anuais por sentimento - Exército

No que se refere ao eWOM negativo, observou-se uma percepção negativa sobre a instituição, com o nome a aparecer maioritariamente neste tipo de comentários. No mesmo sentido, os termos «portugal», «pátria», «defender», «combate», «guerra» e «dinheiro» referem-se a comentários relacionados com a opinião negativa acerca da utilidade pública da instituição e pertinência do erário público atribuído. Concomitantemente, este eWOM é também constituído por termos relacionados com a imagem de empregador, destacando-se «contrato», «mercado de trabalho», «pessoal» e «vida». Estes termos estão normalmente associados às condições contratuais oferecidas, à sua comparação às alternativas existentes no mercado de trabalho e às dificuldades de recrutamento existentes. Salienta-se também o termo «Força Aérea», utilizado em comentários de polaridade negativa que recomendam esse ramo (por oposição ao Exército).

No sentido positivo, a nuvem de palavras na Figura 4.18 destaca os termos relacionados com a formação académica ou profissional proporcionada («curso», «formação», «licenciatura», «academia», «área») sendo congruente com os resultados de estudos nacionais (Silva et al., 2019) e internacionais (Jorlöv and Venngren, 2014), que têm revelado a importância das oportunidades de formação no contexto militar. Simultaneamente, destacam-se características do empregador, nomeadamente, especificidades da profissão das armas («militares», «funções», «campanha», «operação»), o «serviço», as categorias de oficiais («oficiais») e praças («soldado») e as provas de ingresso («ingresso», «correr», «óculos»). Estes resultados são congruentes com dados obtidos por questionário aos jovens no Dia da Defesa Nacional, nos quais foi referido interesse no ramo devido à componente operacional, experiência profissional e possibilidade de continuar os estudos (Reis et al., 2019).

Por último, é possível encontrar termos que embora frequentes não distinguem entre as categorias de sentimento, tais como, «recruta», «especialidade», «militar» ou «sargento».



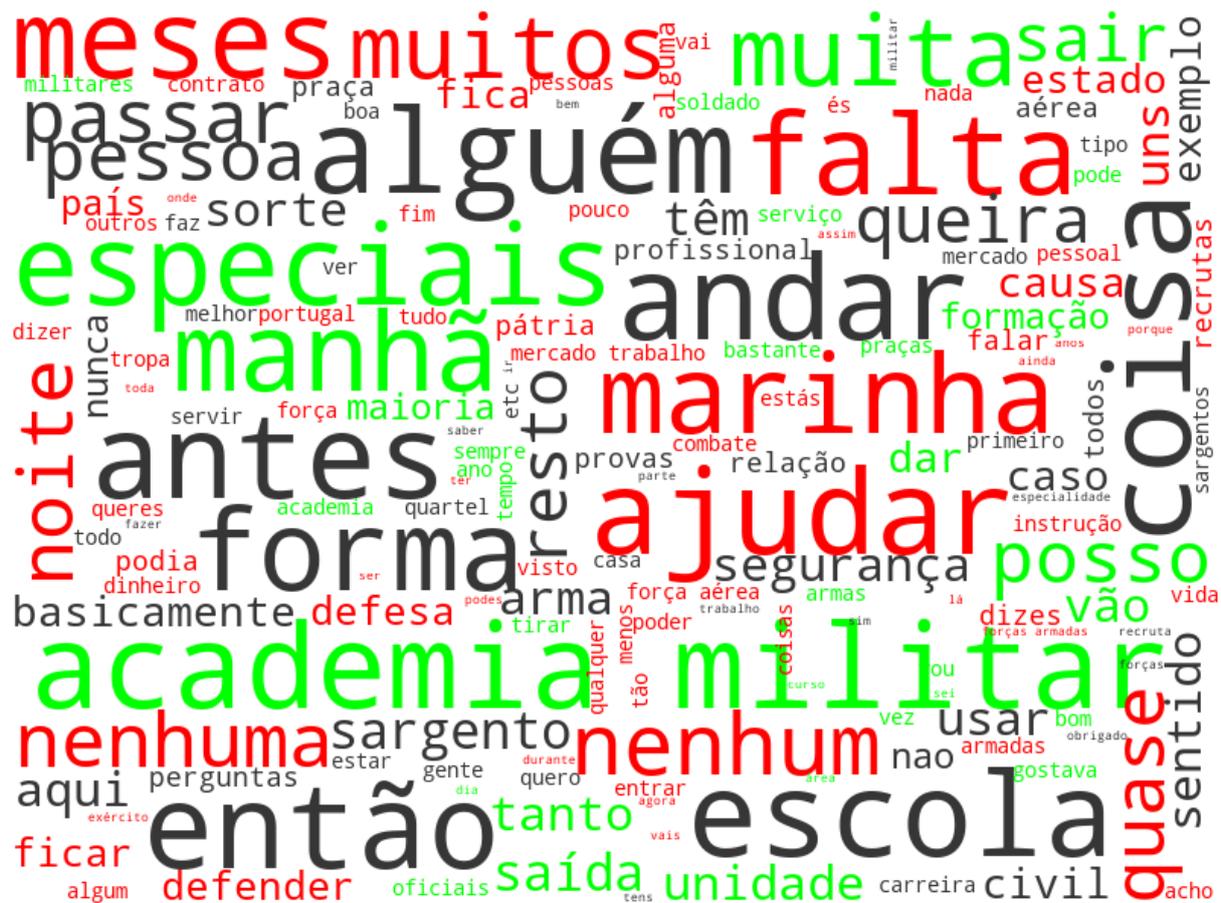


Figura 4.19: Nuvem de palavras (TF-IDF) - Exército

#### 4.5.1.2 Análise de sentimento aplicada à Marinha Portuguesa

Para testar a validade da metodologia foi necessário testar a aplicação da análise de sentimento em novos contextos institucionais. Com esse objetivo e de acordo com o previsto durante os processos de seleção e extração da informação, segue-se a replicação das técnicas anteriores para a análise de sentimento dos outros ramos das Forças Armadas. Tendo sido demonstrada a aplicabilidade de diversas técnicas no caso do Exército, segue-se uma análise mais sumária para os restantes ramos.

No que se refere ao caso da Marinha, embora esta instituição possua características organizacionais bastante distintas do Exército, os comentários obtidos também demonstraram um padrão tendencialmente negativo e congruente com as dificuldades de atratividade registadas nesta instituição. A análise temporal na Figura 4.20 revela que desde 2017 os sentimentos negativos tornaram-se significativamente superiores aos positivos e atingiram uma assimetria significativamente elevada em 2019. Relativamente aos temas que originaram comentários nesse último ano, salientam-se o investimento público na instituição e a divulgação de um vídeo não oficial (filmado durante um exercício militar).



Figura 4.20: Comentários anuais por sentimento - Marinha

Seguidamente, o conteúdo dos comentários foi avaliado através da lista de termos e respetiva frequência, expressos em formato de nuvem de palavras na Figura 4.21. Relativamente aos resultados obtidos, a perceção negativa (a vermelho) revelou opiniões acerca da utilidade pública («dinheiro», «custo», «orçamento» «contra»); condições de trabalho («vida», «merda»), o insucesso tecnológico («drone») e o seu desempenho («míssil», «afundar»). No sentido positivo (a verde), destacaram-se a formação académica (escola «naval», «oficial», «matemática») e características específicas do equipamento e missão da instituição («submarino», «guerra», «atacar», «portugal», «base»). No que se refere aos termos não discriminativos (a preto), salienta-se que apesar da imagem tendencialmente negativa, o termo «Marinha» não denotou uma conotação prejudicial ao Ramo.

Globalmente, a análise de sentimento na Marinha indica uma tendência anual para a deterioração da sua imagem e confirma a capacidade para obter indicadores sobre a imagem e reputação da instituição através desta metodologia. Porém, os resultados obtidos possuem limitações significativas na quantidade e diferenciação da informação, potencialmente relacionada com o baixo número de caracteres existentes nos comentários deste Ramo (ver Secção 4.10).

#### 4.5.1.3 Análise de sentimento aplicada à Força Aérea Portuguesa

Seguidamente, os procedimentos foram replicados na Força Aérea Portuguesa (FAP), tendo sido confirmada a sua reputação positiva e a aplicabilidade da metodologia proposta.

O primeiro passo consistiu na evolução temporal dos conteúdos partilhados com resultados na Figura 4.22. Entre 2017 e 2018, esta instituição apresentou um eWOM tendencialmente negativo, porém, em 2019 - até ao momento da extração de informação - a Força Aérea conseguiu inverter a sua situação. Aparentemente, a informação partilhada atingiu





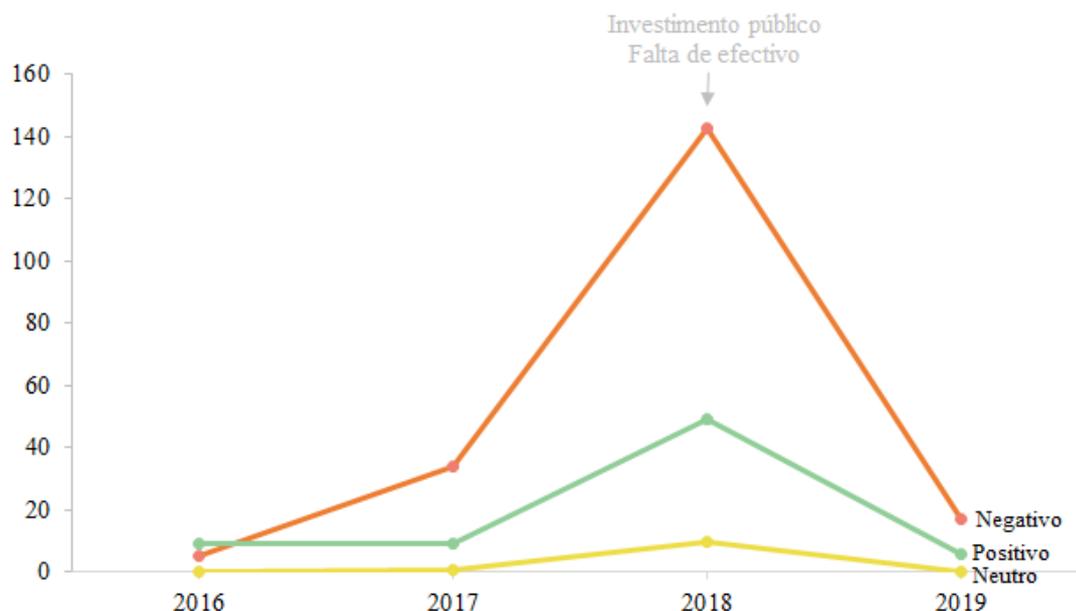


Figura 4.24: Comentários anuais por sentimento - Forças Armadas

#### 4.5.1.4 Análise de sentimento aplicada às Forças Armadas

Além da análise por ramo, importa testar se a reputação das Forças Armadas (FA) pode ser avaliada globalmente através dos dados extraídos. Nesse sentido, foram selecionados os comentários sobre este tema, aplicados os procedimentos descritos anteriormente e concluída a sua aplicabilidade à imagem dos cidadãos sobre o sector da Defesa.

A Figura 4.24 confirma a variação anual no eWOM sobre as Forças Armadas. A informação partilhada foi particularmente negativa nos anos 2017 a 2018, mas melhorou quando se registaram publicações referentes ao investimento estatal nestas Forças, à falta de efetivo militar e à possibilidade de reintrodução do serviço militar obrigatório.

Seguidamente, procedeu-se à análise visual referente à frequência de termos e respetiva polaridade através da nuvem de palavras na Figura 4.25. Esta figura denota a importância do Exército na reputação das Forças Armadas (justificado pelo elevado efetivo militar que exige face aos restantes ramos) e salienta o termo «ser», congruente com os resultados individuais dos ramos e que remete para a personalização da função: «ser» militar. A informação partilhada entre utilizadores destacou também as «forças armadas» globalmente e o serviço militar («serviço militar», «militar obrigatório»).

Relativamente aos comentários com valência negativa, a vermelho encontram-se os termos mais expressivos sobre a opinião negativa das Forças Armadas. Como se pode observar, a utilidade pública relacionada com a importância do serviço militar obrigatório («militar obrigatório», «contra»), missão, pertença à *North Atlantic Treaty Organization* (NATO) e relação custo/ benefício para o erário público («defesa», «portugal», «preciso», «estado», «pagar», «dinheiro», «nato»), a possibilidade de conflito armado («guerra»), a falta de efe-



classificação foi baseada no SentiLex-PT 02 (Silva et al., 2012), um léxico português recomendado para análises deste tipo, cuja composição inclui 7014 *lemmas* (dos quais, 4,779 adjetivos, 1,081 nomes, 489 verbos e 666 expressões da língua portuguesa) juntamente com o respetivo sentimento e polaridade.

Seguidamente, foram utilizadas técnicas de aprendizagem supervisionada, nomeadamente, os classificadores generativos (*Naive Bayes*) e discriminativos (Regressão Logística e *Support Vector Machine*) que através de um conjunto de dados para treino conseguem aprender padrões com capacidade de distinguir sentimentos em novos casos. Foram testados modelos com todos os classificadores selecionados, com extração através *Bag-of-words* e *Term frequency-inverse document frequency* (frequência mínima de 5), separação por unigramas, e normalização do texto por eliminação de afixos (*stemming*). Cada modelo recorreu à divisão do *dataset* entre treino (80%) e teste (20%).

Esta análise aos comentários do Reddit, sugere que o treino de modelos de classificação permite obter uma solução automática capaz de atribuir a polaridade de sentimento em comentários extraídos de *social media*.

### 4.5.3 Análise de tópicos

A Secção de modelação de tópicos pretende testar a capacidade da metodologia para detetar temas emergentes a partir dos comentários do Reddit. Esta hipótese foi testada com recurso às técnicas LSA e LDA e diferentes parâmetros, tendo sido obtido o melhor modelo de tópicos a partir do método de extração IDF (termos com frequência mínima de 3 e sequências até 3 elementos) e redução dimensional (LSA).

Na Tabela 4.11 podem distinguir-se os 5 tópicos que surgem nos comentários sobre os temas em análise e os seus dez termos mais relevantes. O tópico 1 apresenta termos genéricos relacionados com a prestação de serviço militar no Exército. O tópico 2 apresenta maior especificidade e revela a discussão sobre a utilidade das Forças Armadas, na qual se abrange a obrigatoriedade do serviço militar e a pertença à NATO. Relativamente aos tópicos 3 e 4, estes abordam respetivamente a componente operacional das Forças Armadas e os ramos Força Aérea e Marinha. Por último, o tópico 5, embora pouco específico, denota uma imagem negativa acerca das instituições.

Globalmente, a análise de tópicos possui potencialidade para detetar tendências na informação partilhada sobre as instituições em *social media*, porém este tipo de análise requer um uma maior quantidade de dados disponíveis para obter resultados mais significativos.

Tópico 1	Tópico 2	Tópico 3	Tópico 4	Tópico 5
ser	militar	forças	marinha	ainda
militar	serviço	forças armadas	forças	bom
exército	obrigatório	armadas	forças armadas	porque
porque	serviço militar	ser	armadas	xd
ter	militar obrigatório	porque	alguém	ter
serviço	serviço militar obrigatório	força	força	nao
lá	nato	militares	aérea	anos
forças	armadas	mundo	força aérea	dinheiro
fazer	forças armadas	qualquer	marinha portuguesa	merda
onde	forças	armas	dá	mau

Tabela 4.11: Tópicos emergentes (top 10 dos termos mais relevantes)

## 4.6 Avaliação

A fase de avaliação tem como objetivo avaliar e selecionar os resultados e modelos obtidos na análise de sentimento e de tópicos. Inicialmente, esta secção confirma a adequabilidade da avaliação manual neste tipo de informação. Seguidamente, demonstra a potencialidade de treinar modelos de classificação com um desempenho aceitável e encerra com o teste ao modelo de tópicos com a categorização temática de novos comentários.

### 4.6.1 Análise de sentimento manual

Para testar o nível de subjetividade da categorização manual de sentimentos descrita na Secção 4.5.1, procedeu-se à avaliação com recurso a dois avaliadores militares e especialistas na área da psicologia e recursos humanos. Este processo permitiu calcular a concordância entre avaliadores através da medida *Kappa* de Cohen, revelando um resultado ( $k=0.62$ ) moderado e aceitável de classificação inter-avaliador para este tipo de técnica (McHugh, 2012).

### 4.6.2 Análise de sentimento automática

Relativamente ao treino de modelos para classificação automática de sentimento, a Tabela 4.12 revela o desempenho calculado a partir da comparação entre as classificações de referência e as classificações automáticas obtidas, com os valores de Accuracy e as restantes com recurso ao parâmetro de cálculo «macro». O primeiro modelo, referente ao classificador baseado na soma de polaridade dos termos (léxico de sentimentos), revelou um desempenho modesto (62% de Accuracy e 49% de *F1*). Relativamente aos restantes modelos com recurso à aprendizagem automática, estes apresentaram resultados globalmente aceitáveis, com exceção dos modelos com extração TF-IDF, cujo comportamento de algoritmos como o *Naive Bayes* tendem a apresentar um desempenho inferior quando comparado com extração por contagens.

	<b>Accuracy</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-score</b>
<b>Léxico de sentimentos</b>	0.62	0.54	0.45	0.49
<b>Feature Extraction: Bag-of-words</b>				
<i>Naive Bayes</i>	0.70	0.66	0.67	0.67
Regressão logística	0.71	0.69	0.64	0.65
<i>Support-Vector Machine</i>	0.68	0.64	0.62	0.63
<i>Naive Bayes (c/ stemming)</i>	<b>0.74</b>	<b>0.72</b>	<b>0.70</b>	<b>0.71</b>
Regressão logística (c/ stemming)	<b>0.74</b>	<b>0.73</b>	<b>0.66</b>	<b>0.67</b>
<i>Support-Vector Machine (c/ stemming)</i>	<b>0.75</b>	<b>0.73</b>	<b>0.70</b>	<b>0.71</b>
<b>Feature Extraction: TF-IDF</b>				
<i>Naive Bayes</i>	0.68	0.67	0.68	0.60
Regressão logística	0.68	0.73	0.55	0.51
<i>Support-Vector Machine</i>	0.71	0.71	0.62	0.62
<i>Naive Bayes (c/ stemming)</i>	0.70	0.71	0.58	0.55
Regressão logística (c/ stemming)	0.70	0.76	0.58	0.55
<i>Support-Vector Machine (c/ stemming)</i>	0.70	0.69	0.60	0.60

Tabela 4.12: Resultados do treino de modelos de classificação automática

A negrito destacam-se os três melhores modelos obtidos com recurso à extração *Bag-of-Words*, técnica de normalização *stemming* (versão portuguesa do NLTK) e classificadores *Naive Bayes* (parâmetros Standard), SVM (parâmetros: C:1.0, Kernel: 'linear', degree: 3) e Regressão Logística (Algoritmo de otimização: Sag). Conclui-se que os modelos encontrados apresentaram resultados aceitáveis quando comparados com os resultados existentes na literatura. Embora os resultados não possam ser diretamente comparáveis - especialmente porque os conjuntos de teste são diferentes -, a título de exemplo, em [Agarwal et al. \(2011\)](#) a classificação de sentimento em *Tweets* apresentou resultados de *accuracy* entre 71% e 74%.

### 4.6.3 Análise de tópicos

Relativamente ao modelo de tópicos desenvolvido, a sua avaliação foi efetuada a partir de 10 comentários reservados para teste. A partir do melhor modelo obtido, foram estimados os tópicos de pertença para cada novo caso e comparada a sua pertinência face ao conteúdo do comentário. No total, 80% dos comentários pertenciam ao tópico indicado, destacando-se principalmente casos com características do serviço militar no Exército (Tópico 1) ou uma perceção negativa da instituição militar (Tópico 5).

## 4.7 Recomendações

A Secção de recomendações visa transformar o conhecimento extraído em sugestões para monitorização e gestão da imagem institucional, iniciando-se com recomendações dirigidas ao Exército e terminando com sugestões ao nível do Estado-Maior das Forças Armadas.

Relativamente à gestão de *social media* do Exército, observou-se que os seus canais de comunicação abrangem um conjunto restrito de redes sociais e que requerem uma estratégia orientadora, estruturante e unificada. Nesse sentido, sugere-se:

- A unificação das páginas *ExercitoRecrutamento* e *ExercitoPortuguesPRT*, uma vez que possuem gestões distintas (a primeira situa-se no Porto, e a segunda em Lisboa), segmentam os utilizadores, dificultam uma imagem congruente e uma mensagem consistente. Criar uma página comum permitiria diversificar o número de utilizadores, aumentar a possibilidade de interação entre militares e potenciais candidatos, e principalmente, publicar conteúdos promotores de uma imagem institucional consistente.
- A elaboração de normas restritivas da criação de páginas geridas pelas Unidades do Exército, pelos motivos acima referidos.
- O evitamento de técnicas focadas no aumento de tráfego na página, uma vez que, o número de reações e interações obtidos através deste tipo de publicações não significam uma relação de qualidade com o utilizador. Propõe-se a ênfase na divulgação de mensagens que forneçam informação relevante, reflitam a personalidade da instituição e criem uma relação com os utilizadores.
- A promoção de conteúdos tendencialmente positivos, tais como a atividade operacional, apoio à sociedade civil e formação disponibilizada.
- Introdução de temas relacionados com a importância da missão institucional e do serviço militar.
- Adopção de estratégias alternativas para lidar com comentários negativos que cumpram o código de conduta. A eliminação deste tipo de comentários leva à perda de informação útil sobre o que está a ser falado e poderá diminuir a participação dos utilizadores, o seu nível de confiança na fonte de informação e aumentar a procura de outras plataformas de *social media*. Formas alternativas de lidar com estes comentários podem incluir a resposta informativa e a estimulação da participação ativa de utilizadores com imagem positiva.

A partir do estudo realizado considera-se também que as necessidades de capital humano e o nível de eWOM negativo detetado em *social media* sugerem a necessidade de uma gestão estratégica nacional da reputação institucional. Contrariamente à situação vigente, esta estratégia deveria ser desenvolvida por um departamento dedicado, com acesso a informação permanentemente atualizada e capaz de incluir um conjunto de funções, das quais se destacam:

- Implementação de um sistema de extração e análise do eWOM em *social media* como ferramenta de *marketing intelligence*;

- Detecção e apoio à decisão acerca de acontecimentos e eWOM publicado (e.g. notícias) com potenciais consequências na reputação institucional;
- Monitorização constante da informação e sentimentos no eWOM em *social media*, permitindo uma intervenção rápida perante o risco causado por mensagens potencialmente prejudiciais;
- Apoio ao desenvolvimento da imagem interna e externa do Exército de forma congruente e consistente nos diversos meios de comunicação online;
- Promoção de uma imagem transparente e informativa relativamente à importância, atuação e utilidade pública da instituição;
- Elaboração de recomendações relativamente a medidas promotoras das características valorizadas positivamente na imagem institucional e de empregador;
- Elaboração de propostas corretivas para melhoria dos aspetos negativos que estão espelhados na imagem institucional e de empregador;

Além das recomendações referentes às redes sociais e marketing interno/externo, importa também comentar a situação atual da imagem do Exército. Globalmente, o eWOM da instituição remete para uma reputação negativa ao nível da imagem de empregador e utilidade pública. No sentido oposto, as especificidades da profissão militar e o acesso à formação são referidos positivamente na internet. Partindo do pressuposto *ceteris paribus*, os resultados são congruentes com a contínua dificuldade de atrair e reter os recursos humanos necessários e requerem mudanças organizacionais baseadas nos aspetos positivos e negativos espelhados nas opiniões públicas. Assim, sugerem-se medidas que aumentem a competitividade do Exército face ao mercado de trabalho (e.g. melhoria das condições oferecidas), melhorem a perceção de valor por parte dos cidadãos portugueses (e.g. atividades de apoio à sociedade civil), garantam a congruência entre imagem institucional e realidade profissional (e.g. manter a especificidade operacional da profissão) e aumentem as oportunidades de estudo e especialização dos militares nas fileiras.

Apesar da prioridade atribuída que as instituições militares têm dado à divulgação através das redes sociais, o presente estudo revelou que existem outras aplicações de *social media* onde surgem opiniões mais elaboradas, com diversidade de conteúdo, maior participação e variedade de utilizadores envolvidos. Esta observação é congruente com a perceção de imparcialidade em comentários de outras pessoas e não à informação disponibilizada pela instituição (Constantinides, 2014). Nesse sentido, propõe-se o desenvolvimento de uma página independente, em formato de comunidade virtual, que inclua informação relevante e fidedigna, com possibilidade de colaboração dos utilizadores na partilha de publicações diversificadas que aumentem o envolvimento dos cibernautas. Contrariamente ao existente, as publicações seriam provenientes de colaboradores previamente validados

e seguiriam o cumprimento de um código de conduta. A utilização deste tipo de meio permitiria unificar as fontes de informação e utilizar um sistema de extração e análise do eWOM com maior facilidade.

Por último, embora o estudo tenha incidido sobre o Exército, a replicação aos restantes ramos e às Forças Armadas como um todo sugere que estes também poderão beneficiar da gestão da sua imagem. A título de exemplo a análise temporal revela que, além do Exército, também a Marinha e as Forças Armadas apresentaram valores preocupantes de eWOM negativo nos últimos anos, sendo recomendável uma intervenção imediata. No que se refere especificamente às Forças Armadas, verificou-se que a sua imagem é mais do que o conjunto de reputações dos seus ramos. Embora na opinião dos cibernautas a profissão militar mantenha características diferenciadoras, esta apresenta também dificuldades de competitividade face ao mercado. Assim, sugere-se que as entidades neste nível organizacional promovam uma imagem pública positiva e unificada com efeitos positivos sobre a missão da Defesa e a reputação das suas instituições. A título de exemplo, fontes de informação das Força Armadas demonstraram informação demasiado específica do contexto militar (e.g. missões) e recurso a páginas diversas, representadas por entidades distintas (EMGFA e Defesa Nacional) e com estratégias de comunicação diferentes e informação frequentemente redundante.

Conclui-se que a gestão da imagem é uma componente fundamental para a manutenção das instituições públicas e, apesar das suas restrições ou especificidades, que a implementação da metodologia apresentada se mostrou capaz de fornecer recomendações concretas de apoio à decisão.

## 4.8 Sumário

O Capítulo 4 tinha como objetivo demonstrar a aplicação da metodologia proposta e testar a sua adaptação a diferentes organismos públicos. Através dos procedimentos e resultados descritos em cada secção, foi possível obter um conjunto de 19 fontes de informação relativas ao setor da Defesa e extrair 13 358 comentários para análise do eWOM. Desta extração, as fontes não institucionais revelaram uma melhor relação com o público e a fonte Reddit destacou-se pela quantidade e diversidade de conteúdos sobre a instituição em análise.

Na ausência de modelos de classificação automática adequados, foram aplicadas técnicas de análise de sentimento manual que revelaram capacidade para obter conhecimento revelante sobre a reputação do Exército e demonstraram a sua replicabilidade em instituições da Defesa. Para testar a possibilidade de desenvolver modelos baseados em aprendizagem automática, a classificação de sentimento e a análise de tópicos revelaram potencialidade para identificar opiniões e detetar temas emergentes nos comentários em *social media*.

Globalmente, os resultados confirmam o funcionamento e adequação da metodologia desenvolvida. A sua aplicação, revelou uma reputação maioritariamente negativa para a maioria das instituições militares, mas permitiu obter indicadores específicos sobre a opinião pública e elaborar um conjunto de recomendações personalizadas ao contexto da Defesa.



# 5

## Conclusões

Para finalizar a dissertação apresentada, este capítulo visa apresentar as principais conclusões obtidas referentes à análise e teste da metodologia proposta. Assim, inicia-se com a revisão do objetivo em estudo, seguindo-se a resposta às questões de investigação colocadas inicialmente e a confirmação da aplicabilidade da solução encontrada. Por último, descreve as principais limitações do estudo e propõe um conjunto de sugestões para estudos futuros.

A motivação impulsionadora deste estudo centrava-se na falta de respostas face à necessidade de gerir a reputação em instituições do setor público. Para isso, pretendia-se uma solução acessível, prática e replicável que foi efetivada através do objetivo em estudo: desenvolver uma metodologia que extraia informação relevante sobre o eWOM em *social media*, usando estratégias de *text mining* e processamento de língua natural.

O resultado alcançado consistiu numa metodologia que inclui um conjunto de fontes de informação em *social media*, o desenvolvimento de um processo de extração semiautomática, e na aplicação de técnicas não convencionais capazes de extrair conhecimento relevante sobre o eWOM positivo, negativo e sobre as suas características. A solução proposta considerou as especificidades do setor público e traduziu-se numa contribuição que se distingue por ser inovadora, livre de custos, parcialmente automatizada, capaz de aceder a uma amostra extensa e preparada para funcionar perante as dificuldades e especificidades do setor.

A prova de conceito foi realizada para o contexto do Exército Português, uma instituição que enfrenta dificuldades de recrutamento e retenção que poderão influenciar ou resultar da sua reputação. Através do caso em estudo, foram respondidas as questões de investigação colocadas inicialmente e identificadas as potencialidades e as limitações deste tipo de metodologia.

Relativamente à primeira questão, confirmou-se que os social media possuem fontes de informação com potencial para conter informação sobre a reputação de um organismo público. Embora as instituições estatais recorram principalmente a métodos de recolha tradicionais (Purcarea et al., 2015; Le et al., 2019), ficou demonstrado que a técnica de *web scraping* conjugada com um conjunto de diretrizes de seleção possui capacidade para extrair informação com pouca interferência humana sobre o processo de recolha ou sujeitos

em estudo. A partir da informação extraída - e de acordo com a diversidade de fontes aconselhada na literatura (Muneta, 2017)-, ficou também demonstrado que o conjunto de fontes de informação no setor público deve conter fontes institucionais (páginas oficiais da instituição) e não institucionais (fontes noticiosas) presentes em redes sociais (Facebook), fóruns temáticos e comunidades virtuais (Reddit). No caso do Exército, foram selecionadas 19 fontes de informação com comentários realizados entre 2014 e novembro de 2019.

A segunda questão referiu-se ao subconjunto de fontes de informação que se aplica a um determinado organismo em estudo. Definir esse subconjunto requer uma exploração dos dados com ênfase nas métricas: tipo de informação contida, número de comentários e número médio de palavras por comentário. Para o caso em estudo, a fonte de informação Reddit revelou um bom desempenho, caracterizando-se por uma gestão independente e um número significativo de comentários e publicações com opiniões suficientemente extensas e diversificadas.

No que se refere à terceira questão, o conteúdo textual dos comentários revelou potencialidade para a aplicação de modelos de tópicos. Para os comentários em estudo, a técnica LSA foi aquela que apresentou melhores resultados e permitiu extrair 5 tópicos relacionados com características do Exército, com a utilidade das Forças Armadas, com a imagem da Marinha e Força Aérea, e com as percepções acerca da sua operacionalidade e da sua imagem negativa global.

Respondendo à última questão, foi possível detetar sentimentos nos comentários sobre o organismo público avaliado, replicando o sucesso desse tipo de análise no setor privado (Jansen et al., 2009). Os sentimentos nos comentários podem ser detetados de forma manual (através de avaliador) e de forma automatizada (com recurso a aprendizagem automática). Globalmente, esta análise revelou potencial para conhecer a reputação de uma instituição, os fatores que a influenciam ao longo do tempo e o conteúdo das opiniões públicas (positivas e negativas). A sua aplicação ao contexto do Exército veio confirmar que a valência do eWOM partilhado em social media pode fornecer indicadores relevantes, dos quais se destacam:

- A proporção de comentários com sentimentos negativos/ positivos poderá constituir um bom indicador da reputação institucional. Como previsto, a valência do eWOM e a frequência com que ocorre relaciona-se com as dificuldades registadas pelas instituições e permite distingui-las;
- O comportamento do eWOM ao longo do tempo fornece uma visualização dos efeitos produzidos por acontecimentos ou medidas concretas;
- Os termos que mais discriminam o eWOM com sentimento positivo do negativo;
- Os termos mais frequentes permitem compreender a percepção dos cidadãos acerca da instituição. No caso em estudo, foi possível obter indicadores sobre a opinião relacionada com utilidade pública, investimento público e imagem de empregador;

- A segmentação dos comentários por valência emocional pode fornecer informação sobre as características positivas e negativas da reputação institucional.

Relativamente à replicabilidade da metodologia, foi possível confirmar que esta possui flexibilidade e sistematização suficiente para funcionar em diferentes instituições públicas, conforme a sua aplicação ao caso de estudo demonstrou. Com exceção da Força Aérea Portuguesa, as Forças Armadas e os seus ramos possuem uma percentagem não negligenciável de comentários negativos (>60%) e que remetem para uma reputação congruente com a perda de efetivo militar nos últimos anos (Coelho et al., 2019). Globalmente, as instituições da Defesa possuem uma fraca imagem de utilidade pública, referente à perceção da relação custo/ benefício do erário público atribuído e à possibilidade da reintrodução do serviço militar obrigatório. Adicionalmente, observou-se que a confiança dos cidadãos nestas instituições diminuiu perante a divulgação de informação incongruente com os valores militares, nomeadamente através da divulgação de esquemas de corrupção, de exercícios militares pouco capacitados e da falta de resposta perante as necessidades da população. No que se refere à diferença entre instituições, o eWOM das Forças Armadas e do Exército permitiu detetar uma imagem de empregador negativa, em particular, na opinião negativa sobre as condições de trabalho oferecidas e défice de capital humano existente. Neste âmbito, o Exército revelou uma imagem particularmente negativa, sendo desaconselhado pelos cibernautas face aos restantes ramos e ao mercado de trabalho civil. No sentido oposto, a avaliação dos fatores de satisfação permitiu identificar que o acesso a oportunidades de formação e educação são os pontos fortes do Exército e da Força Aérea. Porém, o Exército é também positivamente caracterizado pela sua proximidade às atividades tidas como militares, enquanto que a Força Aérea, consegue melhorar a sua reputação com o apoio realizado à sociedade civil.

Considerando a dificuldade em controlar a informação partilhada online e o custo de reverter o seu efeito negativo (Kirby and Marsden, 2006), conclui-se que a metodologia desenvolvida é útil e aplicável à monitorização do eWOM dos organismos públicos. No que se refere ao contexto da Defesa Nacional, ficou demonstrado que esta poderá beneficiar de medidas interventivas no funcionamento organizacional e na imagem junto dos cidadãos portugueses em geral e, em particular, junto da sua população-alvo e referenciadores (familiares e amigos).

## **Limitações**

A metodologia utilizada possui como principal limitação a sua subjetividade. Os comentários em *social media* caracterizam-se por mensagens relativamente curtas, com diferentes objetivos (informativos ou opinativos), recurso a linguagem específica, utilização de ironia e, por vezes, com conversas secundárias à publicação inicial. Estas características aumentam a subjetividade da metodologia e respetiva interpretação de resultados. Concomitantemente, para as instituições do setor público, o número de comentários disponíveis

é reduzido e com uma distribuição assimétrica ao nível dos sentimentos. Estas limitações dificultam a avaliação e generalização do conhecimento extraído.

### **Sugestões**

De acordo com a aprendizagem proporcionada no desenvolvimento do trabalho, podem realizar-se um conjunto de sugestões futuras. Quanto ao contexto de aplicação, seria pertinente a sua replicação numa instituição do setor público fora do contexto da Defesa: estes resultados permitiriam validar a capacidade de generalização da metodologia ou melhorar as suas componentes. Relativamente à análise dos dados, sugere-se a classificação de um maior número de comentários, a qual permitiria refinar os resultados através da divisão de comentários por nível (principal ou secundário) ou análise de sentimento com diversas categorias de valência. Adicionalmente, estudar a relação entre publicações e respetivos comentários poderia produzir informação relevante para a tomada de decisão no âmbito do marketing institucional.

# Bibliografia

- Agarwal, A., Xie, B., Vovsha, I., Rambow, O., and Passonneau, R. (2011). Sentiment analysis of twitter data. In *Proceedings of the Workshop on Languages in Social Media, LSM '11*, page 30–38, USA. Association for Computational Linguistics.
- Aggarwal, C. C. and Zhai, C. (2012). *An Introduction to Text Mining*. Springer, Boston.
- Anagnostopoulou, S., Buhalis, D., Kountouri, I., Manousakis, E., and Tsekrekos, A. (2020). The impact of online reputation on hotel profitability. *International Journal of Contemporary Hospitality Management*, 32(1):20–39.
- Anderson, E. W. (1998). Customer satisfaction and word of mouth. *Journal of Service Research*, 1(1):5–17.
- Ash, S. (1951). Effects of group pressure upon the modification and distortion of judgments. In Guetzkow, H., editor, *Groups, leadership and men; research in human relations*, pages 177 – 190. Carnegie Press.
- Augusto, M. and Torres, P. (2018). Effects of brand attitude and ewom on consumers' willingness to pay in the banking industry: Mediating role of consumer-brand identification and brand equity. *Journal of Retailing and Consumer Services*, 42:1 – 10.
- Bandura, A. (1971). *Social learning theory*. General Learning Press.
- Banerjee, A. V. (1992). A simple model of herd behavior. *The Quarterly Journal of Economics*, 107(3):797 – 817.
- Barnett, M., Jermier, J., and Lafferty, B. (2006). Corporate reputation: The definitional landscape. *Corporate Reputation Review*, 9:538–541.
- Bengfort, B., Ojeda, T., and Bilbro, R. (2018). *Applied Text Analysis with Python: Enabling Language-Aware Data Products with Machine Learning*. O'Reilly Media, California.
- Cheung, C. M. and Thadani, D. R. (2012). The impact of electronic word-of-mouth communication: A literature analysis and integrative model. *Decision Support Systems*, 54(1):461 – 470.
- Christopher, M., Payne, A., and Ballantyne, D. (2002). *Relationship marketing: integrating quality, customer service and marketing*, pages 148 – 189. Butterworth-Heinemann, Oxford.

- Chu, S.-C. and Kim, Y. (2011). Determinants of consumer engagement in electronic word-of-mouth (ewom) in social networking sites. *International Journal of Advertising*, 30(1):47–75.
- Coelho, A., Reis, C., Borlinhas, V., Carreiras, H., Cardoso, A. I., Rodrigues, J., Tinoco, A., Reis, C., Vultos, D., Capela, M., Nogueira, J., and Nunes, P. (2019). Estudo de caracterização sociodemográfica e de satisfação organizacional dos militares do regime de voluntariado e de contrato dos três ramos das forças armadas. Technical report, Direção-Geral de Recursos da Defesa Nacional.
- Constantinides, E. (2014). Foundations of social media marketing. *Procedia - Social and Behavioral Sciences*, 148:40 – 57.
- Gurau, C. (2008). Integrated online marketing communication: implementation and management. *Journal of Communication Management*, 12(2):169–184.
- Hand, D. J. (1999). Statistics and data mining: Intersecting disciplines. *SIGKDD Explorations Newsletter*, 1(1):16–19.
- He, W., Wu, H., Yan, G., Akula, V., and Shen, J. (2015). A novel social media competitive analytics framework with sentiment benchmarks. *Information and Management*, 52(7):801 – 812.
- Hennig-Thurau, T., Gwinner, K. P., Walsh, G., and Gremler, D. D. (2004). Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the internet? *Journal of Interactive Marketing*, 18(1):38 – 52.
- Hevner, A. R., March, S. T., Park, J., and Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1):75–105.
- Hoye, G. and Lievens, F. (2007). Social influences on organizational attractiveness: Investigating if and when word of mouth matters. *Journal of Applied Social Psychology*, 37(9):38 – 52.
- Jalilvand, M. R., Esfahani, S. S., and Samiei, N. (2011). Electronic word-of-mouth: Challenges and opportunities. *Procedia Computer Science*, 3:42 – 46.
- Jansen, B. J., Zhang, M., Sobel, K., and Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth. *Journal of the American Society for Information Science and Technology*, 60(11):2169–2188.
- Jorlöv, K. and Venngren, E. (2014). Employer attractiveness of the swedish armed forces: A quantitative study of gender differences and the effects of employer brand attributes and social media in the government authority sector. Master’s thesis, Lulea University of Technology.

- Kaplan, A. M. and Haenlein, M. (2010). Users of the world, unite! the challenges and opportunities of social media. *Business Horizons*, 53(1):59 – 68.
- Kirby, J. and Marsden, P. (2006). *Connected Marketing, the viral, buzz and Word of Mouth Revolution*. Butterworth-Heinemann, Oxford.
- Ko, H., Cho, C.-H., and Roberts, M. S. (2005). Internet uses and gratifications: A structural equation model of interactive advertising. *Journal of Advertising*, 34(2):57–70.
- Kousheshi, M., Aali, S., Bafandeh Zendeh, A., and Iranzadeh, S. (2019). The antecedents and consequences of online relationship quality in internet purchases. *Journal of Islamic Marketing*, 11(1):161–178.
- Le, T., Dobele, A., and Robinson, L. (2019). Information sought by prospective students from social media electronic word-of-mouth during the university choice process. *Journal of Higher Education Policy and Management*, 41(1):18–34.
- Lee, M. and Youn, S. (2015). Electronic word of mouth (ewom): How ewom platforms influence consumer product judgement. *International Journal of Advertising*, 28(3):473–499.
- Lo, L. and Lin, S. (2017). An approach to initiating ewom intentions: A study on the effects of reference price and value homophily. *Internet Research*, 27(2):277–302.
- Luoma-aho, V. (2007). Neutral reputation and public sector organizations. *Corporate Reputation Review*, 10:124–143.
- Marktest, G. (2019). Análise sobre o comportamento dos portugueses nas redes sociais: Os portugueses e as redes sociais 2019.
- McHugh, M. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, 3(22):276–282.
- Muneta, M. H. (2017). Spread the word - the effect of word of mouth in e-marketing. In Siebert, G., von Rimscha, M. B., and Grubenmann, S., editors, *Commercial Communication in the Digital Age Information or Disinformation?*, pages 227–245. De Gruyter Saur.
- Paine, K. D. (2011). *Measure What Matters: Online Tools For Understanding Customers, Social Media, Engagement, and Key Relationships*. John Wiley and Sons, New Jersey.
- Peppers, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, 24(3):45–77.
- Peters, K., Chen, Y., Kaplan, A. M., Ognibeni, B., and Pauwels, K. (2013). Social media metrics — a framework and guidelines for managing social media. *Journal of Interactive Marketing*, 27(4):281 – 298.

- Poulis, A., Rizomyliotis, I., and Konstantoulaki, K. (2019). Do firms still need to be social? firm generated content in social media. *Information Technology and People*, 32(2):387–404.
- Purcarea, V., Gheorghe, I., and Gheorghe, C. (2015). Exploring the influential variables of ewom messages with a focus on health care services: A romanian case study. In *Proceedings of the 25th International Business Information Management Association Conference: Innovation Vision 2020: From Regional Development Sustainability to Global Economic Growth*, pages 2863–2867.
- Reis, C., Tinoco, A., Cardoso, A. I., Nogueira, J., and Nunes, P. (2019). Os jovens e as forças armadas. Technical report, Direção-Geral de Recursos da Defesa Nacional.
- Royo-Vela, M. and Casamassina, P. (2011). The influence of belonging to virtual brand communities on consumers' affective commitment, satisfaction and word-of-mouth advertising: The zara case. *Online Information Review*, 35:517–542.
- Saura, J., Reyes-Menendez, A., and Alvarez-Alonso, C. (2018). Do online comments affect environmental management? identifying factors related to environmental management and sustainability of hotels. *Sustainability*, 10(9).
- Schlagwein, D. and Hu, M. (2017). How and why organisations use social media: Five use types and their relation to absorptive capacity. *Journal of Information Technology*, 32(2):194–209.
- Serrat, O. (2017). *New-Age Branding and the Public Sector*, pages 375 – 383. Springer Singapore, Singapore.
- Silva, M., Eusébio, R., Reis, C., Guimarães, C., Andrade, M., and Azinhaes, J. (2019). Employer branding: the army mirrored on it's candidates. In *European Research Group on Military and Society*.
- Silva, M. J., Carvalho, P., and Sarmiento, L. (2012). Building a sentiment lexicon for social judgement mining. In Caseli, H., Villavicencio, A., Teixeira, A., and Perdigão, F., editors, *Computational Processing of the Portuguese Language*, pages 218–228, Berlin. Springer Berlin Heidelberg.
- Sterne, J. (2010). *Social Media Metrics: How to Measure and Optimize Your Marketing Investment*. John Wiley and Sons, New Jersey.
- Vargo, C., Gangadharbatla, H., and Hopp, T. (2019). eWOM across channels: comparing the impact of self-enhancement, positivity bias and vengeance on Facebook and Twitter. *International Journal of Advertising*, 38(8):1153–1172.
- Vázquez-Casielles, R., Suárez-Álvarez, L., and del Rio-Lanza, A.-B. (2013). The word of mouth dynamic: How positive (and negative) wom drives purchase probability. *Journal of Advertising Research*, 53(1):43–60.

- Whiting, A. and Williams, D. (2013). Why people use social media: a uses and gratifications approach. *Qualitative Market Research*, 16(4):362–369.
- Wirth, R. and Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. In *Proc. of the 4th international conference on the practical applications of knowledge discovery and data mining*, pages 29–39. Springer-Verlag London, UK.
- Zavattaro, S. M. (2018). What's in a symbol?: Big questions for place branding in public administration. *Public Administration Quarterly*, 42(1):90 – 119.