

FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO



***Customer Xperience – Análise de  
padrões de comportamento nas redes  
sociais para apoiar o setor do retalho***

**Ana Filipa Pereira da Cunha**

Mestrado Integrado em Engenharia Eletrotécnica e de Computadores

Orientador: Pedro João

16 de Julho de 2018



# Resumo

Atualmente existe uma inúmera quantidade de comentários, produzidos diariamente pelos utilizadores nas redes sociais, acerca dos mais variados tópicos. Estes dados, são uma oportunidade para as empresas acompanharem o que é dito nas redes sociais acerca da sua marca, bem como acerca de marcas concorrentes e perceberem futuras tendências do mercado. A utilização de ferramentas capazes de analisar e tratar os dados não estruturados, produzidos diariamente nas redes sociais, pode ser uma forma de apoio às empresas nas suas tomadas de decisão.

Considerando a possibilidade de integrar uma plataforma de análise das redes sociais, nas soluções já apresentadas aos seus clientes, a *Wipro* sugeriu o tema em que se inclui esta dissertação.

O presente projeto é direcionado ao caso específico de um cliente da *Wipro* que já mostrara interesse na implementação de uma ferramenta de análise das redes sociais nos seus sistemas de informação. Neste sentido, foram analisados os principais processos de negócio deste retalhista em específico, para os quais seria vantajoso trazer informações relevantes, a partir das redes sociais.

De forma a entregar dados com valor ao retalhista em questão, foi planeada a utilização da ferramenta *Data Discovery Platform, (DDP)* - ferramenta da *Wipro*. Esta ferramenta seria utilizada, essencialmente, para que fosse possível usufruir de duas das suas principais funcionalidades: a extração de dados presentes nas redes sociais e a análise de sentimentos dos dados extraídos.

Após a obtenção e armazenamento dos dados obtidos, bem como um primeiro tratamento dos mesmos, este projeto estaria terminado. É de salientar, que um tratamento mais profundo dos dados obtidos e a elaboração de gráficos, são da responsabilidade de duas outras dissertações já definidas e atribuídas a outros dois estudantes.

A utilização do *DDP*, principalmente devido a questões de confidencialidade, apenas foi possível por meio indireto. Deste modo, foram obtidos ficheiros *.csv*, incluindo dados recolhidos semanalmente, das páginas das redes sociais do retalhista em questão e dos seus competidores. Estes ficheiros foram carregados numa base de dados, alojada num servidor, para futuras análises. Desta forma, foi montado um processo manual, combinado com a equipa de *DDP*.

Devido ao facto de nunca ter sido permitido o acesso direto à ferramenta de *DDP* ao longo de todo este projeto, com a intenção de enriquecer o mesmo, recorreu-se ainda à utilização do *Buzzmonitor*. Utilizando a ferramenta *Buzzmonitor*, foi possível construir e configurar diferentes tipos de relatórios com informações relevantes para os vários processos de negócio do retalhista em questão. Para além de ter sido possível transformar os dados em forma de gráficos, através do *Buzzmonitor* foram ainda recolhidos os dados extraídos das redes sociais, em forma de ficheiros *.xls*. Assim, foi possível efetuar uma análise comparativa entre as duas ferramentas.



# Abstract

Currently there is a lot of comments, produced daily by users on social networks, about the most varied topics. This data is an opportunity for companies to follow what is said on social networks about their brand, as well as about competing brands and perceive future market trends. The use of tools capable of analyzing and treating the unstructured data produced daily in social networks can be a way of supporting companies in their decision making.

Considering the possibility of integrating a platform for analysis of social networks, in the solutions already presented to its clients, *Wipro* suggested the theme in which this dissertation is included.

This project is directed to the specific case of a textit *Wipro* client who has already shown interest in implementing a social network analysis tool in their information systems. In this sense, we analyzed the main business processes of this retailer in specific, for which it would be advantageous to bring relevant information, from the social networks.

In order to deliver data with value to the retailer in question, the use of the Data Discovery Platform tool, (*DDP*) - a *Wipro* tool - was planned. This tool would be used so that the extraction of the data present in the social networks as well as the analysis of the feelings of the same were possible. After obtaining and storing the obtained data, as well as a first treatment of data, this project would be finished. It should be noted that the treatment of the data obtained and the drawing of dashboards are the responsibility of two other dissertations already defined and attributed to two other students.

The use of *DDP*, mainly due to confidentiality issues, was only possible through indirect means. In this way, *.csv* files, including weekly collected data, were obtained from the retailer's pages in question and from their competitors. These files were uploaded into a database hosted on a server for future analysis. In this way, a manual process was assembled, combined with the *DDP* team.

Due to the fact that the direct access to the *DDP* tool was never allowed throughout this project, with the intention of enriching it, we also resorted to the use of *Buzzmonitor*. Using the *Buzzmonitor* tool, it was possible to build and configure different types of reports with information relevant to the various business processes of the retailer in question. Using *Buzzmonitor*, it was possible to transform data coming from social networks in the form of dashboards. In addition to this functionality, using *Buzzmonitor*, it was also possible to extract data from social networks in the form of *.xls* files and so it was possible to perform a comparative analysis between the two tools.



# Agradecimentos

Agradeço ao meu orientador, o professor Pedro João, por me ter apoiado ao longo de todo o projeto, especialmente nos momentos mais difíceis. Agradeço à professora Maria Antónia Carravilla que apesar de não estar diretamente relacionada com o meu projeto, esteve sempre disponível quando precisei de apoio.

Agradeço ainda à *Wipro* por me ter proporcionado a possibilidade de ter uma experiência em ambiente empresarial, enriquecendo-me profissional e pessoalmente. Agradeço especialmente à Susana Ribeiro por ter-me acompanhado ao longo deste projeto.

Acima de tudo, agradeço à base mais sólida e importante da minha vida, à minha família, por me ter proporcionado a oportunidade de ter estudado durante estes últimos cinco anos e por todo o apoio que me deram ao longo de toda a minha vida. Ao meu pai, à minha irmã e especialmente à minha mãe, agradeço por sempre ter acreditado em mim, por me ter motivado nos momentos mais difíceis e por ser o maior suporte da minha vida.

Agradeço ao meu namorado, Bruno Conceição, por ter marcado a minha passagem pela FEUP e acima de tudo por ter marcado a minha vida de uma forma tão positiva ao longo dos últimos anos. Agradeço por todo o companheirismo a nível pessoal e também profissional, agradeço pela companhia fantástica que é e por toda a motivação que me deu para nunca desistir.

À minha família, ao meu namorado e aos meus amigos, agradeço por todo o apoio e por todos os bons e simples momentos que fazem a vida valer a pena.

Ana Cunha





*“Everybody is a genius. But if you judge a fish by its ability to climb a tree, it will live its whole life believing that it is stupid.”*

Albert Einstein



# Conteúdo

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	<i>Wipro</i>	1
1.1.1	Wipro Portugal	1
1.1.2	Espírito da <i>Wipro</i>	2
1.2	Contextualização	3
1.3	Descrição do projeto	3
1.4	Objetivos	4
1.5	Estrutura da Dissertação	5
<b>2</b>	<b>Enquadramento Teórico</b>	<b>7</b>
2.1	Inteligência Artificial e <i>Machine Learning</i>	7
2.1.1	Sistemas de Inteligência Artificial e de <i>Machine Learning</i> nas áreas de Retalho	8
2.2	Análise das redes sociais	9
2.2.1	Principais Objetivos e Benefícios de uma Análise das Redes Sociais	9
2.2.2	Etapas da Análise das Redes Sociais	10
2.3	<i>Web Crawler</i>	13
2.3.1	Desafios de um <i>Web Crawler</i>	14
2.3.2	Limitações de um <i>Web Crawler</i>	14
2.3.3	Recolha de Informação	14
2.3.4	Escolha do Nó Inicial	14
2.3.5	Algoritmos de Pesquisa	15
2.3.6	Critério de Paragem	15
2.4	<i>Data mining</i>	16
2.4.1	Hierarquia do <i>data mining</i>	16
2.4.2	<i>Data mining e text mining</i>	17
2.4.3	<i>Opinion mining</i>	17
2.4.4	<i>Web Crawler e Data Mining</i>	18
2.5	Análise de Sentimentos	18
2.5.1	Etapas da Análise de Sentimentos	19
2.5.2	Classificação dos Sentimentos: Abordagens baseadas em Léxico e <i>Machine Learning</i>	21
2.6	Áreas de aplicação de Análise de Sentimentos e de <i>Opinion Mining</i>	23
2.6.1	Compras	23
2.6.2	Entretenimento	24
2.6.3	Política	24
2.6.4	Investigação e Desenvolvimento	25
2.6.5	Marketing	25

2.6.6	Educação . . . . .	25
<b>3</b>	<b>Evolução do Projeto e Ferramentas Utilizadas</b>	<b>27</b>
3.1	<i>Heuristics and Ontology-based Learning Machines and Exponential Systems (HOL-MES)</i> . . . . .	27
3.2	<i>Data Discovery Platform (DDP)</i> . . . . .	28
3.2.1	<i>DDP - Casos de Uso</i> . . . . .	29
3.3	<i>Buzzmonitor</i> . . . . .	30
3.3.1	Análise de Sentimentos . . . . .	31
<b>4</b>	<b>Retalhista A - Necessidade e Viabilidade da Adoção de uma Ferramenta de Análise das Redes Sociais</b>	<b>33</b>
4.1	<i>Retalhista A</i> . . . . .	33
4.1.1	Necessidade de Análise das Redes Sociais . . . . .	33
4.1.2	Análise <i>SWOT</i> . . . . .	34
4.2	<i>Retalhista A</i> - Processos de Negócio . . . . .	35
4.2.1	Mudança nas Tendências de Consumo - Previsão da Procura . . . . .	36
4.2.2	Performance das Lojas e Planeamento do Espaço . . . . .	38
4.2.3	Disponibilidade de produtos na Cadeia de Abastecimento . . . . .	39
4.2.4	Promoções e Preço Adequado para os Produtos . . . . .	40
4.2.5	Reconhecimento do Nome da Marca . . . . .	40
<b>5</b>	<b>Resultados</b>	<b>43</b>
5.1	Resultados obtidos com o <i>DDP</i> . . . . .	43
5.1.1	Descrição do Processo Manual Estabelecido . . . . .	43
5.2	Resultados obtidos com o <i>Buzzmonitor</i> . . . . .	45
5.2.1	Principais tópicos (acerca de moda) mais comentados e sentimentos prevalentes . . . . .	45
5.2.2	Adesão do público às promoções do <i>Retalhista A</i> e dos seus competidores . . . . .	46
5.2.3	Sentimentos mais comuns quando se comenta sobre o <i>Retalhista A</i> e sobre o <i>Concorrente 1</i> . . . . .	47
5.2.4	Principais críticas e elogios dados ao <i>Retalhista A</i> . . . . .	48
5.2.5	Principais reclamações e sugestões sobre uma determinada loja . . . . .	49
5.2.6	Críticas dos clientes relativamente à alocação do <i>staff</i> da loja . . . . .	50
5.2.7	Localidade dos utilizadores que mais comentam sobre a marca . . . . .	50
5.2.8	Reclamações ou elogios relativamente aos preços da marca do <i>Retalhista A</i> . . . . .	51
5.2.9	Redes sociais com maior adesão . . . . .	51
5.2.10	Adesão dos utilizadores por tipo de publicação . . . . .	52
5.2.11	Utilizadores com mais seguidores que mais falam acerca do <i>Retalhista A</i> . . . . .	52
5.2.12	Número de comentários, <i>likes</i> e partilhas por dia. . . . .	52
5.3	Análise de Resultados . . . . .	53
5.3.1	Análise dos Resultados Obtidos com o <i>DDP</i> . . . . .	53
5.3.2	Análise dos Resultados Obtidos com o <i>Buzzmonitor</i> . . . . .	55
5.3.3	Análise Comparativa entre os dados Extraídos pelas duas ferramentas . . . . .	57
<b>6</b>	<b>Conclusões e Trabalho Futuro</b>	<b>59</b>
6.1	Conclusões . . . . .	59
6.2	Trabalho Futuro . . . . .	60

<i>CONTEÚDO</i>	xi
<b>A Desenvolvimento do Projeto - Arquiteturas</b>	<b>63</b>
<b>B Relação entre os Processos de negócio do <i>Retalhista A</i></b>	<b>65</b>
<b>C Resultados obtidos com o <i>DDP</i></b>	<b>67</b>
<b>D Resultados obtidos com o <i>Buzzmonitor</i></b>	<b>69</b>
<b>Referências</b>	<b>81</b>



# Lista de Figuras

1.1	Logótipo da <i>Wipro</i> . . . . .	2
2.1	Etapas existentes na análise das redes sociais e respetivos desafios encontrados em cada uma dessas mesmas etapas (Stieglitz <i>et al.</i> , 2018, [28]) . . . . .	11
2.2	Algoritmos de pesquisa <i>BFE</i> e <i>DFE</i> , <a href="https://www.quora.com/What-are-the-differences-between-DFS-and-BFS">https://www.quora.com/What-are-the-differences-between-DFS-and-BFS</a> , acessado pela última vez em 2018-06-22, 17:03 . . . . .	15
2.3	Hierarquia do <i>data mining</i> (Srinivasa Rao, Satyanarayana Prasad e Venkateswara Rao, 2018, [4]) . . . . .	16
2.4	Etapas consideradas por Rajput e Solanki, (2016, [25]) na Análise de Sentimentos . . . . .	20
2.5	Técnicas de classificação de sentimentos segundo Medhat, Hassan e Korashy, (2014, [26]) . . . . .	21
2.6	Exemplo de uma opinião acerca de um produto adquirido, <a href="http://www.shein.co.uk/">http://www.shein.co.uk/</a> , acessado pela última vez em 2018-04-04, 14:56 . . . . .	24
4.1	Principais Processos de Negócio do <i>Retalhista A</i> . . . . .	36
5.1	Modelo <i>As Is</i> do Processo Manual Estabelecido . . . . .	44
5.2	Comentário classificado, erradamente, como positivo . . . . .	57
A.1	Ideia inicial do projeto, utilizando o HOLMES . . . . .	63
A.2	Alteração da arquitetura inicial . . . . .	64
B.1	Relação entre os processos de negócio do <i>Retalhista A</i> e dados a extrair para cada um dos processos de negócio . . . . .	65
C.1	Tipo de ficheiros recebidos pela equipa de <i>DDP</i> . . . . .	67
D.1	Tipo de ficheiros exportados a partir do <i>Buzzmonitor</i> . . . . .	69
D.2	<i>Hashtags</i> mais utilizadas pelos utilizadores . . . . .	70
D.3	Exemplo de comentários obtidos quando se clica na barra de <i>#fashion</i> do gráfico da Figura D.2 da página 70 . . . . .	70
D.4	<i>Hashtags</i> acerca do tema moda, mais usados pelos utilizadores . . . . .	70
D.5	Número de publicações, por dia, do <i>Retalhista A</i> (a laranja) e do <i>Concorrente 1</i> (a azul) . . . . .	71
D.6	<i>Engagement</i> dos utilizadores para com o <i>Retalhista A</i> e o <i>Concorrente 1</i> . . . . .	71
D.7	Análise de sentimentos expressos pelos utilizadores acerca do <i>Retalhista A</i> . . . . .	72
D.8	Exemplos de comentários acerca do <i>Retalhista A</i> . . . . .	73
D.9	Análise de sentimentos expressos pelos utilizadores acerca do <i>Concorrente 1</i> . . . . .	73
D.10	Exemplos de comentários acerca do <i>Concorrente 1</i> . . . . .	74

D.11 Assuntos mais comentados acerca do <i>Retalhista A</i> , desde 1 de janeiro de 2018 até 14 de maio de 2018 . . . . .	74
D.12 Assuntos mais comentados acerca das lojas do <i>Retalhista A</i> , desde 1 de janeiro de 2018 até 21 de maio de 2018 . . . . .	75
D.13 Comentários exemplo acerca das lojas do <i>Retalhista A</i> . . . . .	75
D.14 Análise geral acerca dos comentários relativos ao <i>staff</i> das lojas do <i>Retalhista A</i> .	76
D.15 Pequena amostra de comentários que é possível visualizar a partir da Figura D.14	76
D.16 Análise mais restrita acerca das críticas relativas ao <i>staff</i> das lojas do <i>Retalhista A</i>	76
D.17 Pequena amostra de comentários que é possível visualizar a partir da Figura D.16	77
D.18 Localização dos utilizadores que mais comentam sobre a marca do <i>Retalhista A</i> .	77
D.19 Reclamações ou elogios relativamente aos preços da marca do <i>Retalhista A</i> . . .	77
D.20 Sentimentos expressos pelos utilizadores relativamente aos preços da marca do <i>Retalhista A</i> . . . . .	78
D.21 Redes sociais com maior adesão dos utilizadores para o caso do <i>Retalhista A</i> . . .	78
D.22 Adesão por tipo de publicação . . . . .	78
D.23 Influenciadores que mais comentam acerca da marca <i>Retalhista A</i> , nas redes sociais	79
D.24 Número de comentários, por dia, na página do <i>Facebook</i> do <i>Retalhista A</i> . . . . .	79
D.25 Número de comentários, por dia, na página do <i>Facebook</i> do <i>Concorrente 1</i> . . . . .	79
D.26 Número de <i>likes</i> , por dia, na página do <i>Facebook</i> do <i>Retalhista A</i> . . . . .	80
D.27 Número de <i>likes</i> , por dia, na página do <i>Facebook</i> do <i>Concorrente 1</i> . . . . .	80
D.28 Número de partilhas das publicações do <i>Retalhista A</i> , por dia . . . . .	80
D.29 Número de partilhas das publicações do <i>Concorrente 1</i> , por dia . . . . .	80



# **Lista de Tabelas**



# Abreviaturas e Símbolos

<i>API</i>	<i>Application Programming Interface</i>
<i>BFE</i>	<i>Breadth First Exploration</i>
<i>DDP</i>	<i>Data Discovery Platform</i>
<i>DFE</i>	<i>Depth First Exploration</i>
<i>ERP</i>	<i>Enterprise Resource Planning</i>
<i>HOLMES</i>	<i>Heuristics and Ontology-based Learning Machines and Exponential Systems</i>
<i>HTML</i>	<i>HyperText Markup Language</i>
<i>IMDb</i>	<i>Internet Movie Database</i>
<i>NoSQL</i>	<i>Not Only SQL</i>
<i>PLN</i>	<i>Processamento de Linguagem Natural</i>
<i>RPA</i>	<i>Robotic Process Automation</i>
<i>Social CRM</i>	<i>Social Customer Relationship Management</i>
<i>SQL</i>	<i>Structured Query Language</i>
<i>TI</i>	<i>Tecnologias de Informação</i>
<i>URLs</i>	<i>Uniform Resource Locators</i>
<i>WinSCP</i>	<i>Windows Secure CoPy</i>



# Capítulo 1

## Introdução

O presente documento tem por objetivo descrever os vários processos da realização desta dissertação, em ambiente empresarial, na empresa *Wipro*, situada na Maia.

Neste primeiro capítulo será feita uma breve contextualização e descrição do tema desta dissertação, serão apresentados os objetivos deste projeto bem a organização deste documento. Adicionalmente, será efetuada a apresentação da instituição onde foi realizada esta dissertação.

### 1.1 *Wipro*

A *Wipro* foi criada em 1945 por Mohamed Premji. Nessa altura, a empresa começou por ser chamada de "*Western India Vegetable Products Limited*", sendo inicialmente uma empresa fabricante de óleos vegetais e refinados, situada na Índia. Mais tarde, no ano de 1966, após o falecimento de Mohamed Premji, o seu filho Azim Premji, de 21 anos, foi forçado a abandonar os seus estudos para ficar encarregue do negócio familiar.

Foi sob o comando de Azim Premji, que entre os anos 1970 e 1980, a empresa apostou em novas oportunidades de negócios, na indústria de TI, que nessa época estava numa fase embrionária na Índia.

Atualmente, a *Wipro* é uma empresa multinacional, líder em TI, consultoria e serviços de processos de negócios. É reconhecida globalmente devido ao seu abrangente portfólio de serviços, forte compromisso com a sustentabilidade e boa cidadania corporativa. Neste sentido, é de salientar o facto da empresa *Wipro* ter sido nomeada pelo *Ethisphere Institute*, em fevereiro de 2018, pelo sétimo ano consecutivo, como sendo a empresa mais ética do mundo, [40]. Promovendo um ambiente multicultural, neste momento, a empresa conta com mais de 160 mil funcionários e opera em mais de sessenta países.

Na Figura 1.1 da página 2 é possível observar-se o atual logótipo da *Wipro*.

#### 1.1.1 *Wipro Portugal*

A *Wipro Portugal* surge a partir da compra da empresa *Enabler*. A *Enabler* foi criada em 1997, como resultado da independência planeada do departamento de Sistemas de Informação da *Modelo*



Figura 1.1: Logótipo da *Wipro*

*Continente Hipermercados*, pertencente ao Grupo *Sonae*. Dois anos após a sua criação, a *Enabler* fez a sua primeira implementação internacional, em Itália, e fez ainda o co-desenvolvimento de recursos *Oracle Enterprise Resource Planning (ERP)* <sup>1</sup> para a *Sonae*. A empresa foi evoluindo, foram feitas implementações nos Estados Unidos da América e na Europa e foram abrindo novos escritórios nomeadamente em Braga, Reading, Paris, Dusseldorf, Curitiba e Minneapolis.

Devido ao elevado conhecimento e domínio na área de retalho e por ser implementador líder do pacote *Oracle Retail ERP*, em 2006, a *Wipro* adquiriu a *Enabler*.

No ano de 2008 foi inaugurado o escritório da *Wipro*, na Maia, que conta com mais de cento e oitenta consultores, com experiência em TI, muitos deles já provenientes da antecedente *Enabler*.

### 1.1.2 Espírito da *Wipro*

A essência da *Wipro* assenta nos seguintes valores :

- Ter como foco principal o sucesso dos clientes;
- Tratar cada pessoa com respeito e cultivar um ambiente onde as pessoas possam ser encorajadas a aprender, partilhar e crescer, promovendo-se a diversidade do pensamento, das culturas e das pessoas;
- Ter sempre presente integridade em tudo o que é feito;
- Ser global e responsável.

---

<sup>1</sup>Sistemas e pacotes de software usados pelas organizações (clientes) para gerir as suas atividades comerciais diárias, (contabilidade, compras, etc)

## 1.2 Contextualização

Com o surgimento e proliferação das redes sociais [24], foi notável uma significativa mudança na interação humana[29]. Assim, as relações sociais do mundo real foram migradas para o mundo virtual, resultando em comunidades *online* que reúnem pessoas de todo o mundo. Deste modo, é possível aos indivíduos compartilhar conhecimento e opiniões uns com os outros e promover diálogos entre diferentes culturas. Contudo, as redes sociais vieram alterar não só a forma como as pessoas se relacionam ou os hábitos que estas têm, vieram também alterar grande parte dos negócios e da economia do mundo. Hoje em dia raras são as empresas que não têm o seu *website*, a sua página de *Facebook* e que expandiram o seu negócio tornando-o *omichannel*<sup>2</sup>[23].

As redes sociais são um meio onde as interações e comentários dos clientes ajudam as empresas a entender o mercado e a ajustar os seus produtos e estratégias [34]. Para além disto, muitas empresas organizam concursos e premeiam os seus seguidores para os entusiasmar a visitar a sua página com mais frequência, isto porque, comparado com outras formas de marketing, a presença nas redes sociais é um meio barato e eficaz para melhorar a imagem e a popularidade da marca.

Da perspetiva do consumidor, as redes sociais são um meio através do qual podem procurar opiniões acerca de um serviço ou produto, antes de o adquirirem [34]. Assim, opiniões de pessoas qualificadas e influentes (e não só), são tidas em conta pelos consumidores, antes de efetuarem uma compra. Para além disto, os clientes procuram, nos dias de hoje, ter uma experiência de compra interessante, quer seja quando procuram informações sobre o artigo quer seja no ato de compra em si [34] [1].

Contudo, os dados presentes nas redes sociais são inúmeros e torna-se difícil para os retalhistas (e também para os consumidores) formar uma opinião geral sobre o que se tem comentado sobre uma determinada marca ou sobre determinados produtos. Neste sentido, torna-se fundamental o desenvolvimento de uma solução que possa extrair dados relevantes das redes sociais, de modo automático, para dar a conhecer ao retalhista o estado do mercado, as tendências e determinados padrões existentes, de forma resumida. Estas informações recolhidas a partir das redes sociais são interessantes e têm valor para o retalhista uma vez que o podem apoiar nas suas decisões de negócio.

Uma ferramenta de análise das redes sociais, que permita a sumarização de opiniões, sobre marcas e produtos, pode ser ainda interessante para os consumidores, que procurando opiniões sobre marcas e produtos se deparam com inúmeros comentários e opiniões diferentes acerca de uma mesma marca, produto ou serviço.

## 1.3 Descrição do projeto

Esta dissertação surgiu da partição do projeto *Customer Xperience*, inicialmente definido pela *Wipro* como um só projeto, destinado a uma só dissertação.

---

<sup>2</sup>Unificação dos vários canais de venda de uma empresa de modo a melhorar a experiência do cliente

Considerando a investigação e a carga de trabalho requerida para o projeto inicialmente arquitetado, e tendo ainda em conta o espaço de tempo disponível para o realizar, a decisão de direcionar o projeto para uma só dissertação foi repensada e o projeto inicial foi tripartido, originando três dissertações distintas.

A presente dissertação consiste no primeiro dos três blocos que definem o projeto *Customer Xperience*. Este primeiro bloco consiste na extração de dados a partir das redes sociais e a análise de sentimentos desses mesmos dados, utilizando, para isso, uma ferramenta configurável, da *Wipro*. Nesta dissertação, os dados extraídos deverão ser ainda armazenados numa base de dados de modo a ficarem acessíveis para o restante desenvolvimento do projeto. Ainda da responsabilidade deste primeiro bloco será a criação de um método automático para que os dados extraídos, das redes sociais, através da ferramenta da *Wipro*, sejam guardados automaticamente numa base de dados.

O segundo bloco do projeto *Customer Xperience* será tratado numa outra dissertação e consistirá no tratamento dos dados já extraídos das redes sociais. Nesta segundo bloco, será extraído todo o texto e informação relevante e será descartado todo o conteúdo cuja existência é desnecessária. Assim, nesta fase, os dados são devidamente preparados para que sejam uma entrada do terceiro bloco.

De forma a finalizar o projeto *Customer Xperience*, existirá uma terceira dissertação que irá tratar de transformar as informações provenientes do segundo bloco, em *dashboards*, de modo a permitir ao retalhista, visualizar os dados de forma organizada e conseqüentemente preparar tomadas de decisão. Analisando os dados recolhidos, o retalhista poderá tomar decisões relativamente ao planeamento do *layout* da sua loja, evidenciando determinados produtos e resguardando outros. Poderá planear preços, *stocks*, novas estratégias de marketing bem como o desenvolvimento de novos produtos e reconhecimento de pontos de melhoria.

É de salientar que o objetivo deste projeto, como um todo, é ser integrado nas soluções de *Oracle Retail* usadas pela *Wipro* em grande parte dos seus projetos. Para tal, a ferramenta de extração de dados, da *Wipro*, apenas será utilizada no primeiro bloco desta dissertação. O tratamento dos dados extraídos pelo *DDP* será efetuado utilizando outro tipo de ferramentas que possibilitem a personalização dos dados, de modo a que estes possam ser integrados nas diferentes soluções da *Wipro*.

## 1.4 Objetivos

Tendo em conta os três blocos descritos, o primeiro bloco tem como objetivos:

- Definir fontes de dados, (*Facebook, Instagram, Youtube, etc*), consoante as capacidades da ferramenta de extração da *Wipro*;
- Definir configurações para efetuar pedidos à ferramenta de extração de dados, da *Wipro*, (definir qual a área de retalho a incidir, as palavras-chave a procurar e a frequência de procura destas informações);



- Definir a melhor forma de armazenar os dados extraídos das redes sociais para futuro tratamento dos mesmos;
- Desenho e criação de um processo automático para que os dados extraídos, a partir da ferramenta da *Wipro*, possam ser automaticamente recebidos e guardados numa base de dados;

## 1.5 Estrutura da Dissertação

Para além da introdução, esta dissertação contém mais 5 capítulos. No capítulo 2, é feito um enquadramento teórico, descrevendo alguns dos conceitos associados aos objetivos desta dissertação.

No capítulo 3, é dado a conhecer, de uma forma mais detalhada, todo o desenvolvimento do projeto. Ainda neste capítulo são dadas a conhecer as ferramentas envolvidas no desenvolvimento desta dissertação.

O capítulo 4 mostra um caso particular, de um retalhista da *Wipro*, para o qual será direcionada esta solução. Ainda no capítulo 4 é feita uma análise *SWOT*, de forma a analisar a viabilidade da adoção de uma ferramenta de análise das redes sociais, para o caso do retalhista em questão, são descritos os processos de negócio desse mesmo retalhista, para os quais será necessário e interessante trazer informações relevantes, a partir das redes sociais.

No capítulo 5 mostram-se os resultados obtidos com duas ferramentas distintas e é feita uma análise dos resultados.

O capítulo 6 trata da conclusão e dos trabalhos futuros a ser realizados.



## Capítulo 2

# Enquadramento Teórico

Atualmente, o mundo é preenchido por grandes quantidades de dados. Exemplo disso são as inúmeras fotografias, músicas, vídeos e textos que estão presentes nas mais variadas plataformas *online*. Todos estes dados podem ser gerados por pessoas ou até por máquinas, e é previsto, que nos anos que se seguem, o seu crescimento não abrande [38]. Assim sendo, torna-se impossível gerir esta grande quantidade de dados manualmente, escrevendo e mudando as regras dos sistemas, adaptando-os ao crescimento e desenvolvimento dos dados. Passa a ser imprescindível desenvolver os sistemas para que possam aprender, automaticamente, a lidar com os dados, quer seja com a grande quantidade de dados quer seja com as mudanças que possam surgir nos mesmos. Neste capítulo, serão abordados conceitos fundamentais associados à recolha, tratamento e sumariação destes dados, de forma automática.

### 2.1 Inteligência Artificial e *Machine Learning*

A inteligência artificial consiste na simulação de inteligência humana por máquinas. Esta simulação traduz-se em processos que incluem *machine learning* (aquisição de informações e regras para usar as informações), o raciocínio (usando as regras para chegar a conclusões aproximadas ou definitivas) e a autocorreção, [37].

Embora não nos consigamos aperceber de imediato, atualmente, a inteligência artificial e os sistemas de *machine learning* estão cada vez mais presentes no nosso dia a dia [38].

Quando tiramos uma fotografia e o rosto das pessoas é reconhecido pela câmara, quando navegamos no *Youtube* e há um vídeo que nos é recomendado, baseado nas nossas pesquisas e gostos, ou quando efetuamos uma pesquisa no *Google*, estamos perante sistemas de *machine learning*.

Quando utilizamos o *Google* para efetuar uma simples pesquisa, há uma série de sistemas de *machine learning* que são responsáveis por perceber o que escrevemos, por encontrar resultados e ajustá-los às nossas áreas de interesse, decidindo em função disso que resultados mostrar primeiro.

Se pesquisarmos no *Google "Java"* <sup>1</sup>, dependendo se somos experientes em programação ou apreciadores de café, vão aparecer-nos diferentes resultados como resposta a esta simples pesquisa.

*Machine Learning* está presente em áreas como o reconhecimento de imagem, deteção de falhas, sistemas de recomendação, assim como está presente em áreas em que é necessário entender texto e discurso. Todas estas capacidades podem ser aplicadas a uma série de campos, desde a saúde, retalho e até transportes, (estacionamento e condução automática).

De acordo com Yufeng Guo [38] podemos definir *machine learning* como um sub campo da inteligência artificial que utiliza dados para responder a perguntas. Utilizar dados significa treinar os dados, para que numa fase posterior possam ser feitas previsões, ou seja, para que seja possível responder a perguntas.

Na fase de treino dos dados, são recolhidos os dados para treino, que serão a entrada de um modelo. Os dados de treino podem assentar em diversas características. Supondo que se pretende que a nossa máquina classifique se uma bebida é cerveja ou vinho, as características podem ser, por exemplo, a cor da bebida e a sua percentagem de álcool. Após terem sido recolhidas a cor de diversas bebidas e a respetiva percentagem de álcool (dados de treino), estes dados devem ser devidamente ajustados, em cada ciclo do treino, de modo a treinar devidamente o nosso modelo, até chegar-se à situação ideal de que há uma linha que separa os dados, classificando-os corretamente, neste caso em vinho ou cerveja. Quando se fizer uma pergunta ao modelo, dando como entrada, uma determinada cor e respetiva percentagem de álcool será esperado que o modelo classifique corretamente esses dados, (neste caso em vinho ou cerveja).

Utilizando sistemas de *machine learning*, é possível que as tarefas dos humanos sejam mais fáceis e rápidas. Em específico na área do retalho, partimos do princípio de que hoje em dia, qualquer empresa deve ter o seu próprio *website*, no futuro será esperado que qualquer empresa tenha associados aos seus produtos e serviços sistemas de *machine learning* [38].

### 2.1.1 Sistemas de Inteligência Artificial e de *Machine Learning* nas áreas de Retalho

A abertura da *Amazon Go*, no início de 2018, em Seattle, nos Estados Unidos da América, mostra o quão os sistemas de inteligência artificial já estão presentes nas áreas do retalho e de uma forma tão evoluída ([10], [5]).

A recente loja da *Amazon* é totalmente automatizada, não existem nem empregados na caixa, nem sequer uma caixa *self-service* para o cliente registar as suas contas. O processo de compra dos clientes é simplificado ao máximo, de modo que para adquirirem os seus produtos, basta que, à entrada da loja, passem o telemóvel (onde está instalada a *Amazon Go app*) por um sensor e, a partir desse momento, as câmaras instaladas na loja seguem todos os seus movimentos. À medida que o cliente vai recolhendo produtos das prateleiras, estes vão sendo adicionados à sua conta *Amazon*. Quando um produto é devolvido à prateleira, o sistema reconhece e elimina-o da conta. À saída da loja, é emitida uma fatura eletrónica.

---

<sup>1</sup>Conceito que pode ser relativo a uma marca de café ou a uma linguagem de programação

Em maior ou menor grau, para além da *Amazon*, outras empresas já começam a introduzir sistemas de inteligência artificial, (como por exemplo *chatbots*) nos seus produtos e serviços. Assim, a forma como as cadeias de retalho se relacionam com os consumidores será, cada vez mais diferente do habitual, mais do que isso, é previsto que no futuro haja uma transformação nos empregos, existentes nos dias de hoje. Paula Bellizia, presidente da *Microsoft* Brasil, apresenta uma postura positiva relativamente ao aparecimento e desenvolvimento de áreas de inteligência artificial:

"Como ocorreu nas outras revoluções tecnológicas, passaremos por um ciclo em que de fato alguns empregos deixarão de existir, mas outros serão criados, [17]".

## 2.2 Análise das redes sociais

Os meios de comunicação social como a televisão, a rádio ou os jornais, transmitem informações numa única direção, o que significa que não é possível ao público que tem acesso a essas informações interagir, com os meios de comunicação social, expressando as suas próprias opiniões acerca das informações transmitidas.

Contudo, os meios digitais, como as redes sociais, blogues ou fóruns tornam possível uma comunicação bidirecional que permite que os indivíduos interajam entre si e relativamente às informações transmitidas nas redes sociais, nos blogues ou nos fóruns *online*.

Este tipo de comunicação bidirecional, através das várias plataformas interativas *online*, criou oportunidades para muitas empresas que podem, deste modo, comunicar diretamente com os seus clientes.

### 2.2.1 Principais Objetivos e Benefícios de uma Análise das Redes Sociais

A análise das redes sociais consiste na análise dos dados presentes nas várias plataformas interativas *online*, com o objetivo de com esses dados tomar decisões de negócios. Para tal, é necessário converter os dados qualitativos, recolhidos das várias plataformas, em dados quantitativos, recorrendo-se, para esse efeito, à ajuda de diferentes técnicas de *data mining* e Processamento de Linguagem Natural (PLN).

As recomendações "boa-a-boca" de antigamente, hoje em dia migraram para as redes sociais, para os fóruns *online* e para tantas outras plataformas interativas. Assim, tal como antigamente, as empresas dependem destas recomendações, mas hoje em dia, quando algum conteúdo é postado numa rede social tem um alcance muito maior do que o alcance tido antigamente pelas recomendações "boca-a-boca". Considerando o alcance dos comentários efetuados por via das redes sociais, torna-se uma oportunidade para as empresas estar a par do que é dito sobre si na *Web*, de forma a definir ou redefinir as suas estratégias de negócios, e para além disso, estar presente nesse mundo, interagindo com os seus clientes.

Os principais objetivos e benefícios das análises das redes sociais incluem:

- Ter conhecimento das principais tendências e tópicos da atualidade que mais interesse despertam no público e quais os assuntos menos interessantes para os mesmos;
- Ter uma percepção do sentimento do público relativamente à marca própria e relativamente a marcas ou produtos concorrentes;
- Perceber quais os principais influenciadores dentro de uma determinada área de negócio;
- Perceber quais os competidores da marca e saber quando surgem novas marcas ou produtos concorrentes;
- Identificar áreas no processo de negócios ou produtos que precisam de ser desenvolvidos ou melhorados;
- Entender o caminho que leva à “fidelização de clientes”.
- Construir uma comunidade de embaixadores da marca;
- Determinar áreas que poderiam ser uma “ameaça” para a empresa;
- Avaliar a eficácia de diferentes campanhas de marketing nas redes sociais;
- Melhorar características de um produto ou serviço baseado nas opiniões dos consumidores;
- Encontrar oportunidades de vendas.

### 2.2.2 Etapas da Análise das Redes Sociais

As várias plataformas de redes sociais oferecem muitas possibilidades de formatos de dados, incluindo dados de texto, imagens, vídeos, som e até dados de geolocalização. Estes dados podem ser classificados em dados estruturados e não estruturados. Podemos considerar que o conteúdo textual é um exemplo de dados não estruturados, enquanto que as relações estabelecidas entre amigos são um exemplo de dados estruturados.

No que diz respeito à análise das redes sociais, esta consiste em várias etapas, das quais a análise de dados é apenas uma dessas etapas. Antes de serem analisados, os dados precisam de ser descobertos, coletados e preparados (Stieglitz *et al.*, 2018, [28]).

A análise de redes sociais pode ser entendida como “*um emergente campo de pesquisa interdisciplinar que visa combinar, ampliar e adaptar métodos de análise de dados provenientes das redes sociais*” (Stieglitz *et al.*, 2018, [28]). Tendo em consideração todas as etapas existentes aquando da realização de uma análise às redes sociais, ter uma visão geral dos desafios existentes, é uma mais valia, para que seja possível gerir a complexidade da realização destas análises, isto é, é necessário entender que desafios se têm de enfrentar aquando da descoberta, coleta e preparação dos dados para análises posteriores.

De acordo com Stieglitz *et al.*, (2018, [28]) o processo de análise das redes sociais assenta nos quatro passos mostrados na Figura 2.1. O primeiro passo consiste na descoberta de padrões

de interesse, extraindo-se de forma automática os tópicos relevantes a partir de um determinado conjunto de dados. Na fase seguinte à descoberta, existe a fase de *tracking*, onde são decididas quais serão as fonte de dados (por exemplo, *Twitter*, *Facebook*), onde armazenar os dados e quais as abordagens e métodos a seguir. Seguidamente à fase de *tracking* existe a fase de preparação e tratamento dos dados para que, posteriormente, possam ser analisados. Dependendo do propósito, de acordo com a Figura 2.1 da página 11, existem vários métodos disponíveis para análise de dados.

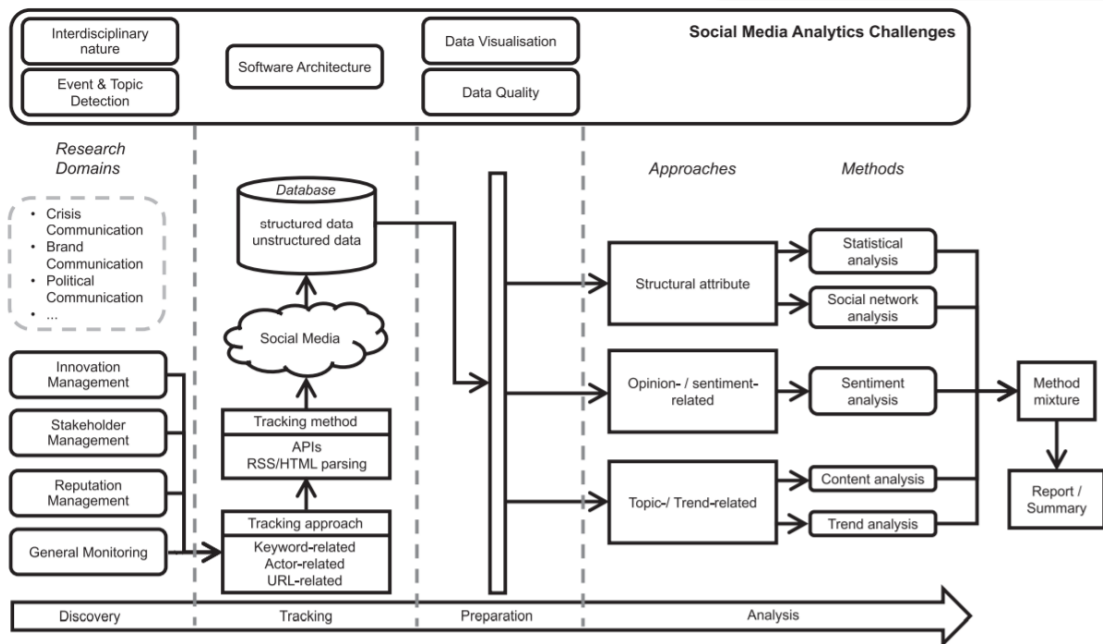


Figura 2.1: Etapas existentes na análise das redes sociais e respetivos desafios encontrados em cada uma dessas mesmas etapas (Stieglitz *et al.*, 2018, [28])

Segundo Stieglitz *et al.*, (2018, [28]) pode-se argumentar que os dados das redes sociais compartilham muitas características com os *big data*. Por serem dados em grandes quantidades e provenientes de redes sociais tem-se utilizado o termo *social media big data* para designar estes tipo de dados.

A análise de *big data* difere das análises tradicionais devido às seguintes razões:

- **Volume:** os *big data* ocupam muito espaço e estão sempre a surgir novos dados, todos os dias, especialmente no caso das redes sociais. Por este motivo são necessários *hardware* e *software* escaláveis para que seja possível acompanhar este constante surgimento de novos dados;
- **Velocidade:** o surgimento de novos dados ocorre a uma velocidade elevada e responder a este crescimento é um desafio;
- **Variedade:** os dados assumem diferentes formas (dados de texto, vídeos, imagens, etc).

- Veracidade: existe incerteza no que diz respeito à qualidade dos dados, especialmente no caso das redes sociais.

Na fase de descoberta de dados, como mostra na Figura 2.1 da página 11, os desafios mais comuns são relativos à natureza interdisciplinar dos dados e à dificuldade na descoberta de padrões relevantes.

O grande volume de dados torna difícil a descoberta de tópicos relevantes, tendências e eventos na tão dinâmica comunicação existente nas redes sociais.

O facto dos dados provenientes das redes sociais serem dados com uma natureza interdisciplinar, leva a que esses mesmos dados sejam analisados por profissionais de diferentes áreas. Neste sentido, é de salientar a lacuna existente entre as ciências sociais e as ciências de computação (Stieglitz *et al.*, 2018, [28]). Esta lacuna torna-se evidente ao longo de todo o processo de pesquisa, pois os profissionais das áreas de ciências sociais não possuem métodos à sua disposição, necessários para descobrir, coletar e preparar dados relevantes provenientes das redes sociais. Por outro lado, muitos dos profissionais das áreas de ciências de computação, que aplicam abordagens computacionais poderiam beneficiar de uma base mais sólida nas suas abordagens, tirando partido da teoria das ciências sociais.

Na fase de *tracking*, e com mostra a Figura 2.1 da página 11, o principal desafio encontrado é a escolha de uma arquitetura de *software* e tecnologia de armazenamento apropriadas.

O volume e a velocidade dos dados tornam necessária a escolha de arquiteturas de *software* apropriadas para a fase de coleta de dados. Em configurações convencionais que lidam com pequenas quantidades de dados, uma única máquina, ou um pequeno grupo de máquinas, executa um sistema de gestão de base de dados relacional que implementa o padrão *Structured Query Language (SQL)* (Stieglitz *et al.*, 2018, [28]). Perante *big data*, este tipo de soluções torna-se insuficiente.

O constante crescimento dos *big data* e a sua natureza não estruturada torna necessária a utilização de bases de dados *Not Only SQL, (NoSQL)*, que ao contrário das bases de dados *SQL*, são capazes de lidar com dados não estruturados e com grandes quantidades dos mesmos (Ghotiya, Mandal e Kandasamy, 2017, [19]).

Tal como é possível observar ainda na Figura 2.1 da página 11, na fase de preparação dos dados, os maiores desafios têm a ver com a obtenção de dados com qualidade e a visualização dos mesmos.

No que diz respeito à qualidade, a veracidade dos dados leva a problemas na etapa de preparação dos mesmos. Os dados obtidos a partir das redes sociais são geralmente incompletos, com grande quantidade de informação irrelevante e, por vezes, não confiáveis. Para além disto, pode haver informação em falta, porque o utilizador escolheu não fornecer-la, ou porque o custo financeiro ou computacional é alto demais para coletá-la efetivamente (Stieglitz *et al.*, 2018, [28]).

A visualização de dados na fase de preparação é dificultada devido ao volume e à variedade de dados. As visualizações podem ser cruciais quando as decisões precisam de ser tomadas rapidamente.



## 2.3 Web Crawler

Um *Web Crawler* é um programa que funciona de forma automática, navegando pela Internet de maneira sistemática [36]. Assim, um processo de *Web Crawling* consiste em examinar as palavras-chave nas páginas *Web*, o tipo de conteúdo de cada página e os *links*, antes de retornar as informações ao mecanismo de pesquisa [36].

Ao digitarmos uma consulta de pesquisa, esses *Web Crawlers* verificam todas as páginas *Web* relevantes, (que contêm as palavras que pesquisarmos) e transformam-nas num índice enorme[36].

O *Google* é um dos motores de pesquisa mais conhecidos e utilizados do mundo, como em qualquer outro motor de pesquisa, estão presentes, por trás do funcionamento do *Google*, conceitos de *Web Crawling*. As pesquisas que efetuamos no *Google* não são pesquisas efetuadas realmente na *Web*, são pesquisas efetuadas no índice da *Web* do *Google* [16]. O *Googlebot* (também conhecido como *robot*, *bot*, ou *spider*) [15], começa por obter algumas páginas *Web*, depois segue os *links* dessas páginas, obtendo mais *links* que apontam para novas páginas, e assim sucessivamente até que seja indexada uma boa parte da *Web*, ou seja, quando forem indexadas vários milhões de páginas armazenadas em milhares de máquinas [16]. Assim, quando pesquisamos no *Google*, por exemplo : “velocidade da chita”, o *Googlebot* pesquisa no índice para localizar todas as páginas que incluem os termos de pesquisa e neste caso aparecem centenas de milhares de resultados possíveis [16].

O *Google* decide que documentos mostrar para uma pesquisa específica, fazendo centenas de perguntas do tipo: quantas vezes esta página contém as palavras-chave? As palavras-chave aparecem no título? A página tem sinónimos das palavras-chave? Este *Web site* é de boa ou má qualidade? Qual a classificação da página? [16]

No que se refere à classificação das páginas *Web*, existe uma fórmula inventada pelos fundadores da *Google*, Sergey Brin e Larry Page, que avalia a importância de uma página *Web* através do número de *links* externos que remetem para ela e a relevância desses *links* [16].

Por fim, combinam-se todos esses fatores para gerar a classificação geral de cada página e devolver resultados de pesquisa em menos de um segundo após ter-se clicado no *enter* [16]. É de destacar ainda que os *Web Crawlers* são configurados para monitorizar a *Web* regularmente de modo a que os resultados gerados sejam constantemente atualizados [36].

O mesmo conceito de *Web Crawler* aplica-se ao caso específico da extração e pesquisa de dados nas redes sociais (Wong *et al.*, 2014, [13]).

No caso de uma rede social, um *Web Crawler* começa com uma lista de páginas de utilizadores a visitar. À medida que o *Web Crawler* visita essas páginas, ele identifica todas as páginas de amigos incorporadas na página atual e, em seguida, segue as páginas identificadas apenas para descobrir mais páginas. Este processo continua a ser executado até que seja atingido um determinado critério, (Wong *et al.*, 2014, [13]).

### 2.3.1 Desafios de um *Web Crawler*

De acordo com Wong *et al.*, (2014, [13]) quando se tenta aplicar *Web Crawlers* a redes sociais há alguns desafios presentes. Alguns dos desafios apontados por Wong *et al.*, (2014, [13]) são a complexidade das tecnologias da *Web*, por exemplo, *JavaScript* e *Ajax*, que dificultam a interpretação do conteúdo de uma página.

Outra das dificuldades encontradas foram as restrições de acesso da maioria das redes sociais. Os requisitos de *login*, a visualização limitada e os limites de consulta das *Application programming interfaces (API)* são exemplos de restrições de acesso apontadas por Wong *et al.*, (2014, [13]) . As políticas de controlo de privacidade também não permitem que um *Web Crawler* tenha acesso a toda a rede social.

### 2.3.2 Limitações de um *Web Crawler*

A estrutura de uma rede social pode ser modelada como um grafo  $G$ , tal que:  $G=(V,E)$ , onde  $V$  é um conjunto de nós (utilizadores) e  $E$  é um conjunto de arestas (amizades). No *Facebook*, todo o  $G$  não é acessível para todos, devido ao controlo de privacidade existente, isto é, os utilizadores do *Facebook* podem efetuar determinadas configurações na sua conta, permitindo que determinadas informações sejam apenas exibidas a pessoas autorizadas (por exemplo, a sua rede de amigos). Deste modo, excluindo os amigos da rede do utilizador mais ninguém conseguirá visualizar essas informações (incluindo um *Crawler*). Portanto, em geral, só se pode ter acesso à parte pública de  $G$ , denotada como  $G'$ . Portanto, é preciso entender que somente o  $G'$  pode ser rastreado de maneira legítima.

### 2.3.3 Recolha de Informação

Segundo Wong *et al.* (2014, [13]) antes de se pensar em procurar dados nas redes sociais é importante, numa primeira fase, identificar quais as fontes nas quais queremos procurar os dados e que dados pretendemos procurar. Isto porque, o *design* de um *Crawler* depende do tipo de dados que queremos procurar. O tipo de dados a serem procurados afetam ainda a condição de paragem do processo de procura (Wong *et al.*, 2014, [13]). Portanto, saber os dados que queremos procurar e onde procura-los permite otimizar o algoritmo de um *Crawler*.

### 2.3.4 Escolha do Nó Inicial

De acordo com Wong *et al.* (2014, [13]), escolher como nó inicial os utilizadores com muitas ligações para outros utilizadores possibilita a aceleração de todo o processo de procura. Estes utilizadores a escolher para nó inicial, podem ser identificados pelo número de amigos que possuem. No caso de ser necessário selecionar os utilizadores de uma determinada região é possível identifica-los usando as funções de pesquisa já existentes nas redes sociais. Por exemplo, o *Facebook* possui um recurso para mostrar um número de utilizadores selecionados aleatoriamente de

uma determinada rede regional. O resultado desta pesquisa pode ser um ponto de partida para o *Crawler* a construir.

### 2.3.5 Algoritmos de Pesquisa

O processo de *crawling* do gráfico social começa com um nó inicial e a exploração é dada iterativamente. Em cada uma das iterações, é visitado um nó e são descobertos os seus vizinhos diretos. De acordo com Wong *et al.* (2014, [13]) dependendo de qual será o vizinho seguinte a visitar, existem duas abordagens diferentes de pesquisa. *Breadth First Exploration (BFE)* é uma abordagem que inicia no nó inicial e explora todos os seus vizinhos. Posteriormente, para cada um dos nós não visitados explora seus vizinhos. Outra abordagem existente é o *Depth First Exploration, (DFE)* que começa no nó inicial e a exploração dá-se ao longo de cada ramo, tanto quanto possível antes de existir qualquer retorno.

Na Figura 2.2 é possível observar a diferença existente entre estes dois algoritmos. Na abordagem *BFE* a pesquisa é feita por camadas, começando no nó inicial, neste caso o elemento 0, seguem-se os elementos 1 e 2 e por último os elementos da última camada, 3, 4, 5 e 6. Na abordagem *DFE* a pesquisa começa igualmente pelo nó inicial, neste caso o elemento 0, seguem-se os elementos 1, 3, 4 e só depois se passa para o outro ramo e seguem-se na pesquisa os elementos 2, 5 e 6 [14].

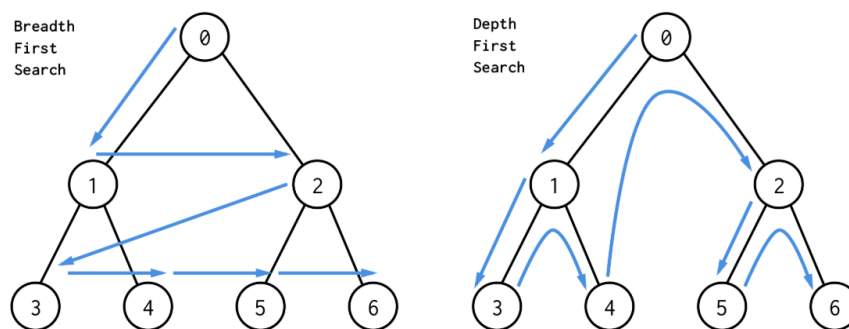


Figura 2.2: Algoritmos de pesquisa *BFE* e *DFE*, <https://www.quora.com/What-are-the-differences-between-DFS-and-BFS>, acessado pela última vez em 2018-06-22, 17:03

Comparando estes dois algoritmos em termos de desempenho, foi provado que o *BFE* é mais rápido que o *DFE* [14].

### 2.3.6 Critério de Paragem

Normalmente, o processo para quando o número de amostras é suficiente para a pesquisa em questão ou quando o resultado das amostras é saturado. Existem várias maneiras de avaliar se o processo de *crawling* foi bem sucedido. Uma das formas de medir se o processo foi bem sucedido é analisar se o número de amostras coletadas foi suficiente para o caso em estudo.

## 2.4 Data mining

Desde que a informática surgiu nas nossas vidas, elevados volumes de informação têm sido sistematicamente coletados e armazenados. O simples armazenamento e recuperação desta informação já traz um grande benefício face à procura de informação em volumosos e ineficazes arquivos em papel, existente no passado.

Contudo, apenas recuperar informação não propicia todas as vantagens possíveis. O processo de *data mining* permite que se investigue dados à procura de padrões que tenham valor, [20].

Dado o crescente volume de dados estruturados e não estruturados, produzidos pelas diversas vias *online*, do conceito de *data mining* surgiram outro tipo de conceitos.

### 2.4.1 Hierarquia do *data mining*

Como referido no início deste capítulo, a enorme quantidade de dados de texto, que são produzidos a partir de diversas fontes, como redes sociais, áreas da saúde, (como é o caso dos registos dos pacientes) e até notícias *online*, leva à necessidade do desenvolvimento de sistemas para automatizar a análise, sumarização e classificação de dados.

No desenvolvimento destes sistemas automatizados estão envolvidas áreas como a estatística, *machine learning* e inteligência artificial.

O conceito de *data mining* tal como os conceitos de si derivados, presentes na Figura 2.3, surgem com o objetivo de explorar grandes quantidades de dados de modo a que nesses mesmos dados sejam encontrados padrões relevantes.



Figura 2.3: Hierarquia do *data mining* (Srinivasa Rao, Satyanarayana Prasad e Venkateswara Rao, 2018, [4])

Nas secções seguintes serão abordados, em maior pormenor, os conceitos derivados de *data mining*, com maior interesse para esta dissertação.

### 2.4.2 Data mining e text mining

A área de *text mining* tem vindo a tornar-se cada vez mais popular devido à necessidade de sistemas e técnicas automáticas, responsáveis por extrair informação relevante a partir de grandes quantidades de dados.

Madni, Anwar e Shah (2017, [12]) definem *text mining* como sendo uma técnica de *data mining*. Observando a Figura 2.3 da página 16 é possível constatar, a partir da forma como estão representados, que o conceito de *text mining* deriva do conceito de *data mining*.

*Data mining* é considerado um sub campo da ciência de computação, que de acordo com koti (2013, [11]), pode ser definido como um processo de exploração e análise, por meios automáticos ou semiautomáticos, de grandes quantidades de dados, com o objetivo de descobrir padrões significativos.

A principal diferença existente entre *data mining* e *text mining*, uma técnica específica de *data mining*, é que enquanto *data mining* tem como objetivo descobrir padrões a partir de texto estruturado, extraindo conhecimento útil a partir do mesmo [32], *text mining* tem por objetivo descobrir padrões e extrair conhecimento a partir de texto semi-estruturado ou não estruturado [32]. No caso do *data mining* pressupõe-se que a informação já está numa forma desejável para extração. Já no caso do *text mining* a informação está na forma de linguagem natural, livremente escrito. Essa característica dá ao *text mining* uma natureza incerta. Assim, o *text mining* é um campo multidisciplinar baseado na recuperação de informação, *data mining*, *machine learning*, estatística, e linguística computacional (Talib *et al.*, 2016, [31]).

É importante referir que *data mining* e *text mining* não são abordagens nem técnicas únicas para a descoberta de conhecimento a partir de dados, conforme o caso em questão, existe uma variedade de métodos e tecnologias existentes que é possível aplicar [32].

### 2.4.3 Opinion mining

De acordo com Talele MECSE e Chandrashekhkar Badgujar, (2012, [33]) podemos classificar texto em duas categorias principais, factos e opiniões. Os factos são geralmente declarações objetivas sobre entidades e eventos. As opiniões são definidas como declarações subjetivas que refletem sentimentos ou perceções sobre as entidades e eventos.

*Opinion mining* é considerada uma área baseada em *text mining*, responsável por extrair opiniões expressas em texto escrito em linguagem natural [30].

Dado o crescente volume e valor das opiniões presentes na Web, surge o conceito de *opinion mining*. Observando a Figura 2.3 da página 16 verifica-se que o conceito de *opinion mining* manifesta-se no seguimento de *Web content mining*, que é uma sub-área de *Web mining*, responsável por extrair informações úteis a partir do conteúdo de páginas e documentos da Web, sítios onde hoje em dia são expressas inúmeras opiniões acerca de variados tópicos.

O facto de que as pessoas recorrem cada vez mais a meios como as redes sociais, o *Youtube* e a *blogs* quer seja para expressar a sua opinião quer seja para procurar opiniões acerca dos mais variados tópicos, de forma a suportarem uma tomada de decisão faz com que o interesse em *opinion*

*mining* tenha vindo a crescer progressivamente. Assim, torna-se útil a existência de ferramentas capazes de resumir qual a opinião prevaiente acerca de determinado assunto.

No caso das empresas, o interesse pela sumarização de opiniões, surge porque, sabendo, de forma resumida, qual a opinião que os consumidores têm acerca da sua marca, dos seus produtos ou serviços e sabendo ainda a opinião dos consumidores acerca de marcas concorrentes, as empresas podem elaborar estratégias, melhorar a sua eficiência e eficácia, tornando-se cada vez mais competitivas.

No entanto, os comentários e opiniões acerca de um produto podem chegar a ser centenas ou mesmo milhares, no caso de um produto popular. Isto dificulta que um potencial cliente consiga ler cada comentário de forma a tomar uma decisão de compra informada. Para além disso torna-se também difícil para o fabricante do produto acompanhar as opiniões dos clientes acerca do mesmo.

Portanto, neste caso específico, um resumo das opiniões, para além de ter interesse para as empresas é também bastante útil para os indivíduos em particular, quando procuram por uma opinião resumida e não tendenciosa de entre as tantas opiniões disponíveis, acerca de um serviço ou produto que pretendem adquirir.

*Opinion mining* é muitas vezes confundida com análise de sentimentos - um antecessor de *opinion mining* - que analisa como as pessoas se sentem relativamente a um determinado tópico. Contudo, *opinion mining* vai mais além, tentando entender os motivos pelos quais as pessoas se sentem de uma determinada forma [39].

#### **2.4.4 Web Crawler e Data Mining**

Tendo em mente os conceitos de *Web Crawler* e de *Data Mining* é possível, nesta fase, entender-se que estes dois conceitos trabalham em conjunto [36].

Depois dos dados serem coletados, a partir de várias fontes, pelo *Web Crawler*, continuarão numa forma não estruturada [36].

Podemos concluir que o processo de *Web Crawling* é o primeiro passo no processo de *data mining*. Os conceitos de *data mining* bem como os conceitos de si derivados, vêm à tona durante o processo de extração dos dados (*Web crawling*).

### **2.5 Análise de Sentimentos**

De acordo com Vohra e Teraiya, (2013, [3]) o surgimento de grandes quantidades de dados provenientes das redes sociais fez surgir um novo campo de pesquisa que pode denominar-se *opinion mining* ou análise de sentimentos.

*Opinion mining* e análise de sentimentos são estudos computacionais acerca das opiniões, emoções e atitudes das pessoas relativamente a uma determinada entidade.

Segundo Medhat, Hassan e Korashy, (2014, [26]) apesar de ser considerado que expressam um significado mútuo, há alguns autores que ainda assim distinguem análise de sentimentos de *opinion mining*, considerando que *opinion mining* trata de extrair e analisar a opinião das pessoas

acerca de uma entidade, enquanto análise de sentimentos trata de identificar e analisar o sentimento expresso num texto.

Vohra e Teraiya, (2013, [3]) definem a análise de sentimentos como a extração automatizada de atitudes, opiniões e emoções a partir de texto, discurso ou a partir de bases de dados, por meio de PLN. A análise do sentimento envolve a classificação de opiniões em categorias como "positivo", "negativo" ou "neutro".

De acordo com Medhat, Hassan e Korashy, (2014, [26]), há três tipos de classificação de sentimentos: classificação ao nível do documento, ao nível da frase e ao nível do aspeto.

A classificação ao nível do documento, visa classificar um documento como expressando uma opinião ou sentimento positivo ou negativo relativamente a uma determinada entidade. A classificação de sentimentos ao nível de frases tem como objetivo classificar o sentimento expresso em cada frase, sendo que primeiramente se identifica se a frase é subjetiva ou objetiva. Se a frase for subjetiva, será determinado se a frase expressa opiniões positivas ou negativas. Contudo, não há diferenças significativas entre as classificações de nível de documento e frase, porque as frases não são mais do que documentos curtos.

No entanto, classificar texto ao nível do documento ou ao nível da frase muitas vezes não fornece as informações necessárias sobre todos os aspetos relativos a uma determinada entidade, que são necessários em muitas situações. Assim sendo, surge a classificação ao nível do aspeto, cujo objetivo consiste em classificar o sentimento relativamente a aspetos específicos alusivos às entidades em questão. Neste caso, é relevante que sejam identificadas as entidades e os seus aspetos numa fase antecedente à classificação do sentimento.

### 2.5.1 Etapas da Análise de Sentimentos

Na Figura 2.4 da página 20, são mostradas as várias etapas existentes na análise de sentimentos, consideradas por Medhat, Hassan e Korashy, (2014,[26]).

A análise de sentimentos começa com uma primeira fase em que é efetuada a extração dos dados a analisar. No caso específico desta dissertação, os dados que se pretendem extrair e analisar são dados provenientes das redes sociais.

O *Twitter* é o exemplo de uma rede social que permite a utilização da sua *API* para extração de dados. No entanto, nem todas as redes sociais disponibilizam uma *API* para o efeito pretendido, assim, a extração de dados pode ainda ser feita recorrendo a *web crawlers* criados para o efeito.

Seguidamente à extração de dados, deverá ser realizado o pré-processamento que consiste no processo de limpeza dos dados, preparando o texto para que posteriormente possa ser classificado.

A existência do pré-processamento deve-se ao facto de que os textos retirados das redes sociais geralmente contêm muito conteúdo que não é relevante. Nesse sentido, o pré-processamento dos dados reduz o "ruído" existente nos textos extraídos, ajudando a melhorar o desempenho do processo de classificação (Medhat, Hassan e Korashy, 2014,[26]). Alguns processos a considerar na fase de pré-processamento, segundo Medhat, Hassan e Korashy, (2014,[26]) são os seguintes:

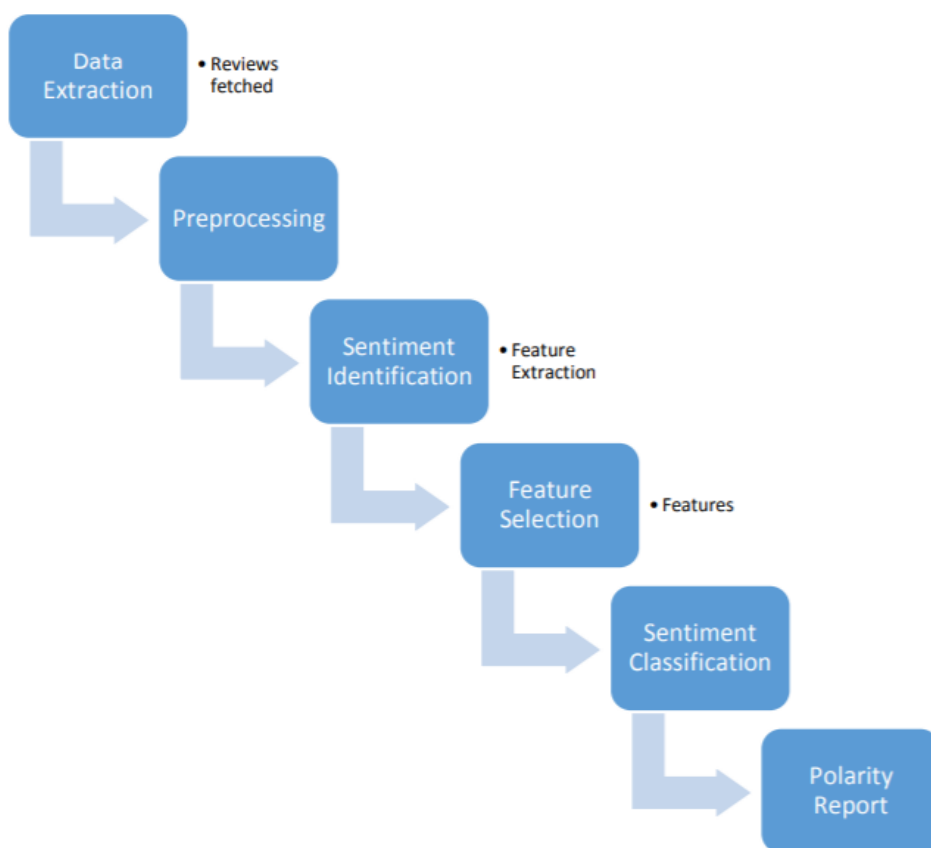


Figura 2.4: Etapas consideradas por Rajput e Solanki, (2016, [25]) na Análise de Sentimentos

- Remoção de *Uniform Resource Locators* (URL), *hashtags* e quaisquer tipos de caracteres especiais;
- Tradução da gíria para palavras simples com um mesmo significado;
- Remoção de letras em excesso nas palavras - por exemplo, a palavra: “*Happyyyyyy*” é reduzida para “*Happy*”.

Pode ainda considerar-se, na fase de pré-processamento os seguintes processos (De França e Oliveira, 2013, [6]):

- Tratamento da negação - dar equivalência de expressões como “não é bom” à expressão “mau”;
- Remoção de *stop words*<sup>2</sup>;
- *Stemming* - processo que visa a obtenção da raiz morfológica, eliminando os prefixos e sufixos que carregam a informação gramatical ou lexical da palavra.

<sup>2</sup>São palavras que podem ser consideradas irrelevantes para o conjunto de resultados a ser exibido numa consulta efetuada num motor de busca. Exemplos de *stop words*: as, e, os, de, para, etc.



As fases de identificação de sentimentos e seleção de características podem consistir segundo Medhat, Hassan e Korashy, (2014, [26]) em encontrar adjetivos que possam ser potenciais indicadores de opiniões, frases, palavras ou expressões utilizadas para expressar opiniões.

Quanto à classificação de sentimentos, como mostra a Figura 2.5, as técnicas de classificação de sentimentos podem ser divididas, essencialmente, em dois grupos distintos: as técnicas de *machine learning* e as técnicas baseadas em léxico.

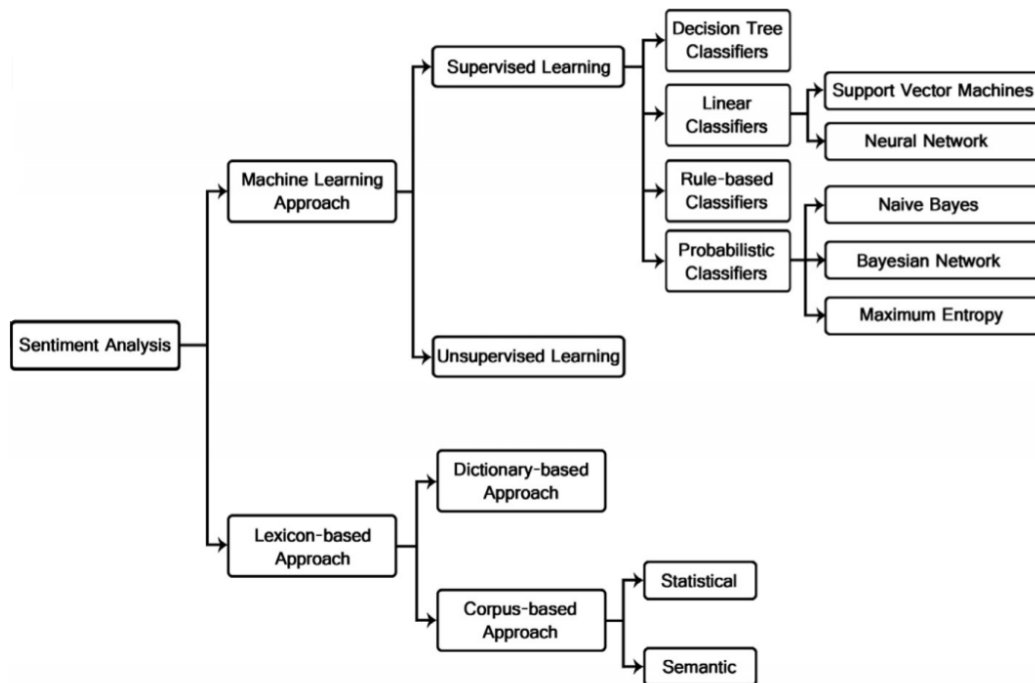


Figura 2.5: Técnicas de classificação de sentimentos segundo Medhat, Hassan e Korashy, (2014, [26])

Embora não estejam representadas na Figura 2.5, podem existir ainda técnicas híbridas, (misturando técnicas de *machine learning* e técnicas baseadas em léxico).

Na Figura 2.5 são ainda mostradas as sub-categorias das técnicas de classificação de sentimentos, bem como algumas das técnicas mais conhecidas e utilizadas (Medhat, Hassan e Korashy, 2014, [26]).

### 2.5.2 Classificação dos Sentimentos: Abordagens baseadas em Léxico e *Machine Learning*

Como mostra a Figura 2.5, existem duas abordagens principais para a análise de sentimentos: abordagem baseada em léxico e abordagem baseada em *machine learning*. O método baseado em léxico recorre a dicionários de sentimento, que contêm palavras de opinião, e combina-os, com os dados a avaliar, para determinar uma polaridade aos mesmos. Às palavras de opinião, presentes

nos dicionários de sentimento, são atribuídas pontuações descrevendo o quão positivas, negativas ou objetivas essas palavras são.

As abordagens de *machine learning* utilizam técnicas de classificação, recorrendo a algoritmos de *machine learning* para classificar texto.

### 2.5.2.1 Abordagens baseadas em léxico

Nas abordagens baseadas em léxico a classificação de sentimentos é feita comparando as características de um determinado texto com os dicionários de sentimento que contêm listas de palavras (normalmente adjetivos) e expressões utilizadas para expressar sentimentos e opiniões subjetivas das pessoas.

De acordo com Vohra e Teraiya, (2013, [3]) os passos básicos das técnicas baseadas em léxico são os seguintes:

1. Pré-processar o texto (remover tags, *HyperText Markup Language (HTML)*, caracteres irrelevantes, etc);
2. Inicializar a pontuação total do sentimento de um excerto de texto a analisar:  $s=0$ .
3. Efetuar o processo de *Tokenization* - processo que consiste em quebrar um fluxo de texto em palavras, frases, símbolos ou outros elementos significativos chamados de *tokens*. Para cada *token*, verificar se está presente no dicionário de sentimentos. Se o *token* estiver presente no dicionário, e for positivo, então  $s=s+w$ . Se o *token* for negativo, então  $s=s-w$ .
4. Verificar a pontuação total do sentimento do texto. Se  $s >$  limite, então o texto é classificado como positivo. Se  $s <$  limite, então o texto pode ser classificado como negativo.

Na Figura 2.5 da página 21 estão representados dois métodos para construir um dicionário de sentimentos: métodos baseados em corpus e métodos baseados em dicionários. Para além destes dois métodos, Vohra e Teraiya, (2013, [3]) considera existir um terceiro método: método manual que consiste numa tarefa difícil e demorada.

No método baseado em dicionário o processo começa com um pequeno conjunto de palavras de opinião, recolhidas manualmente e com as orientações conhecidas. A ideia deste método é fazer crescer esse conjunto de palavras inicial recorrendo a dicionários como o *WordNet*, que contém sinónimos e antónimos. Quando novas palavras são encontradas, são adicionadas à lista de palavras inicial e o processo termina quando já não são encontradas novas palavras.

O método baseado em dicionário apresenta uma grande desvantagem que é a incapacidade de identificar palavras de opinião que diferem a sua orientação em determinados contextos. Por exemplo, a orientação da palavra “velho” depende do contexto. Se é aplicada à qualidade do vinho, habitualmente é considerada como positiva. No entanto, se estivermos a referir-nos a um idoso pode ser considerada como uma expressão negativa.

Os métodos baseados em corpus (coleção de documentos) ajudam a resolver o problema de encontrar palavras de opinião com orientações específicas do contexto. Existem dois métodos nas abordagens baseadas em corpus: os métodos estatísticos e os métodos semânticos.

Nos métodos estatísticos se uma palavra aparece frequentemente entre textos positivos, então a sua polaridade é positiva. Se aparece frequentemente entre textos negativos, então a sua polaridade pode ser considerada negativa. Se tiver frequências iguais, então pode ser considerado como palavra neutra.

Nos métodos semânticos inicialmente extraem-se os termos de sentimento do texto não estruturado e calcula-se a polaridade dos termos. A maioria dos termos que carregam sentimento possuem mais do que uma palavra, por exemplo, “bom filme”, “boa cinematografia”, “bons atores”, etc.

### 2.5.2.2 Abordagens de *Machine Learning*

As abordagens de *machine learning* aplicáveis à análise de sentimentos pertencem, principalmente, à classificação supervisionada (Vohra e Teraiya, 2013, [3]).

Nas técnicas de *machine learning* são necessários dois tipos de dados: dados para efetuar o treino de um classificador, permitindo-o aprender, e um conjunto de dados para testar esse mesmo classificador, verificando o seu desempenho. Ao longo do tempo várias técnicas de *machine learning*, para classificação de texto, têm vindo a ser utilizadas. Segundo Vohra e Teraiya, (2013, [3]) as técnicas de *Naive Bayes*, *Maximum Entropy* e *Support Vector Machines* destacam-se por alcançarem grande sucesso na análise de sentimentos.

As técnicas de *machine learning* começam, normalmente, com uma recolha de um conjunto de dados de para treino. De seguida, utilizam-se os dados de treino anteriormente recolhidos para treinar um classificador. Por fim, o classificador é avaliado, utilizando-se para esse efeito, dados de teste.

## 2.6 Áreas de aplicação de Análise de Sentimentos e de *Opinion Mining*

Nos seguintes subcapítulos são apresentadas áreas de aplicação dos conceitos de Análise de Sentimentos e de *Opinion Mining*, destacando-se as áreas de compras, entretenimento, política, investigação e desenvolvimento, marketing e educação.

### 2.6.1 Compras

Atualmente, a pesquisa por opiniões de *bloggers* e *youtubers*, bem como de tantas outras pessoas comuns que via *online* expressam a sua opinião acerca de produtos e serviços, antecede e suporta, cada vez mais, a decisão de compra dos clientes.

Para além destas fontes de opinião, podem considerar-se ainda as opiniões expressas por clientes, na própria página da marca, avaliando o produto que outrora adquiriram. Como exemplo desta situação temos um comentário no site da *SheIn*<sup>3</sup>, ilustrado na Figura 2.6.

Neste caso, a cliente em questão adquiriu um vestido e posteriormente postou uma fotografia vestindo o artigo e dando a sua opinião pessoal acerca do mesmo.

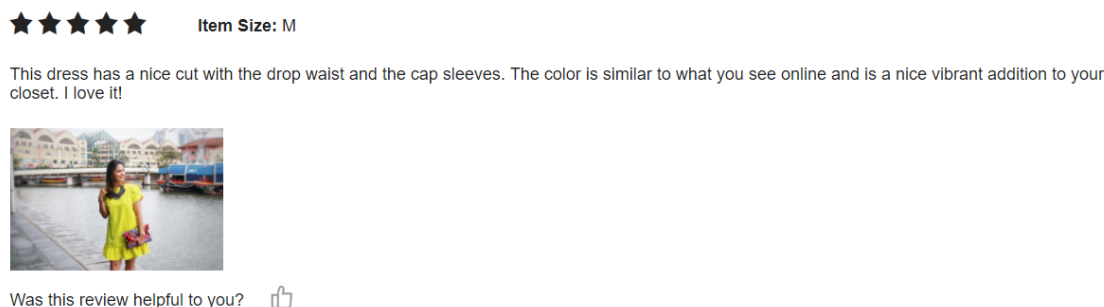


Figura 2.6: Exemplo de uma opinião acerca de um produto adquirido, <http://www.shein.co.uk/>, acessado pela última vez em 2018-04-04, 14:56

A *Amazon* é outro exemplo, assim como outras tantas lojas *online*, que também reservam um espaço nos seus *sites* para que o cliente possa exprimir a sua opinião acerca dos produtos.

### 2.6.2 Entretenimento

De acordo com Talele MECSE e Chandrashekar Badgujar, (2012, [33]) técnicas de *opinion mining* também se podem aplicar nas áreas de entretenimento.

Quando procuram um filme ou uma série de televisão para assistir, muitos são os utilizadores que previamente consultam opiniões e avaliações acerca desses mesmos filmes e séries. O *Internet Movie Database (IMDb)* é uma excelente fonte de informações, avaliações e opiniões acerca de séries e filmes.

No sentido de agrupar e sumarizar os comentários, opiniões e avaliações relativamente a determinados tópicos na área de entretenimento é relevante recorrer a técnicas de *opinion mining*.

### 2.6.3 Política

Atualmente, o público em geral, reage aos partidos políticos através das redes sociais. Devido ao grande número e diversidade de publicações, estas importantes reações públicas não são processadas para formular informações úteis que podem ser usadas para formar uma imagem melhor da voz pública relativamente ao cenário político.

Recorrendo a técnicas de *opinion mining* é possível dar a conhecer tanto aos candidatos eleitorais como aos cidadãos comuns, as opiniões sociais prevalecentes acerca do cenário político. Deste modo, os cidadãos comuns ficarão melhor informados e tais opiniões poderão leva-los a

<sup>3</sup>Loja *online* que vende roupa feminina

pesquisar melhor acerca de possíveis dúvidas e até suportar a sua decisão de voto. No que diz respeito aos candidatos eleitorais, estes podem ficar a conhecer melhor a opinião prevalecente do público acerca de si próprios e acerca dos seus concorrentes, sendo-lhes permitido, deste modo, identificar as suas forças e fraquezas.

#### 2.6.4 Investigação e Desenvolvimento

As *reviews* de produtos são úteis para potenciais compradores. No entanto, são também bastante pertinentes para os fabricantes e equipas de desenvolvimento desses mesmos produtos.

Perante uma *review* de um produto, são normalmente apontadas características positivas e negativas, bem como possíveis melhorias. Todas estas opiniões do público em geral, são úteis para perceber de que forma o produto poderá tornar-se melhor. Neste sentido, são disponibilizadas aplicações web que oferecem plataformas para que os clientes possam projetar produtos e enviar os seus projetos para as empresas responsáveis pelo fabrico. Este tipo de abordagens pode ajudar significativamente no estabelecimento de recursos que são apreciados pelos clientes [33].

#### 2.6.5 Marketing

As empresas agora podem economizar nas despesas de marketing solicitando revisões em seus sites e sites de revisão por pares. Isso elimina a necessidade de consultores de negócios realizarem pesquisas, pois as empresas agora podem ter todos os dados de que precisam on-line. O advento da internet trouxe consigo novas formas de marketing. A família e os amigos podem agora recomendar produtos / serviços uns aos outros ou buscar mais conhecimento sobre um produto ou serviço antes de se comprometerem. É análogo ao tradicional marketing boca-a-boca de produtos e serviços. Para incentivar postagens e recomendações entre os pares, os profissionais de marketing normalmente oferecem incentivos como descontos para recomendações que se transformam em compras [14]. Consideremos a seguinte recomendação para um resort turístico: “É uma terra de contrastes e majestade, a África mais selvagem e inexplorada [12]”. Opiniões positivas deste tipo podem melhorar a classificação dos produtos. WAYN (onde você está agora) é uma rede social que permite amigos acompanharem uns aos outros durante suas viagens, postando fotografias e descrevendo suas aventuras e experiências.

#### 2.6.6 Educação

As opiniões e sentimentos dos alunos relativamente aos métodos de ensino e relativamente aos próprios docentes pode ajudar a melhorar a prestação de serviços. Os coordenadores das unidades curriculares podem saber o que os alunos pensam sobre os membros de sua equipe, solicitando-lhes que efetuem avaliações *online* como parte da exigência do curso.

A *Curtin University of Technology*<sup>4</sup> oferece espaços *online* onde alunos devem submeter revisões semanais e também disponibiliza fóruns de discussão. Estes meios de exposição de opiniões

---

<sup>4</sup>Universidade Australiana de tecnologia

agem como fontes ricas de conteúdo gerado pelos alunos e que pode ser posteriormente extraído, constituindo uma medida de avaliação da qualidade do ensino.

## Capítulo 3

# Evolução do Projeto e Ferramentas Utilizadas

Ao longo deste capítulo serão descritas as várias fases pelas quais passou este projeto e serão ainda descritas as ferramentas utilizadas (e as que se ponderou utilizar) para atingir os resultados.

### 3.1 *Heuristics and Ontology-based Learning Machines and Exponential Systems (HOLMES)*

A ideia inicial do projeto, representada na Figura A.1 da página 63, mostra três blocos. Segundo esta abordagem, o bloco 1 seria relativo à extração de dados a partir das redes sociais, definindo uma certa frequência para a recolha dos dados, montando para isso um *Robotic Process Automation*, (*RPA*), o bloco 2 trataria da análise desses dados recolhidos e o bloco 3 consistiria na elaboração de *dashboards* com base nos dados do bloco 2. Esta dissertação trataria do bloco 1.

A divisão tripartida mostrada na Figura A.1 da página 63 supunha que o *HOLMES* seria capaz de efetuar a extração e o tratamento dos dados extraídos. Nesta fase do projeto foram realizadas pesquisas acerca de conceitos de inteligência artificial, *machine learning*, análise de sentimentos, sobre *RPA* e sobre formas de armazenar os dados extraídos. Todos estes conceitos seriam fundamentais para o desenvolvimento de uma solução. Paralelamente tentava-se procurar informação disponível sobre o *HOLMES*, plataforma de inteligência artificial da *Wipro*. Contudo, encontrar qualquer informação mais específica acerca do *HOLMES* não se assemelhou tarefa fácil, isto porque, apesar do *HOLMES* ser propriedade da *Wipro*, é uma ferramenta que está sob o controlo de uma equipa da *Wipro*, na Índia. Para além da existência de informação extremamente confidencial, a comunicação com as equipas da Índia não foi tarefa fácil. Mais difícil assemelhava-se a obtenção de acesso à própria ferramenta - que seria fundamental para extrair os dados provenientes das redes sociais (no caso deste primeiro bloco).

Tendo-se passado perto de um mês sem mais informações acerca das capacidades e funcionalidades do *HOLMES* e sem a possibilidade de utilizar uma ferramenta alternativa, (a intenção seria utilizar uma ferramenta da empresa para efetuar este passo) o trabalho até então efetuado

assemelhava-se bastante teórico e voltado para os vários conceitos associados a uma ferramenta de análise das redes sociais.

Todo o tempo dispensado na procura de informações relativas às ferramentas da *Wipro* bem como todo o tempo dispensado na arquitetura desta solução, justifica-se pelo facto de se tratar de uma solução com bastante potencial de ser integrada nas soluções de *Oracle Retail* usadas pela *Wipro* na grande maioria dos seus projetos.

De modo a melhor detalhar o projeto e a fim de efetuar uma melhor separação entre os dois primeiros blocos, a *Wipro* redefiniu a arquitetura desta solução, mostrada na Figura A.2 da página 64.

Continuando sem mais informações acerca das capacidades do *HOLMES* a pesquisa voltou-se para a procura de soluções alternativas (já existentes no mercado).

### 3.2 *Data Discovery Platform (DDP)*

Ao fim de precisamente um mês e um dia do início deste projeto e ainda sem nenhuma informação relevante acerca das capacidades do *HOLMES* foi conseguido um contacto com uma equipa da Índia que nos deu a conhecer que o *HOLMES* era um *software* com inúmeras funcionalidades e que para este projeto em concreto, o ideal seria utilizar uma das ferramentas do *HOLMES*, o *DDP*.

O *DDP* é uma plataforma integrada da *Wipro* capaz de recolher e gerir dados de forma a produzir medidas acionáveis, através de análises avançadas.

Para este caso o mesmo processo se repetiu, procurou-se informação acerca do *DDP*, mas a obtenção de informação específica acerca desta ferramenta (tal como aconteceu para o caso do *HOLMES*) não foi possível. O acesso ao *DDP* também nunca foi possível, uma vez mais, tudo isto, devido, principalmente, a questões de confidencialidade.

Sem informações mais específicas acerca do *DDP*, a solução começou a ser direcionada para um dos clientes da *Wipro* que outrora mostrara interesse num projeto deste tipo. Sem conhecimento profundo sobre as funcionalidades do *DDP*, foi-se pensando e selecionando informações úteis a extrair das redes sociais de modo a trazer informações relevantes para os processos de negócio do *Retailista A*, mostrados na Figura 4.1 da página 36.

Em meados de abril a equipa da Índia, detentora do *DDP*, fez uma apresentação explicando melhor em que consistia esta ferramenta. Embora esta apresentação tenha esclarecido algumas das funcionalidades desta ferramenta, que de facto iam de encontro ao que era esperado, nunca foi possível o acesso direto à ferramenta em questão. O contacto tido com esta plataforma de extração de dados, ao longo deste projeto, foi apenas do ponto de vista de cliente, o que significa que conforme pedidos, os dados eram recolhidos pela equipa de *DDP* e eram-nos posteriormente enviados.

Tendo em conta o projeto *Customer Xperience* como um todo, é de salientar que o *DDP* deverá ser utilizado apenas para o bloco 1. No bloco 1, devem ser utilizadas as capacidades de extração de dados do *DDP* bem como as suas capacidades de correção de texto, atribuição de palavras-chave e análise de sentimentos dos dados extraídos, entre outras funcionalidades deste



tipo. Numa fase posterior, os dados extraídos pelo *DDP* devem ser guardados numa base de dados, deverá ser efetuada uma primeira análise dos mesmos e estes devem ser disponibilizados ao segundo bloco de continuação deste projeto. É de referir ainda que todo o tipo de tratamento e análise dos dados extraídos, bem como a transformação dos mesmos em *dashboards*, deverá ser feita de forma independente ao *DDP*.

### 3.2.1 *DDP* - Casos de Uso

O *DDP* oferece soluções em diversas áreas, desde o setor energético - possibilitando uma melhor gestão da rede elétrica, fornecendo previsões da procura e geração de energia solar, utilizando as tendências meteorológicas passadas e as previsões meteorológicas - até áreas financeiras, de saúde e áreas de tecnologia - prevendo problemas para implementar medidas preventivas e manutenção ou substituição de componentes. Para além destas áreas a plataforma *DDP* possui ainda soluções nas áreas do retalho.

De seguida são descritos alguns casos de usos do *DDP* nas áreas do retalho.

- Otimização de *Packs*

Com o objetivo de identificar que produtos se costumam comprar em conjunto, com a intenção de determinar que produtos levam à venda de outros produtos e de modo a desvendar os padrões de gasto do consumidor para que seja possível associar dois ou mais produtos em *packs* e levar ao planeamento de outro tipo de estratégias de mercado, são utilizados detalhes dos produtos bem como as suas respetivas informações de venda. Assim, o *DDP* é capaz de projetar futuras promoções ou ofertas, assim como outro tipo de estratégias de mercado, permitindo ao retalhista conseguir planear atividades promocionais eficientes e bem organizadas, selecionando promoções e estratégia de mercado eficazes.

- Desempenho da Loja

Com o objetivo de categorizar e analisar fatores internos e externos à loja, de modo a determinar a relevância de cada um dos fatores para o desempenho da mesma, assim como sugerir melhorias, são utilizados dados internos e externos das lojas. Deste modo, o *DDP* é capaz de analisar os números de visitantes, os custos operacionais, as vendas, etc. em cada loja, a fim de comparar o desempenho de cada uma das lojas e decidir onde efetuar investimentos adicionais de realizar melhorias.

- Segmentação de Clientes

Com o objetivo de dividir os clientes em diferentes segmentos, de acordo com a sua localização, compromisso para com a marca, preferência por determinados canais de compra, de acordo com as diferentes necessidades de cada cliente, comportamentos e preferência por determinados produtos.

São utilizados os dados detalhados dos produtos e os dados acerca dos clientes,

Assim, através de análises detalhadas, é possível entender que segmentos de clientes são mais lucrativos para o retalhista.

O *DDP* é, deste modo, capaz de desenvolver modelos para criar campanhas de marketing segmentadas, é capaz de desenvolver estratégias diferenciadas de atendimento aos clientes e direcionar as perspectivas com o maior potencial de desenvolvimento.

Atingidos todos os objetivos, é possível um melhor atendimento no que diz respeito ao cliente, é possível descobrir necessidades não atendidas e entregar melhores produtos e serviços a novos grupos de pessoas.

- Campanhas direcionadas a determinados clientes

Com o objetivo de melhorar as taxas de resposta e rentabilidade dos programas de marketing direcionando-as para clientes que irão, mais provavelmente, responder às campanhas, são efetuadas, pelo *DDP*, análises preditivas para identificar compradores mais prováveis para determinadas atividades específicas. Para tal, são utilizados dados demográficos dos clientes e dados do produto.

Com este tipo de análise, o *DDP* permite ao retalhista identificar novas oportunidades e minimizar os recursos dedicados a metas e atividades aparentemente pouco atraentes.

- Preferência do canal do cliente

Utilizando dados detalhados do produto, dados demográficos do cliente e dados acerca de campanhas passadas, o *DDP* permite ao retalhista identificar quais os canais mais utilizados pelos seus clientes de modo a planear o modo de interação com o cliente, bem como campanhas de marketing.

- Análise Social Competitiva

O *DDP* é capaz de extrair publicações relevantes das redes sociais e analisar esses dados para identificar tópicos, géneros, localização, sentimentos etc. usando técnicas de previsão e PNL.

Deste modo, o *DDP* permite que as organizações analisem a resposta dos seus consumidores, nas várias redes sociais, aos seus produtos e serviços, em comparação com os seus concorrentes, permitindo assim que as organizações possam identificar os segmentos de mercado que estão em risco ou os que representam oportunidades de negócio. É ainda possível às organizações identificar forças e fraquezas das suas marcas.

### 3.3 *Buzzmonitor*

O *Buzzmonitor* é um *software de social media monitoring* da empresa *E.Life*. Esta ferramenta surge nesta dissertação para que exista efetivamente o contacto direto com uma ferramenta de análise das redes sociais, sendo possível, deste modo, enfrentar alguns dos principais desafios

aquando da extração de dados, a partir das redes sociais. Utilizando diretamente o *Buzzmonitor* é ainda possível efetivamente testá-lo e compará-lo com a ferramenta de *DDP*, identificando possíveis aspetos de melhoria no *DDP*.

A *E.Life* desenvolveu o *Buzzmonitor* no sentido de democratizar o acesso à informação proveniente das redes sociais [27] pois até então, a informação recolhida a partir das redes sociais era entregue apenas a grandes empresas. Assim, foi criado o conceito de *social media* para todos. O *Buzzmonitor* surge neste sentido e permite que pequenas e médias empresa saibam o que se diz sobre elas nas redes sociais. Com o *Buzzmonitor*, a *E.Life* pretende ir mais além na análise das redes sociais. Mais do que trabalhar apenas com *likes* e *retweets*, o *Buzzmonitor* trabalha os dados intrínsecos das redes socais, [27].

Considerando uma cadeia de comida saudável, até agora com uma ferramenta de *social media monitoring* tradicional, esta cadeia de comida saudável conseguia perceber o que as pessoas diziam sobre a sua marca, os *likes* e *retweets* que faziam das suas publicações. Com o *Buzzmonitor*, para além disso é possível perceber como as pessoas avaliam o preço dos seus produtos e dos produtos concorrentes. É possível perceber que novos produtos ou que receitas mais têm sido comentadas nas redes socais. Com o acesso a esta informação, a cadeia de comida saudável poderia decidir incluir as receitas mais faladas nas redes sociais, nos seus menus.

No fundo, o *Buzzmonitor* permite retirar informação gerada pelo consumidor nas redes sociais em benefício do próprio negócio [27].

O *Buzzmonitor* é uma ferramenta que consiste em três áreas fundamentais:

- *Social Customer Relationship Management (Social CRM)* - O *Social CRM*, permite gerir a interação das páginas de uma determinada marca com os seus clientes. Esta funcionalidade requer que se esteja com a conta do *Twitter* ou com a conta do *Facebook* sincronizada com a conta do *Buzzmonitor*. O *Social CRM* permite que marcas e consumidores conversem sem hierarquias, criando uma relação de proximidade;
- *Social Analytics* - Permite analisar a evolução das páginas das redes sociais da marca e de marcas concorrentes, permite ainda perceber qual o compromisso do cliente para com a marca em questão (assim como para marcas concorrentes);
- *Monitoring*- Permite criar relatórios específicos como por exemplo, relatórios dos assuntos mais falados acerca de uma determinada marca, sentimentos predominantes perante uma determinada marca, entre outros tantos relatórios personalizáveis.

### 3.3.1 Análise de Sentimentos

O *Buzzmonitor* utiliza um conjunto de classificadores permitindo extrair o sentimento dos utilizadores, expresso através de publicações nas redes sociais, face a uma ou várias entidade ou face a determinados tópicos.

A classificação de sentimentos das publicações das redes sociais, efetuada pelo *Buzzmonitor*, consiste, numa primeira fase, no cálculo de probabilidades para um conjunto de dados de treino.

Estes dados de treino traduzem-se num conjunto de frases, já pré-classificadas em positivas e negativas, por um analista humano [9]. Baseando-se nestes dados de treino o *Buzzmonitor* fica preparado para classificar novos dados, que surgem diariamente a partir das redes sociais.

Os classificadores utilizados pelo *Buzzmonitor* são classificadores estatísticos, essencialmente classificadores com base no *Teorema de Bayes* [9]. A Equação 3.1 expressa matematicamente o *Teorema de Bayes*, tendo em conta que A e B são eventos e que a probabilidade de B é diferente de zero.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (3.1)$$

Para o caso particular de análise do sentimento em comentários provenientes das redes sociais, o teorema diz que podemos calcular a probabilidade do sentimento de uma frase se soubermos as probabilidades dos sentimentos de cada termo nela presente, tendo ainda em consideração a ordem em que os termos ocorrem na frase.

Supondo que numa frase existe o termo "bem", sozinho esse termo tem uma probabilidade de sentimento positivo de 0.8. No entanto, supondo que nessa mesma frase, logo a seguir a "bem", encontra-se o termo "mal", o algoritmo já possui, pré-calculado (na sua fase de aprendizagem),  $pPOS(bem\mal)$ , ou seja, a probabilidade do sentimento ser positivo, dado que na frase ocorre "bem" e dado que em seguida se encontra o termo "mal". Esse valor é muito provável, em exemplos reais  $pPOS(bem\mal) = 0.1$  o que significa que "bem mal" vai empurrar o resultado final de classificação de sentimento dessa frase para o lado negativo [9].

No caso de numa frase se encontrarem os termos "bem bom",  $pPOS(bem\lbom)$ , ou seja, a probabilidade do sentimento de uma determinada frase ser positiva, sabendo que se encontrou a palavra bem e em seguida a palavra bom, é num caso real, bastante alta, resultando no valor 0.9.

O sentimento de uma qualquer publicação encontrada nas redes sociais é uma fórmula que concatena todas estas probabilidades e dá um resultado entre 0 e 1, naturalmente probabilístico. Com base na probabilidade final, o sistema classifica uma publicação como positiva, neutra ou negativa.

Os classificadores utilizados, como já foi referido, são classificadores estatísticos, o que significa que os resultados acarretam um erro associado e por isso devem ser tratados como indicadores e nunca como verdade absoluta.

## Capítulo 4

# ***Retalhista A* - Necessidade e Viabilidade da Adoção de uma Ferramenta de Análise das Redes Sociais**

Neste capítulo, será definido qual o retalhista para o qual esta solução será direcionada. Devido a questões de confidencialidade, o retalhista em estudo, será denominado de *Retalhista A*. Ao longo deste capítulo serão ainda definidas as necessidades que levaram este cliente da *Wipro* a ponderar a adoção de uma ferramenta de análise das redes sociais. Neste sentido, é feita uma análise *SWOT* de modo a analisar a viabilidade de adoção de uma solução de análise das redes sociais, no caso do *Retalhista A*. Por fim, são dados a conhecer alguns processos de negócio do *Retalhista A*, outrora definidos e selecionados para este estudo e para os quais seria útil recolher informações provenientes das redes sociais de forma a colocar o *Retalhista A*, numa posição avançada no mercado, face aos seus concorrentes.

### **4.1 *Retalhista A***

Este estudo irá incidir sobre a área de moda de um cliente da *Wipro*, o *Retalhista A*, que se trata de uma multinacional, com sede no Reino Unido e que vende produtos desde vestuário de mulher, homem e criança até brinquedos, mobiliário, produtos eletrónicos, acessórios e produtos de beleza.

Este retalhista possui marca própria, no entanto, para além de produtos de marca própria oferece ainda aos seus clientes produtos das mais conceituadas e diversas marcas.

#### **4.1.1 Necessidade de Análise das Redes Sociais**

No caso do *Retalhista A* a maior necessidade de adotar uma ferramenta de análise das redes sociais surge devido ao facto deste ser um retalhista bastante dependente de dados históricos e internos para gerir os vários processos da empresa.

Para o *Retalhista A*, a previsão da procura, na área de moda, torna-se pouco precisa, pois nesta área é ainda mais desafiante efetuar previsões da procura, isto porque, a procura é extremamente variável e regida não só pela sazonalidade mas também por tendências. Assim sendo, prever a procura utilizando para isso, dados internos e históricos pode ter uma precisão satisfatória se estivermos perante produtos básicos, cujo consumo não é influenciado por tendências ou mudanças repentinas na procura, (como é o caso de arroz, massa, roupa básica, etc). Nestes casos, utilizar dados internos e históricos para prever a sua procura é um método bastante preciso. A área da moda, sendo uma área regida por tendências e pela sazonalidade, requer, cada vez mais, previsões mais complexas do que simplesmente assumir que tudo acontecerá dentro do que aconteceu em anos anteriores e segundo dados internos relativos unicamente à própria empresa.

#### **4.1.2 Análise SWOT**

Uma análise *SWOT* identifica os pontos positivos, indica quais os pontos que devem ser melhorados, mostra as oportunidades de crescimento e aponta alguns riscos e ameaças de uma empresa. Assim, é possível identificar elementos chave para a gestão de uma empresa, possibilitando um diagnóstico do estado da mesma e permitindo planeamentos estratégicos.

Contudo, esta ferramenta não serve apenas para elaborar planeamentos estratégicos e visualizar o posicionamento de uma empresa no mercado, serve também para a elaboração de novos projetos, [7]. Neste caso específico recorreu-se à análise *SWOT* para analisar, do ponto de vista do *Retalhista A*, a decisão de adotar uma solução tecnológica para analisar as rede sociais de modo a poder, a partir dessas mesmas análises, perceber padrões e comportamentos sociais a fim de ajudar este mesmo retalhista em determinadas atividades relacionadas com a previsão da procura (preço, reaprovisionamento, alocação de *stocks*, *layouts* de loja, experiência do cliente em processos do *Omni-channel* e processos de campanhas direcionadas).

Assim, foram formulados pontos fortes, pontos fracos, oportunidades, e ameaças relativamente à decisão de adotar uma solução que permite a recolha, tratamento e elaboração de *dashboards* utilizando dados das redes sociais.

##### **4.1.2.1 Strengths (Forças)**

- Entender, de uma forma resumida, o que os clientes pensam e sentem acerca da marca (bem como acerca de marcas concorrentes);
- Perceber o nível de compromisso dos clientes para com a marca (e para com marcas concorrentes);
- Perceber a presença da própria marca e de marcas concorrentes nas redes sociais;
- Monitorizar a atividade de competidores percebendo os seus pontos fortes e fracos;
- Analisar as principais críticas e elogios dados à marca e a marcas concorrentes;
- Perceber futuras tendências na área da moda;

- Possibilidade de combinar os dados internos da empresa com os dados externos (extraídos a partir das redes sociais) para efetuar previsões e suportar tomadas de decisão.

#### 4.1.2.2 *Weaknesses* (Fraquezas)

- Dados nem sempre fiáveis;
- Classificação de sentimentos imperfeita;
- Recolha de dados nem sempre perfeita.

#### 4.1.2.3 *Opportunities* (Oportunidades)

- Possibilidade de um melhor planeamento da procura e melhor planeamento de atividades da empresa (melhor planeamento de preço dos artigos, reaprovisionamento, alocação de *stocks*, *layouts* de loja, processos de campanhas direcionadas, etc) ;
- Possibilidade de repensar a proximidade com o cliente;
- Possibilidade de identificar oportunidades de melhoria;
- Possibilidade de identificar crises.

#### 4.1.2.4 *Threats* (Ameaças)

- Com o crescimento de ferramentas deste tipo poderá crescer o número de comentários falaciosos acerca de determinadas marcas e produtos induzindo a análises poucos fiáveis.

## 4.2 *Retalhista A* - Processos de Negócio

A ideia de adotar uma ferramenta de análise das redes sociais tem por objetivo extrair das redes sociais dados relevantes, que possam ser transformados em medidas acionáveis, que permitam ao *Retalhista A*, utilizar essa ferramenta de análise das redes sociais, para lhe dar suporte nas tomadas de decisão.

A Figura 4.1 da página 36 mostra cinco dos principais processos de negócio do *Retalhista A*.

Combinando os dados extraídos das redes sociais com dados internos à própria empresa, o *Retalhista A* terá uma boa base de sustentação para as suas tomadas de decisão.

Analisando os processos de negócio do *Retalhista A*, verifica-se que existem alguns desses processos que contêm áreas demasiado específicas e cuja extração de dados provenientes das redes sociais pode ser injustificável quer seja devido à baixa frequência com que determinados tópicos são partilhados nas redes sociais, quer seja devido ao facto de ser extremamente complicado filtrar dados que ocorrem tão poucas vezes.

Analisando os processos de negócio bem como algumas das suas áreas mais específicas verifica-se ainda que de facto, a maior parte desses processos dependem de previsões, nomeadamente de

## Data Discovery Platform – Embedding insights in key business processes



Figura 4.1: Principais Processos de Negócio do *Retailista A*

previsões de procura (*Mudança nas Tendências de Consumo*). Na Figura B.1 da página 65 é mostrado um esquema que representa a dependência existente entre os processos de negócio. Analisando a Figura B.1 da página 65 verifica-se que apenas o processo de *Reconhecimento do Nome da Marca* é realmente independente dos demais. Pelo facto do processo de *Mudança nas Tendências de Consumo* provocar dependências na maioria dos processos, este processo será abordado em primeiro lugar. É de referir que na Figura B.1 da página 65 são mostradas as várias entradas (a cinzento) que se pensou ser necessárias extrair das redes sociais para cada processo (a azul) e que serão detalhadas nos seguintes sub-capítulos.

### 4.2.1 Mudança nas Tendências de Consumo - Previsão da Procura

A previsão da procura refere-se a fazer estimativas sobre a procura futura dos clientes por determinados produtos ou serviços, utilizando, para isso, dados históricos assim como outro tipo de informações úteis para este efeito. A previsão adequada da procura fornece às empresas informações valiosas sobre o seu potencial no atual mercado e noutra tipo de mercados, para que os responsáveis pela gestão da empresa possam tomar decisões informadas sobre preços, estratégias de crescimento de negócios e potencial de mercado. Sem previsão da procura, as empresas correm o risco de tomar decisões erradas sobre os seus produtos e mercados-alvo.

De acordo com Jorge Caiado, (2016, [18]) "*em gestão de empresas, como em todos os domínios de atividade humana, a previsão desempenha um papel fundamental no processo de tomada de decisão e tem implicações na definição dos planos operacionais de cada uma das áreas funcionais da empresa*".

A previsão da procura segundo Brusque e Zucatto, (2015,[22]) é fundamental para a tomada de decisões nas empresas, pois a partir da previsão da procura é possível utilizar os recursos de



forma adequada.

Considerando o ponto de vista destes dois autores, pode considerar-se que uma boa previsão da procura é a chave para que um negócio, acompanhado de uma boa gestão, seja bem sucedido. Neste sentido, as pequenas, médias ou grandes empresas tentam aplicar nas suas organizações modelos para prever a procura de produtos, tornando possível à empresa, gerir grande parte das suas atividades (marketing, produção, transporte, etc), com sucesso. Esses modelos de previsão da procura podem tratar-se de modelos de previsão qualitativos ou quantitativos. Os modelos qualitativos baseiam-se na subjetividade, sendo os dados julgados por pessoas chave (especialistas), capazes de opinar acerca da procura futura. Os métodos qualitativos, como descrito em (Nunes Dias et al, 2015, [21]) podem ser pesquisas de mercado, simulação de diferentes cenários futuros, de acordo com a opinião de especialistas e pode ainda recorrer-se ao método de *Delfi*. O método de *Delfi* consiste em duas etapas: na primeira etapa é realizado um *brainstorming* com os vários participantes, sendo as suas opiniões recolhidas e analisadas por um moderador, na segunda etapa, chega-se a um acordo em torno das opiniões sugeridas na etapa de *brainstorming*. No que diz respeito aos métodos quantitativos, estes são descritos em (Nunes Dias et al, 2015, [21]) como sendo métodos que utilizam padrões históricos dos dados para prever a procura futura. Os métodos quantitativos podem ser subdivididos em dois grandes grupos: as técnicas baseadas em séries temporais e as técnicas baseadas em correlações. Os métodos de séries temporais assumem que o padrão da procura no passado se reproduzirá no futuro, assim, analisam-se dados passados para prever as tendências para o futuro através de técnicas estatísticas. As previsões baseadas em correlação procuram prever a procura de determinado produto com base na previsão de outra variável que esteja relacionada com o produto.

De acordo com Braga, Filho e Alegre, ( 2015, [35]) a forma de previsão da procura mais promissora é a que integra os métodos quantitativos e qualitativos. Ambos os métodos possuem vantagens exclusivas, o que torna a integração particularmente interessante.

Analisar as redes sociais, procurando perceber tendências e padrões dos consumidores, a fim de prever a procura, não se enquadra em tentar perceber a procura tendo como base dados históricos, é por isso que esta solução se trata de uma solução inovadora, permitindo a um determinado retalhista estar um passo mais à frente dos demais.

A fim de alcançar o objetivo de prever a procura e de conseqüentemente trazer informações úteis e relevantes para os processos de negócio da Figura 4.1 da página 36, procuram-se extrair, a partir das redes sociais, os seguintes dados:

- Principais tópicos (acerca de moda) mais comentados e sentimentos prevalecentes;
- Adesão do público às promoções do *Retalhista A* e dos seus concorrentes;
- Sentimentos mais comuns quando se comenta sobre a marca;

Obtendo, a partir das redes sociais, os assuntos mais comentados acerca de moda (cores, padrões, estilos, eventos e peças de roupas) bem como os respetivos sentimentos acerca dos mesmos, é possível ao retalhista repensar no seu plano de compras, nos artigos que dispõe nas lojas e na

forma como os irá dispor. Com base nos assuntos mais comentados acerca de moda o retalhista poderá ainda repensar em promoções ou em novas estratégias de preço de modo a vender artigos que já não são tendência.

A adesão do público às campanhas efetuadas pela marca, nas redes sociais, pode ser um importante fator a considerar quando se pensa em prever a procura de um produto. Assim, se a marca promover, nas redes sociais, uma determinada peça de roupa poderá perceber qual o interesse por parte dos seus clientes nessa mesma promoção e dessa forma prever a procura desse artigo. Para além de analisar a adesão do público relativamente às suas próprias campanhas o *Retalhista A* pode ainda analisar a adesão do público relativamente a campanhas concorrentes, e dessa forma, prever a procura pelos seus próprios produtos, tendo em conta a procura pelos produtos das marcas concorrentes.

Por fim, perceber o que os clientes pensam acerca da própria marca, poderá influenciar na procura por produtos da mesma marca. Se no mês atual o público expressa críticas e sentimentos negativos relativamente à marca, pode ser um indicador de que a procura por produtos dessa marca poderá decrescer nos próximos tempos. Neste sentido, poderá ser uma oportunidade para a empresa perceber o que há de errado e em função disso o que deverá ser melhorado para que as opiniões do público face à própria marca melhorem.

#### **4.2.2 Performance das Lojas e Planeamento do Espaço**

A primeira coluna da Figura 4.1 da página 36 representa o processo de *Performance das Lojas e Planeamento do Espaço*, este processo é constituído pelas três sub-áreas mostradas na Figura 4.1 da página 36: a área de *Desempenho das Lojas*, a área de *Localização de Categorias de Produtos* e a área de *Otimização do Staff nas Lojas*. A sub-área de *Localização de Categorias de Produtos*, que consiste em perceber padrões de compra do consumidor de forma a melhor otimizar a disposição dos produtos numa loja, mostra ser complexo, através da análise das redes sociais, perceber os hábitos de compra que um mesmo cliente tem. Descobrir os hábitos de compra de cada cliente, através da recolha de dados das redes sociais torna-se uma informação bastante específica e nem sempre fácil de filtrar. Sendo uma informação cuja frequência de partilha é pouco comum torna-se pouco relevante uma análise sobre esta informação.

No entanto, a partir de dados internos, é perfeitamente possível, e suficiente, identificar hábitos comuns de compra e a partir disso otimizar a alocação das categorias de produtos nas lojas.

Para a sub-área de "*Desempenho das Lojas*" seria interessante extrair das redes sociais os seguintes dados:

- Principais críticas ou sugestões dadas acerca das lojas.

Sabendo quais as principais reclamações ou sugestões dos clientes relativamente ao desempenho das lojas será possível, ao *Retalhista A*, perceber pontos de melhoria.

A sub-área de "*Otimização do Staff nas Lojas*" diz respeito à identificação de horários e períodos de tempo em que não será compensatório ter tantos funcionários na loja ou, pelo contrário,

identificar determinados momentos em que deveriam existir mais funcionários para atender clientes. Para esta sub-área seria interessante extrair das redes sociais os seguintes dados:

- Adesão dos clientes às promoções da marca;
- Críticas dos clientes relativamente à alocação do *staff* da loja.

Uma elevada adesão dos clientes às promoções, divulgadas, via redes sociais, pela marca, pode ser um indicador de que no dia em que se vai fazer aquela promoção poderá ser necessário alocar mais ou menos funcionários na loja.

Críticas relativamente à alocação dos funcionários nas lojas poderão ajudar o retalhista a perceber períodos de tempo em que são feitas alocações de *staff* pouco compensatórias, estando a ser desperdiçado dinheiro em recursos que não são necessários ou, pelo contrário, ser necessário investir em recursos em determinados períodos.

### **4.2.3 Disponibilidade de produtos na Cadeia de Abastecimento**

O processo de *Disponibilidade do Produto na Cadeia de Abastecimento* está bastante relacionado com a previsão da procura. A sub-área de *Stock Out e Vendas Perdidas* deste processo, está intimamente relacionada com a previsão da procura. *Stock Out* significa precisamente a indisponibilidade de um determinado produto para venda. Esta indisponibilidade deve-se a uma previsão da procura que falhou. As vendas perdidas são também resultado de uma má previsão da procura. Uma venda perdida pode ser, por exemplo, um tamanho que simplesmente não existe na gama de produtos do retalhista. É da responsabilidade do retalhista perceber as vendas perdidas e adequar o seu negócio às necessidades dos seus clientes. Para a sub-área de *Stock Out e Vendas Perdidas* será interessante recolher das redes sociais os dados já definidos em 4.2.1 para melhor perceber a previsão da procura de modo a permitir ao retalhista otimizar a seleção e disponibilidade de artigos nas prateleiras.

A sub-área de *Otimização dos packs* consiste em perceber os hábitos dos clientes e o volume das suas compras de modo a construir *packs* ótimos que satisfaçam o cliente.

Assim como a sub-área de *Stock Out e Vendas Perdidas*, a sub-área de *Otimização dos packs* está bastante relacionada com a previsão da procura e por isso será interessante extrair, para este caso, os dados já definidos em 4.2.1.

Para esta área será interessante perceber quais os tamanhos mais comprados em determinadas regiões, o que se refere a hábitos de compra dos clientes, que como já foi referido em 4.2.2, não é uma informação que seja partilhada com frequência nas redes sociais e cuja recolha poderá ser injustificada devido ao baixo volume de dados disponíveis. A partir de dados internos à própria empresa será bastante fácil obter estas informações. Mesmo não conjugando tais informações internas com dados externos, os dados internos, por si só, serão dados bastante ricos.

Para a sub-área de *Localização de Outlets*, no que se refere a dados a extrair a partir das redes sociais, será também interessante extrair das redes sociais:

- Localidade dos utilizadores que mais comentam sobre a marca;

Contudo, este tipo de informação pode nem sempre ser disponibilizada pelos utilizadores das redes sociais.

#### 4.2.4 Promoções e Preço Adequado para os Produtos

Este processo inclui as sub-áreas de *Otimização de Preços*, *Markdown Optimization* e *Planeamento de Promoções*.

Este processo e as suas respetivas sub-áreas estão relacionados com a previsão da procura, na medida em que, para efetuar o planeamento de preços de produtos de modo a que haja o maior lucro possível é importante perceber a procura pelos produtos para que o preço atribuído seja estratégico.

Especialmente para as áreas de *Otimização de Preços* e *Markdown Optimization* é importante retirar das redes sociais os dados já definidos em 4.2.1, para melhor perceber a procura. Para o caso destas duas áreas será ainda interessante recolher, a partir das redes sociais, os seguintes dados:

- Reclamações ou elogios relativamente aos preços da marca do *Retalhista A*.

No que diz respeito à área de *Planeamento de Promoções* será interessante retirar das redes sociais os seguintes dados:

- Adesão às promoções do *Retalhista A* e a promoções do *Concorrente 1*;
- Redes sociais com maior adesão;
- Adesão dos utilizadores por tipo de publicação (link, vídeo, imagem, etc).

#### 4.2.5 Reconhecimento do Nome da Marca

Este processo de negócio, como mostra a Figura 4.1 da página 36 inclui as áreas de *Social Advocacy*, *Compromisso para com a marca* e *Media Mix Optimisation*.

A área de *Social Advocacy* diz respeito ao investimento que deverá ou não ser feito em influenciadores digitais para amplificar a mensagem da marca. Para esta área será útil recolher das redes sociais os seguintes dados:

- Utilizadores com mais seguidores que mais falam acerca do *Retalhista A*.

O *Compromisso para com a marca* diz respeito à relação de fidelidade que os clientes têm para com a marca. Para melhor perceber o compromisso dos clientes para com a marca, será útil extrair as seguintes informações das redes sociais:

- Sentimentos mais comuns quando se comenta sobre a marca;
- Principais críticas e elogios dados ao *Retalhista A*;

- Número de comentários, *likes* e partilhas por dia.

Sabendo quais as principais críticas e sentimentos prevaletentes do público relativamente à marca do *Retalhista A* e relativamente a marcas concorrentes é possível ao *Retalhista A* delinear estratégias de melhoria.

No que diz respeito à gestão das redes sociais perceber qual a adesão do público às suas páginas das redes sociais bem como às suas publicações, efetuando uma comparação com marcas concorrentes permitiria ao *Retalhista A* repensar nas suas estratégias de marketing.

Para o caso da área de *Media Mix Optimisation*, os dados a ser recolhidos das redes sociais são, os dados referidos em 4.2.4 para o caso do *Planeamento de Promoções*. Assim, percebendo quais as redes sociais com maior número de seguidores, analisando o tipo de publicação às quais os utilizadores mais aderem, bem como analisar a própria adesão dos utilizadores às publicações do *Retalhista A* e a publicações do *Concorrente 1* será possível planear estratégias de marketing de modo a otimizar o retorno neste tipo de investimentos.



## Capítulo 5

# Resultados

### 5.1 Resultados obtidos com o *DDP*

A ferramenta de *DDP*, foi utilizada, para extrair os dados das redes sociais e para efetuar uma primeira análise desses dados, incluindo a análise de sentimentos aos comentários extraídos.

É de salientar que devido a questões de confidencialidade e de sensibilidade do próprio *software*, o contacto direto com esta ferramenta não foi permitido, assim como também não foi permitida a partilha de qualquer documentação específica acerca desta mesma ferramenta.

Assim sendo, o contacto tido com o *DDP* foi unicamente do ponto de vista do cliente, o que significa que os dados foram obtidos a partir do contacto com um elemento da equipa de *DDP*, que por sua vez tinha contacto direto com a restante equipa, assim como com a ferramenta em questão. Através deste contacto, foram transmitidos os *URLs* das páginas das redes sociais, que incluem dados que se pretendiam extrair. Todo o contacto necessário a ter com a ferramenta de *DDP*, para efetivamente extrair os dados provenientes das redes sociais, foi da responsabilidade da equipa de *DDP*. Deste modo, foram obtidos dados relativamente ao *Retalhista A* e aos seus principais concorrentes. A Figura C.1 da página 67 é um exemplo dos ficheiros que são enviados pela equipa de *DDP*. Nesse caso em específico são mostrados dados acerca do *Retalhista A*, no entanto, para os seus concorrentes, os ficheiros têm exatamente o mesmo formato, (contendo apenas dados diferentes). É de notar que alguns dos dados não foram mostrados, de forma a garantir a confidencialidade do cliente da *Wipro*, em questão.

Para que a equipa de *DDP* extraísse os dados pretendidos foram-lhes fornecidos os *URLs* das páginas do *Retalhista A*, assim como dos seus concorrentes, de onde se pretendiam extrair os dados.

#### 5.1.1 Descrição do Processo Manual Estabelecido

Foi acordado, com a equipa de *DDP*, um processo manual, mostrado na Figura 5.1 da página 44. Este processo começa pelo envio, à equipa de *DDP*, dos *URLs* das páginas específicas, das redes sociais, de onde se pretende extrair dados. Tendo recebido os dados, cabe à equipa de

*DDP* enviar, semanalmente, três ficheiros *.csv* relativos ao *Retalhista A* e a dois dos seus concorrentes. Como mostra a Figura 5.1, após ter-se recebido, por email os ficheiros *.csv*, estes são carregados num servidor *Linux*. Para isso, é utilizado o *Windows Secure CoPy*, (*WinSCP*), que se trata de um programa capaz de efetuar, neste caso, a transferência segura de ficheiros entre o computador utilizado e o servidor *Linux*. Após serem transferidos para o servidor *Linux* os ficheiros são carregados numa base de dados *SQL*, através do *Putty* - software de emulação de terminal.

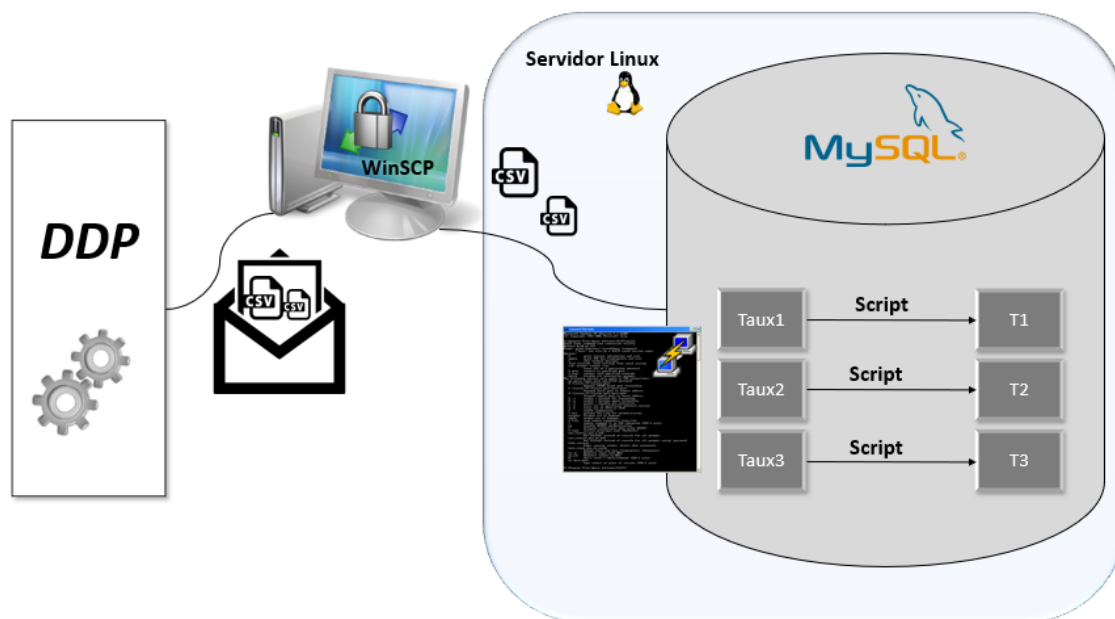


Figura 5.1: Modelo *As Is* do Processo Manual Estabelecido

É de ter em conta ainda, que os ficheiros recebidos em cada semana, contêm todos os dados já recebidos em semanas anteriores. Na Figura 5.1 estão representadas dentro da base de dados três tabelas auxiliares relativas ao *Retalhista A* e a cada um dos seus dois concorrentes. Essas três bases de dados servem para receber os dados dos ficheiros enviados pela equipa de *DDP*. Após terem sido devidamente tratados, estes dados são transferidos para as tabelas *T1*, *T2* e *T3*, representadas na Figura 5.1. Em cada semana este processo repete-se atualizando-se sempre os dados já armazenados nas tabelas *T1*, *T2* e *T3* e apagando-se os dados das tabelas auxiliares, para que em cada semana estas estejam sem quaisquer dados.

O tratamento efetuado nos dados não estruturados presentes nas tabelas auxiliares, consiste em eliminar os dados já existentes nas tabelas *T1*, *T2* e *T3* e consiste em eliminar comentários efetuados pela própria página do *Retalhista A*.

No final de cada semana, as tabelas *T1*, *T2* e *T3* possuem dados atualizados, não repetidos e produzidos apenas por utilizadores.

A vantagem de se utilizar um servidor para armazenar os dados, é que deste modo o segundo bloco poderá, posteriormente aceder aos dados guardados na base de dados.



## 5.2 Resultados obtidos com o *Buzzmonitor*

As análises a efetuar através do *Buzzmonitor* serão feitas da melhor forma possível para ir de encontro aos dados que se pretendem extrair das redes sociais, definidos ao longo do Capítulo 4.

O primeiro passo a dar no sentido da monitorização do *Retalhista A* bem como do *Concorrente 1*, através do *Buzzmonitor* é criar as chamadas "*brands*".

Na criação de "*brands*" há uma secção destinada a seleccionar de onde se pretende procurar informação acerca dessa marca, nessa secção é possível adicionar as páginas oficiais (somente do *Facebook* e *Youtube*) dos retalhistas a analisar, é possível seleccionar as línguas e os países de onde se pretende procurar informações e ainda é possível procurar informações em blogues. Como neste caso se está a utilizar um pacote *freemium* não é possível recolher informações de blogues.

Foram criadas duas "*brands*", a marca do *Retalhista A* e a marca do *Concorrente 1*. Aquando da definição das "*brands*" foram fornecidas informações acerca do *Retalhista A* e do *Concorrente 1*. Foi fornecido o *URL* das páginas oficiais, do *Facebook*, dos dois retalhista, foi disponibilizado o *URL* do canal do *Youtube* de cada um dos retalhistas e para além disto foi ainda fornecido o *URL* de determinadas páginas de fãs, do *Facebook*. Foram ainda formuladas algumas palavras-chave para serem procuradas nas redes sociais. Para cada um dos retalhistas, as palavras-chave definidas foram bastante gerais, por exemplo, para o caso do *Retalhista A* e do *Concorrente 1* foram formuladas *queries* como as seguintes:

- *Retalhista A* OR *RA*.
- *Concorrente 1*.

O *Buzzmonitor* é capaz de lidar com variações ortográficas, por isso, aquando da formulação de *queries* não é necessário fazer a distinção entre maiúsculas e minúsculas ou entre cedilhas e acentos. Contudo, é de ter em conta a importância dos parêntesis para restringir determinadas pesquisas.

Para o caso de se pretender usufruir do *Social CRM* (abordado em 3.3) será necessário sincronizar a conta do *Buzzmonitor* com contas do *Twitter* e *Facebook*. Como neste estudo não se pretende comunicar com clientes não fará sentido ter em conta esta funcionalidade. Serão apenas exploradas as funcionalidades de *Social Analytics* e de *Monitoring* do *Buzzmonitor*.

### 5.2.1 Principais tópicos (acerca de moda) mais comentados e sentimentos prevalentes

Para esta análise, foram procurados os *hashtags* mais utilizados pelos utilizadores das redes sociais, para isso, foram aplicados os seguintes filtros:

- Foram seleccionadas como fonte de pesquisa as redes sociais *Twitter*, *Facebook*, *Youtube* e *Instagram*;
- Foram recolhidos dados produzidos nos últimos 30 dias;

- Foram recolhidos comentários produzidos por todo o género de utilizadores (masculino, feminino e desconhecido);
- Foram recolhidos comentários com todo o tipo de sentimentos (positivo, negativo e neutro);

Filtros aplicados ao *Facebook*, *Instagram*, *Twitter* e *Youtube*:

- Foram selecionados comentários e publicações efetuadas apenas por utilizadores;
- Foram selecionados todos os idiomas;
- Foram selecionados todo o tipo de publicações (fotografia, vídeo, *URLs*, etc).

No que diz respeito às variáveis, estas foram agrupadas por *hashtag* e divididas consoante o sentimento. No gráfico Figura D.2 da página 70 é ainda possível visualizar o número de comentários existentes que contêm cada *hashtag*.

Na Figura D.2 da página 70 mostra-se uma análise geral acerca dos *hashtags* mais usados pelos utilizadores bem como o sentimento prevalente. Analisando a Figura D.2 da página 70 verificam-se termos que são interessantes para esta análise, nomeadamente, "*fashion*" e "*trend*". Sendo o *Buzzmonitor* uma ferramenta interativa, é possível clicar em cima dessas duas barras e melhor analisar os comentários que contêm as *hashtags* "*fashion*" e "*trend*".

Contudo, foi feita uma análise mais específica, adicionando no motor de pesquisa a *query* "*fashion OR trend*". O resultado desta análise é mostrado na Figura D.4 da página 70.

Analisando os dois gráficos verifica-se ser fundamental o facto do *Buzzmonitor* permitir uma análise mais profunda dos resultados obtidos nos gráficos. Através dos gráficos é possível visualizar informações bastante gerais e nem sempre sugestivas. Clicando nas barras do gráficos é possível encontrar comentários mais sugestivos como o comentário presente na Figura D.3 da página 70.

### 5.2.2 Adesão do público às promoções do *Retalhista A* e dos seus competidores

O *Buzzmonitor* tem uma funcionalidade, indisponível para utilizadores *freemium*, que permite seguir determinadas publicações da marca, no *Facebook*. O que significa que seria possível analisar constantemente determinadas publicações da marca, ou de marcas concorrente. Neste caso, esta funcionalidade seria útil para seguir publicações promocionais, comparando essas publicações com publicações promocionais dos concorrentes, analisando a adesão do público às mesmas.

Uma vez que está a ser utilizado o pacote *freemium* do *Buzzmonitor*, ter-se-á de procurar uma alternativa à solução acima mencionada.

Na Figura D.5 da página 71 são mostrados os resultados da análise do número de publicações efetuadas, por dia, nas páginas do *Facebook* do *Retalhista A* e do *Concorrente 1*.

Para esta análise utilizaram-se as duas *brands* inicialmente criadas e foram configurados filtros da seguinte forma:

- Como fonte de pesquisa foi utilizado o *Facebook*;

- Foram recolhidos dados produzidos nos últimos 30 dias;
- Foram recolhidas apenas as publicações efetuadas pelas marcas, na sua página de *Facebook*;

No que diz respeito às variáveis estas foram configuradas do seguinte modo:

- Os dados são agrupados por dia;
- São mostrados no gráfico o número de publicações efetuadas por cada uma das marcas, por dia;
- Os dados são divididos por marca.

Esta análise possibilita que, clicando em cada barra, laranja ou azul, da Figura D.5 da página 71, se visualizem as publicações em concreto, por ordem de "engagement". Como mostra na Figura D.5 da página 71 rodeado a encarnado, é possível ordenar as publicações por ordem de "engagement". Desta forma, é possível comparar, diariamente, as publicações do *Retalhista A* e do *Concorrente 1*.

Utilizando a secção de *Social Analytics*, para o *Facebook*, é possível ainda a criação do gráfico mostrado na Figura D.6 da página 71. Este gráfico é relativo ao *engagement* que se trata de um conceito utilizado para medir o quanto os fãs e utilizadores das redes sociais se envolvem com uma marca. No *Facebook*, observar o *engagement*, significa acompanhar a participação dos utilizadores nas publicações de uma marca. É através do total de *likes*, partilhas e comentários a uma publicação que se quantifica esse envolvimento. Quanto maior o envolvimento dos fãs com uma publicação, maior o *engagement*, [8]. A equação 5.1 mostra como é calculado o *engagement* pelo *Buzzmonitor*, [8].

$$Engagement = \frac{\text{Interações}(\text{likes} + \text{coments} + \text{partilhas}) * 100}{\frac{\text{N}^\circ \text{de posts no período}}{\text{Total de seguidores no período}}} \quad (5.1)$$

### 5.2.3 Sentimentos mais comuns quando se comenta sobre o *Retalhista A* e sobre o *Concorrente 1*

Para as análises relativas aos sentimentos mais comuns quando se comenta sobre o *Retalhista A* e o *Concorrente 1*, foram selecionadas as marcas já criadas numa fase anterior (a marca *Retalhista A* e a marca *Concorrente 1*) bem como toda a informação submetida aquando da criação destas duas marcas e foram efetuadas as seguintes configurações nos filtros:

- Como fonte de pesquisa foi utilizado o *Twitter*, *Facebook*, *Youtube* e *Instagram*;
- Foram recolhidos dados produzidos entre 1 de janeiro de 2018 até 1 de maio de 2018;
- Foram recolhidos comentários produzidos apenas por utilizadores;

- Foram recolhidos comentários de todo o género de utilizadores (masculino, feminino e desconhecido);
- Foram recolhidos comentários com todo o tipo de sentimentos (positivo, negativo e neutro);
- No motor de pesquisa foi procurado o nome "*Retalhista A*" ou "*Concorrente 1*" (conforme o caso).

No que diz respeito às variáveis estas foram agrupadas por sentimento e divididas por marca, são ainda mostrados no gráfico o número de publicações efetuadas para cada tipo de sentimento.

A Figura D.7 da página 72 mostra a análise de sentimentos dos clientes acerca do *Retalhista A*.

Como mostrado na Figura D.7 da página 72, a grande maioria dos comentários efetuados acerca deste retalhista são positivos. Clicando em cima de cada uma das três barras da Figura D.7 da página 72 é possível visualizar cada comentário em específico, bem como aceder à publicação original. Neste sentido e após ter-se clicado em cada uma das barras da Figura D.7 da página 72 mostram-se, como exemplo, na Figura D.8 da página 73 três comentários recolhidos acerca do *Retalhista A*, sendo possível, através das três cores mostradas, para cada comentário presente na Figura D.8 da página 73, perceber a polaridade do sentimento em cada um dos comentários. Assim, analisando as cores relativas à classificação de sentimentos, nos três comentários, verifica-se que o primeiro comentário é classificado como positivo (verde), o segundo comentário é neutro (laranja) e o terceiro é negativo (vermelho).

Na Figura D.9 da página 73 são mostrados os resultados da análise aos sentimentos dos clientes acerca do *Concorrente 1*. Tal como para a análise dos sentimentos dos clientes acerca do *Retalhista A*, é também possível, clicando em cada uma das barras da Figura D.9 da página 73 analisar cada comentário efetuado.

Analisando o gráfico presente na Figura D.9 da página 73 verifica-se que existem mais comentários positivos sobre o *Concorrente 1*, do que negativos. No entanto, comparando com os resultados obtidos para o *Retalhista A*, a partir das mesmas fontes de dados, verifica-se que no caso do *Concorrente 1*, existem muito menos comentários recolhidos.

É de salientar que as fontes de dados configuradas para cada um dos casos, (para o caso do *Retalhista A* e do *Concorrente 1*), são as mesmas, ou seja, para ambos os casos a informação foi retirada das mesmas redes sociais, no entanto, as páginas oficiais das respetivas marcas e as palavras-chave, para cada um dos casos são diferentes.

Para o caso do *Concorrente 1*, foram também selecionados três comentários, dos vários comentários extraídos para este caso e são mostrados na Figura D.10 da página 74, estes, são comentários obtidos clicando em cada uma das barras da Figura D.9 da página 73.

#### **5.2.4 Principais críticas e elogios dados ao *Retalhista A***

No sentido de tentar descobrir as críticas e elogios dados ao *Retalhista A* foi feita a análise mostrada na Figura D.11 da página 74.

Para efetuar esta análise configurou-se os filtros da ferramenta, da seguinte forma:

- Como fonte de pesquisa foi utilizado o *Twitter, Facebook, Youtube e Instagram*;
- Foram recolhidos dados produzidos entre 1 de janeiro de 2018 até 1 de maio de 2018;
- Foram recolhidos comentários produzidos apenas por utilizadores;
- Foram recolhidos comentários de todo o género de utilizadores (masculino, feminino e desconhecido);
- Foram recolhidos comentários com todo o tipo de sentimentos (positivo, negativo e neutro);
- No motor de pesquisa foi procurada *query "Retalhista A OR RA"*;

Os dados recolhidos foram agrupados pelos termos e divididos pelo sentimento.

Os tópicos expostos no gráfico não são muito sugestivos. No entanto, sendo o *Buzzmonitor* uma plataforma interativa, clicando nas barras da Figura D.11 da página 74 é possível visualizar os comentários em maior detalhe. A parte das barras a vermelho será relativa a críticas, pois é representada comentários cujo sentimento é negativo, a parte da barra a verde será relativa a elogios pois representa comentários com sentimento positivo. A cor laranja representa comentários cujo sentimento é neutro.

### 5.2.5 Principais reclamações e sugestões sobre uma determinada loja

A fim de descobrir as principais reclamações e sugestões relativamente ao desempenho das lojas do *Retalhista A* foi feita a análise mostrada na Figura D.12 da página 75.

Para efetuar esta análise foram feitas configurações nos filtros, como as seguintes:

- Como fonte de pesquisa foi utilizado o *Twitter, Facebook, Youtube e Instagram*;
- Foram recolhidos dados produzidos desde 1 de janeiro de 2018 até 21 de maio de 2018;
- Foram recolhidos comentários produzidos apenas por utilizadores;
- Foram recolhidos comentários de todo o género de utilizadores (masculino, feminino e desconhecido);
- Foram recolhidos comentários com todo o tipo de sentimentos (positivo, negativo e neutro);
- No motor de pesquisa foi procurado o termo *"store"*;

No que diz respeito às variáveis estas foram agrupadas pelos termos e divididas pelo sentimento.

Os tópicos expostos no gráfico não são muito sugestivos. No entanto, tal como na análise efetuada em 5.2.4, clicando nas barras da Figura D.12 da página 75 é possível visualizar os comentários em maior detalhe. A parte das barras a vermelho será relativa a críticas ou reclamações,

pois representa comentários cujo sentimento é negativo, a parte da barra a verde será relativa a elogios ou sugestões pois representa comentários com sentimento positivo. A cor laranja representa comentários cujo sentimento é neutro.

Na Figura D.13 da página 75 estão presentes três comentários exemplo, recolhidos da análise presente na Figura D.12 da página 75. Estes comentários foram recolhidos da barra relativa ao termo "*store*" da Figura D.12 da página 75.

### 5.2.6 Críticas dos clientes relativamente à alocação do *staff* da loja

De modo a descobrir quais as principais críticas relativas à alocação do *staff* nas lojas do *Retailista A* começou-se por efetuar uma análise mais geral, mostrada na Figura D.14 da página 76.

Para efetuar esta análise configurou-se os filtros da seguinte forma:

- Como fonte de pesquisa foi utilizado o *Twitter*, *Facebook*, *Youtube* e *Instagram*;
- Foram recolhidos dados produzidos desde 1 de janeiro de 2018 até 1 de junho de 2018;
- Foram recolhidos comentários produzidos apenas por utilizadores;
- Foram recolhidos comentários de todo o género de utilizadores (masculino, feminino e desconhecido);
- Foram recolhidos comentários com todo o tipo de sentimentos (positivo, negativo e neutro);
- No motor de pesquisa foi procurado o termo "*staff*";

Quanto às variáveis estas foram agrupadas pelos termos e divididas pelo sentimento.

Uma vez mais, os tópicos expostos no gráfico não são muito sugestivos. No entanto, clicando nas barras da Figura D.14 da página 76 é possível visualizar os comentários em maior detalhe. Na Figura D.15 da página 76 é possível visualizar uma pequena amostra dos comentários que surgem nesta análise. Tendo sido feita uma análise aos comentários que surgiram, verifica-se que são comentários muito gerais relativamente ao *staff* das lojas. Tentou-se, através de uma análise mais restrita, obter resultados mais específicos, para isso no motor de pesquisa colocou-se a *query* "(*staff* AND *wait*) OR (*staff* AND *time*)". A Figura D.16 da página 76 mostra esta análise. Clicando em cada uma das barras, como mostrado na Figura D.17 da página 77 é possível encontrar comentários mais específicos.

### 5.2.7 Localidade dos utilizadores que mais comentam sobre a marca

Devido a políticas de privacidade apenas foi possível a recolha da localidade dos utilizadores para o caso do *Twitter*. Para esta análise, os filtros foram configurados da seguinte forma:

- Como fonte de pesquisa foi utilizado o *Twitter*;
- Foram recolhidos dados produzidos desde 1 de janeiro de 2018 até 31 de maio de 2018;

- Foram recolhidos comentários produzidos apenas por utilizadores;
- Foram recolhidos comentários de todo o género de utilizadores (masculino, feminino e desconhecido);
- Foram recolhidos comentários com todo o tipo de sentimentos (positivo, negativo e neutro);
- No motor de pesquisa foi procurado o termo "*staff*";

No que diz respeito às variáveis, os dados recolhidos foram agrupados por localidade e no gráfico é mostrado o número de publicações para cada comentário respetivo a cada localidade. Na Figura D.18 da página 77 é mostrado o resultado da análise relativa à localidade dos utilizadores que mais comentam sobre a marca.

### 5.2.8 Reclamações ou elogios relativamente aos preços da marca do *Retalhista A*

A fim de descobrir as principais reclamações e elogios relativamente ao preço dos produtos do *Retalhista A* foi feita a análise mostrada na Figura D.19 da página 77.

Para efetuar esta análise foram aplicados os filtros seguintes:

- Como fonte de pesquisa foi utilizado o *Twitter, Facebook, Youtube e Instagram*;
- Foram recolhidos dados produzidos desde 1 de janeiro de 2018 até 1 de junho de 2018;
- Foram recolhidos comentários produzidos apenas por utilizadores;
- Foram recolhidos comentários de todo o género de utilizadores (masculino, feminino e desconhecido);
- Foram recolhidos comentários com todo o tipo de sentimentos (positivo, negativo e neutro);
- No motor de pesquisa foi procurado o termo "*price*";

Para esta análise, os dados recolhidos foram agrupados pelos termos e divididos pelo sentimento.

Os tópicos expostos no gráfico não são muito sugestivos. No entanto, clicando nas barras da Figura D.19 da página 77 é possível visualizar os comentários em maior detalhe.

Na Figura D.20 da página 78 é feita uma análise semelhante à realizada em 5.2.3. Para este caso no motor de pesquisa foi introduzida a *query "price"* e foram obtidos os resultados mostrados na Figura D.20 da página 78.

### 5.2.9 Redes sociais com maior adesão

A Figura D.21 da página 78 mostra o volume de comentários e publicações, dos utilizadores, em cada rede social. Assim, é possível identificar, para o caso do *Retalhista A*, em que redes sociais o público está mais presente.

- Como fonte de pesquisa foi utilizado o *Twitter, Facebook, Youtube e Instagram*;
- Foram recolhidos dados produzidos nos últimos 30 dias;
- Foram recolhidos comentários produzidos apenas por utilizadores;
- Foram recolhidos comentários de todo o género de utilizadores (masculino, feminino e desconhecido);
- Foram recolhidos comentários com todo o tipo de sentimentos (positivo, negativo e neutro);

Os dados recolhidos foram divididos por rede social e agrupados por dia.

### **5.2.10 Adesão dos utilizadores por tipo de publicação**

Utilizando a secção de *Social Analytics* foi possível obter os resultados presentes na Figura D.22 da página 78, que mostra o tipo de publicações, do *Retalhista A* com maior adesão por parte dos utilizadores. O tipo de publicação pode ser *link*, vídeo, imagem, evento, etc.

### **5.2.11 Utilizadores com mais seguidores que mais falam acerca do *Retalhista A***

Para efetuar esta análise, cujo resultado é mostrado na Figura D.23 da página 79, foram aplicados os seguintes filtros:

- Como fonte de pesquisa foi utilizado o *Twitter, Facebook, Youtube e Instagram*;
- Foram recolhidos dados produzidos nos últimos 30 dias;
- Foram recolhidos comentários produzidos apenas por utilizadores;
- Foram recolhidos comentários de todo o género de utilizadores (masculino, feminino e desconhecido);
- Foram recolhidos comentários com todo o tipo de sentimentos (positivo, negativo e neutro);
- Foram selecionados utilizadores com um número de seguidores maior do que 25000.

Os dados foram divididos por sentimento, para perceber se os utilizadores falam de uma forma positiva ou negativa, do *Retalhista A*. No gráfico da Figura D.23 da página 79 é possível ainda verificar que os dados foram agrupados por utilizador.

### **5.2.12 Número de comentários, *likes* e partilhas por dia.**

Para efetuar esta análise, foram configurados os filtros da seguinte forma:

- Como fonte de pesquisa foi utilizado o *Facebook*;
- Foram recolhidos dados produzidos nos últimos 30 dias;



- Foram recolhidos comentários produzidos apenas por utilizadores;
- Foram recolhidos comentários de todo o género de utilizadores (masculino, feminino e desconhecido);
- Foram recolhidos comentários com todo o tipo de sentimentos (positivo, negativo e neutro);

A inovação aplicada a esta análise consiste na forma como as variáveis são tratadas. Ao editar o relatório, é possível seleccionar a dimensão “*Atributo de Post*” e uma série de atributos como comentários, *likes* e partilhas são possíveis. É de referir que em todos os casos as variáveis foram agrupadas por dia.

Para o caso do *Facebook* foram analisados os comentários (Figura D.24 da página 79 e Figura D.25 da página 79), *likes*, (Figura D.26 da página 80 e Figura D.27 da página 80) e partilhas (Figura D.28 da página 80 e Figura D.29 da página 80) efetuados por dia, pelos utilizadores, relativamente ao *Retalhista A* e ao *Concorrente 1*.

## 5.3 Análise de Resultados

Nos seguintes sub-capítulos serão efetuadas as análises dos resultados obtidos com cada uma das duas ferramentas utilizadas neste projeto.

### 5.3.1 Análise dos Resultados Obtidos com o *DDP*

Quanto às fontes de dados, o *DDP* apenas permite a extração de dados a partir do *Twitter* e *Facebook*.

No que diz respeito aos dados recebidos, analisando o ficheiro mostrado na Figura C.1 da página 67, foram detetados alguns problemas. Relativamente ao campo “*Category*”, verificou-se que por vezes os comentários são classificados em categorias incoerentes. Embora não seja ideal a existência de alguns erros na classificação da categoria de comentários, era esperado que a categorização não fosse perfeita. Por observação manual, a ocorrência de erros no processo de classificação de categorias dos comentários, parece não ser muito frequente.

A análise de sentimentos dos comentários apresenta a classificação do sentimento do comentário como positivo, negativo ou neutro, sem nenhum tipo de pontuação ou nível de confiança associado. Assim, um comentário ligeiramente positivo ou negativo é considerado como 100% positivo ou negativo, respetivamente. Este tipo de análise perde profundidade e valor.

Observando manualmente a classificação de sentimentos, verifica-se que o algoritmo, classifica erradamente, alguns dos comentários. No entanto, é de esperar, que a classificação de sentimentos seja um indicador e não uma verdade absoluta.

Outra falha inerente ao *DDP* é que existem alguns comentários que aparecem sem o seu contexto, estes são comentários que estariam relacionados com outras publicações cuja fonte não é

disponibilizada e portanto é inacessível. Assim sendo, há alguns comentários que se tornam irrelevantes. Saber qual a fonte à qual o comentário está relacionado ajudaria a tornar a análise mais precisa, já que alguns comentários só fazem sentido sabendo o seu contexto.

Apesar de tudo, o maior problema encontrado nos dados recebidos foi o facto de que quando os comentários excedem um certo limite de tamanho (115 caracteres), são cortados e é colocado, seguidamente ao comentário, o *URL* do mesmo. Esta falha dos dados recebidos dificultará a possibilidade de automatização do processo de análise, uma vez que muitos comentários excedem o tamanho limite e não são mostrados por completo. No que diz respeito à análise de sentimentos, à classificação da categoria e a atribuição de uma palavra-chave ao comentário, estas são feitas para o comentário completo.

No que toca ao estado civil, ao sexo, à localização e ao estado parental do utilizador, verifica-se que existem bastantes comentários extraídos que não possuem tais informações acerca do autor do comentário. De facto, muitos utilizadores das redes sociais podem simplesmente escolher não partilhar este tipo de informações. Embora possa ser um fator limitador nas análises, a verdade é que quando se trabalha com dados provenientes das redes sociais, como já foi referido em 2.3.2, só se tem acesso às informações públicas dos utilizadores e só essas informações podem ser extraídas e analisadas de forma legítima.

O campo de "*Spell Corrected Text (Twitter)*" corrige os erros ortográficos que possam existir no campo "*Comment*". Como já foi referido, quando os comentários passam dos 115 caracteres são cortados e é colocado o *URL* desse mesmo comentário. Para além desta correção ser efetuada apenas para parte de um comentário (até 115 caracteres), é uma correção imperfeita na medida em que quando existem palavras incompreendidas, estas são eliminadas ou corrigidas de forma errada. Nomes de marcas e de pessoas são, muitas vezes, corrigidos erradamente. No que diz respeito à pontuação existente nos comentários, esta é, na maior parte das vezes, eliminada.

Nos comentários recebidos foi percebido que vários comentários não provinham de clientes, mas sim de funcionários que interagem com os clientes, nas páginas das redes sociais da marca. Como o objetivo é focar a análise nos comentários e opiniões de clientes, os comentários provenientes de funcionários terão de ser eliminados.

Apesar dos pontos negativos, a plataforma *DDP* tem a si associados pontos positivos. Permite extrair a data e hora de quando o comentário é efetuado, sendo esta uma informação bastante útil para a análise. A capacidade de converter a localização em coordenadas geográficas é algo que também pode ser útil quando se tenta utilizar comentários para localizar a procura. O número de amigos do utilizador que comenta é uma informação valiosa, na medida em que pode ser um indicador de quantas pessoas conseguem visualizar o comentário. O número de *likes* e as partilhas de um determinado conteúdo podem também ser considerados indicadores do número de pessoas com acesso a esse mesmo conteúdo. O campo "*Category*" é bastante interessante pois classifica os comentários em determinadas categorias/palavras-chave sendo possível perceber, de modo geral, do que se trata o comentário.

### 5.3.2 Análise dos Resultados Obtidos com o *Buzzmonitor*

Esta análise de resultados irá basear-se na experiência obtida aquando da utilização do *software Buzzmonitor*, tendo sido utilizada uma versão *free* do mesmo.

Utilizando a secção de *Social Analytics* foi possível analisar o desempenho, essencialmente do *Facebook* e dos canais de *Youtube* das marcas em análise. Através das análises utilizando o campo de *Social Analytics* foi possível perceber a evolução do número de seguidores nas páginas (comparando a evolução de páginas de diferentes marcas). Na secção de *Social Analytics* foi ainda possível perceber o compromisso dos utilizadores para com a marca e para com cada tipo de publicação efetuada pela mesma. O compromisso dos utilizadores para com uma marca é medido com base numa fórmula desenvolvida e mostrada em 5.1. No que diz respeito à secção de *Social Analytics*, seria interessante que para as restantes redes sociais existissem análises igualmente interessantes. No caso do *Twitter* apenas é possível perceber quais são as publicações com mais *retweets*, poderiam existir análises igualmente interessantes como as que existem para o *Facebook*.

No que diz respeito ao campo de *Monitoring* foi possível elaborar relatórios personalizáveis, mostrando gráficos com informações relevantes para os casos em estudo. Em cada gráfico criado, foi possível exportar os dados (utilizados para elaborar os gráficos) para um ficheiro *.xls*. É de referir que os dados poderiam ter sido exportados de outra forma, nomeadamente na forma de *PDF*. Contudo, exportou-se os dados na forma de *.xls* para que fosse possível uma melhor comparação com os dados obtidos pelo *DDP*. Na Figura D.1 da página 69 é possível observar um tipo de ficheiro exportado do *Buzzmonitor*. É de referir que mesmo para diferentes tipos de relatórios os ficheiros são semelhantes, contendo apenas e naturalmente, informações diferentes.

No que concerne à extração de dados, o *Buzzmonitor* permite a extração a partir dos *URLs* inicialmente fornecidos (aquando da criação das marcas) e possibilita ainda a extração de dados através da pesquisa de palavras-chave. O facto de permitir a procura de dados por palavras-chave é um ponto bastante positivo pois permite a seleção de dados realmente relevantes para determinadas análises excluindo a restante informação que não interessa para o estudo.

Para o caso do *Facebook*, apenas se verificaram dados extraídos a partir dos *URLs* das páginas, que foram inicialmente fornecidos. Este facto pode significar que o *Facebook* possui diversas restrições de acesso a dados, o que terá levado à dificuldade de construção de métodos de *crawling* mais elaborados, possibilitando a extração de mais informações a partir da rede social, para além das informações apenas presentes nos *URLs* de páginas específicas. Apesar destas restrições, quando se configura o *Buzzmonitor* para que sejam construídos os gráficos, é possível restringir esses dados (extraídos de páginas específicas). Efetuando pesquisas por palavras-chave, a partir dos inúmeros dados extraídos de uma página específica, é possível selecionar dados relevantes para as análises em questão. Ou seja, imagine-se que inicialmente são dados os *URLs* de três páginas de *Facebook* específicas. Imagine-se que posteriormente se pretende elaborar um gráfico que mostre o que as pessoas comentam acerca do preço do *Retalhista A*, procurando por "preço" é possível obter comentários provenientes dessas páginas, cujos *URLs* foram inicialmente fornecidos e que incluam a palavra "preço".

Para o caso do *Twitter* e do *Instagram* a pesquisa foi efetuada, desde início (aquando da criação das marcas), com base em palavras-chave. Para o caso destas duas redes sociais, para além de ter sido possível a extração de conteúdo, que inclui as palavras-chave, identificadas através de *hashtags* ou de arroba, foi ainda possível a extração de conteúdo em que eram mencionadas as palavras-chave mesmo sem estarem identificadas de alguma forma. Aquando da elaboração dos gráficos, assim como aconteceu para o caso do *Facebook*, foi possível restringir ainda mais a pesquisa, uma vez que nessa fase é também possível procurar dados por palavras-chave.

Resumidamente, inicialmente são dadas algumas fontes de pesquisa (por *URLs* ou por palavras-chave mais genéricas, como por exemplo pesquisar apenas o nome do retalhista) e posteriormente, (na elaboração dos gráficos) é possível restringir ainda mais os dados.

No que diz respeito à elaboração dos relatórios personalizáveis, foi possível efetuar uma série de configurações de modo a obter os dados da forma desejada. Existem diversas combinações possíveis no que diz respeito ao agrupamento e divisão dos dados nos gráficos assim como existem alguns tipos de gráficos a utilizar. Aquando da criação de gráficos é ainda possível efetuar comparações entre marcas, apresentando juntamente, gráficos com dados relativos a cada uma das marcas. Ainda na elaboração dos gráficos é possível aplicar inúmeros filtros aos dados a recolher. Para além de ser possível, nesta fase, restringir ainda mais os dados, através da pesquisa por palavras-chave é ainda possível customizar as datas dos dados que se deseja extrair. É possível configurar a extração de modo a que sejam extraídos só dados produzidos pelos utilizadores ou só dados da própria página, no que diz respeito a este tópico, um ponto negativo do *Buzzmonitor* é que mesmo configurando os relatórios para que sejam extraídos apenas comentários por parte do utilizador, por vezes, foram encontrados comentários da própria página em resposta a comentários de determinados utilizadores. Nesta fase de configuração dos filtros a aplicar aos dados há inúmeras formas de filtrar os dados, entre outros pontos positivos destacam-se o facto de ser possível configurar a plataforma de modo a extrair dados produzidos por utilizadores com um determinado número de seguidores e por ser possível procurar dados produzidos por pessoas de uma determinada região ou de um determinado género. É de destacar ainda que é possível procurar os comentários pelo seu sentimento.

No que se refere à análise de sentimentos, o sistema não é perfeito, mas como foi já referido, a análise de sentimentos deve ser tida em conta apenas como um indicador. Ainda assim, é de destacar que a análise de sentimentos falha, maioritariamente, perante ironias. É de ter em consideração que esta falha é um grande desafio para a grande maioria dos sistemas de análise de sentimentos - até para nós humanos torna-se difícil, por vezes, perceber ironias e sarcasmos.

Quanto à classificação dos sentimentos dos comentários, esta aparece associada a cada comentário de forma simplesmente qualitativa ou seja, cada comentário é classificado como positivo, negativo ou neutro, sem que lhe esteja associado algum tipo de pontuação ou nível de confiança. Este tipo de análise perde profundidade e valor. Na Figura 5.2 é mostrado um comentário classificado, erradamente como positivo. Analisando a fórmula de classificação de sentimentos utilizada pelo *Buzzmonitor* e mostrada em 3.3.1, percebe-se que um comentário como o representado na Figura 5.2 tem potencial para ser classificado como positivo e como negativo pois contém palavras

negativas como "bad" mas contém ainda palavras positivas como "best" e "agree". A existência de um nível de confiança ou de uma pontuação associada permitiria perceber que um comentário como o da Figura 5.2 não é 100% positivo, (nem perto disso).



Figura 5.2: Comentário classificado, erradamente, como positivo

No que concerne a dados pessoais dos utilizadores como por exemplo a localização, o sexo ou a própria fotografia de perfil do utilizador em questão, verificou-se que apenas foi possível retirar-se essa informação do *Twitter*. Ainda assim, deve ter-se em consideração que muitas informações podem ser, por escolha do próprio utilizador, privadas. Uma vez mais se comprova que o acesso e extração dos dados do *Facebook* se assemelha uma tarefa mais difícil do que, por exemplo, para o *Twitter*.

Por último, é de salientar o facto de se estar a utilizar uma versão *free* desta plataforma, isto porque, uma versão *premium* permitiria que os resultados fossem obtidos em tempo real e esatria a ser constantemente atualizados. Assim, os gráficos obtidos nas análises efetuadas ao longo deste projeto seriam ainda mais interessantes e valiosas.

### 5.3.3 Análise Comparativa entre os dados Extraídos pelas duas ferramentas

Esta análise comparativa será efetuada comparando os ficheiros resultantes da utilização de cada uma das ferramentas, mostrados na Figura C.1 da página 67 e na Figura D.1 da página 69. Nesta fase não seria justo comparar as duas ferramentas de outra forma. Numa fase posterior, quando tiver sido concluído o terceiro bloco deste projeto, poder-se-á efetuar uma análise comparativa considerando para isso todo o potencial de análise de dados e de criação e personalização de relatórios do *Buzzmonitor*.

Analisando os dois ficheiros extraídos verificou-se que o *Buzzmonitor* apresenta alguns pontos positivos face aos dados extraídos pelo *DDP*. Na ferramenta de *DDP* foi notória a existência de comentários associados a um contexto, porém a fonte às quais os comentários estariam associados não é guardada. No ficheiro exportado do *Buzzmonitor* verificou-se que quando existem comentários associados a outros comentários a fonte é guardada. Observando a Figura D.1 da página 69 e analisando o conteúdo do próprio ficheiro verificou-se que o campo *URL* mostra o comentário extraído, o campo *ORIGIN PAGE URL* mostra o comentário ao qual o comentário extraído está a responder, sendo a partir daí possível visualizar toda a discussão. Quando os comentários extraídos pelo *DDP* excedem os 115 caracteres e à sua frente é apresentado um *URL* para que seja possível ler o comentário por completo, é possível, a partir desse *URL* perceber (caso seja exista e seja necessário) o contexto no qual o comentário está inserido. No entanto, mais do que um ponto positivo isto trata-se de uma falha.

Relativamente ao facto de que, utilizando o *DDP*, os comentários são cortados quando excedem os 115 caracteres tal facto não se verifica nos dados extraídos do *Buzzmonitor*. Para o caso do *Buzzmonitor* os comentários são extraídos por completo e aparece um *URL* à frente do comentário caso este esteja associado a uma imagem ou vídeo. O *DDP* para além de não guardar *URLs* de imagens ou vídeos associados aos comentários corta os comentários quando estes possuem demasiados caracteres.

Os pontos positivos do *DDP* face ao *Buzzmonitor* são o facto do *DDP* possuir uma secção denominada *Keywords*, nessa secção são atribuídas palavras-chave aos comentários. Para além disto, o *DDP* classifica os comentários em categorias (classificação mais geral do que a classificação por palavras-chave).

## Capítulo 6

# Conclusões e Trabalho Futuro

Neste capítulo são apresentadas as conclusões relativas a todo o trabalho efetuado nesta dissertação bem como é apresentado o trabalho a realizar futuramente.

### 6.1 Conclusões

Analisando os dois ficheiros provenientes das duas ferramentas utilizadas neste projeto, verificou-se que existem alguns desafios quando se tenta recolher dados provenientes das redes sociais. Em primeiro lugar, na verdade, recolhem-se os dados tornados públicos pelos utilizadores. Em diversos casos foi verificado que dados pessoais como o género ou a localização não eram partilhados pelos utilizadores e por isso tornou-se impossível aceder a este tipo de informações, (pelo menos de forma legítima). Assim, a informação a recolher é limitada ao que é tornado público. Para além deste tipo de limitações as próprias redes sociais mostraram nem sempre permitir fácil acesso aos dados. Especialmente para o caso do *Facebook*, verificou-se que não se assemelhava fácil procurar palavras-chave na rede social, talvez por motivos de proteção de dados, apenas foi possível aceder aos dados presentes na própria página do *Facebook* das marcas analisadas neste projeto (utilizando o *Buzzmonitor*).

Quanto às ferramentas propriamente ditas, verificaram-se falhas especialmente a nível da análise de sentimentos. Nos dias de hoje, grande parte dos sistemas não são (ainda) suficientemente evoluídos para perceber, por exemplo, ironias e sarcasmos. Uma frase que contenha mais palavras positivas do que negativas, mesmo que no geral seja uma frase negativa poderá ser classificada como positiva. Portanto, as análises de sentimentos efetuadas pelas duas ferramentas apenas devem ser tidas como um indicador e não como verdade absoluta.

A nível de métodos de pesquisa, verificou-se que o *Buzzmonitor* possibilita métodos de pesquisa bastante interessantes, possibilitando pesquisas por palavras-chave e por *URLs* específicos. O *DDP* mostrou-se bastante mais limitado, possibilitando apenas efetuar pesquisas pelos *URLs* de páginas específicas. Para além disto, o *DDP* apenas recolhe dados do *Facebook* e do *Twitter*. O *Buzzmonitor* permite a recolha dos dados a partir do *Facebook*, *Twitter*, *Instagram* e *Youtube*, bem como a partir de blogs (apesar dessa funcionalidade não estar disponível em versões *freemium*).

No que se refere aos dados extraídos, é de referir que o *Buzzmonitor* permitiu a extração de comentários por inteiro, apresentado ainda o *URL* para imagens ou vídeos associados aos comentários. O *DDP* apenas possibilitou a extração de comentários até um limite de 115 caracteres. Comentários com mais de 115 caracteres eram cortados e era apresentado o *URL* para o comentário completo à frente do mesmo.

Relativamente a comentários que estavam incluídos em contexto, o *Buzzmonitor* mostrou um bom método para identificar a sua fonte, tal não aconteceu com o *DDP*. Deste modo, os comentários inseridos em determinados contextos perdiam o significado quando extraídos pelo *DDP*.

Apesar de tudo, o facto do *DDP* atribuir categorias aos comentários e o facto de atribuir ainda palavras-chave a esses mesmos comentários assemelhou-se um ponto positivo face ao *Buzzmonitor* que não apresentou esta funcionalidade no ficheiro extraído.

Esta análise comparativa entre estas duas ferramentas pode ser útil para melhorar alguns aspectos menos positivos do *DDP* face ao *Buzzmonitor*.

Quanto aos relatórios efetuados com o *Buzzmonitor* conclui-se que o facto do *Buzzmonitor* se tratar de uma ferramenta interativa acrescenta valor aos gráficos elaborados. Apesar dos gráficos nem sempre serem sugestivos, a possibilidade de analisar os dados de uma forma mais detalhada torna-se um ponto forte desta ferramenta. O facto do *Buzzmonitor* permitir restringir os dados recolhidos e de ser possível personalizar os relatórios de uma forma bastante variada faz com que esta ferramenta seja bastante competitiva no mercado. No final do projeto *Customer Xperience*, (no fim do terceiro bloco), será possível efetuar comparações com o *Buzzmonitor*, analisando-se, para tal, as suas capacidades de tratamento de dados e elaboração de *dashboards*. Na fase em que o projeto se encontra de momento, apenas foi possível comparar os ficheiros *.xls* com os dados extraídos, para cada relatório criado com os ficheiros com dados extraídos pelo *Buzzmonitor*.

Por fim, conclui-se que a continuação do projeto *Customer Xperience*, utilizando a ferramenta de *DDP* como base, é possível. Apesar da existência de algumas limitações inerentes ao *DDP*, é possível trabalhar os dados que nos chegam, de modo a entregar ao retalhista informações de valor para as suas tomadas de decisão.

## 6.2 Trabalho Futuro

Considerando o ponto de situação em que ficou esta dissertação no que diz respeito à utilização da ferramenta de *DDP*, considera-se que no futuro deverá ser automatizado o processo manual, descrito em 5.1.1. Como não são conhecidas as capacidades e características do *DDP* em detalhe, apenas se pode concluir que o trabalho futuro a realizar poderá incidir na automação do processo manual, recebendo-se os ficheiros por parte do *DDP*, de forma automática, sem ser necessário o envio de emails semanais e sem ser necessário o carregamento dos ficheiros utilizando-se o *WinSCP*. Esta ligação do *DDP* poderá ser feita através de uma aplicação que permita receber ou ir buscar os dados à ferramenta.

Como já foi referido, quando os comentários excedem os 115 caracteres são cortados pelo *DDP* e é colocado à sua frente o *URL* para esse comentário. No processo de tratamento dos dados



não estruturados, (descrito em 5.1.1) seria de incluir a separação desse *URL* do restante comentário, guardando-o numa coluna adicional. Dessa forma, seria possível em análises posteriores aceder ao comentário total.



## Anexo A

# Desenvolvimento do Projeto - Arquiteturas

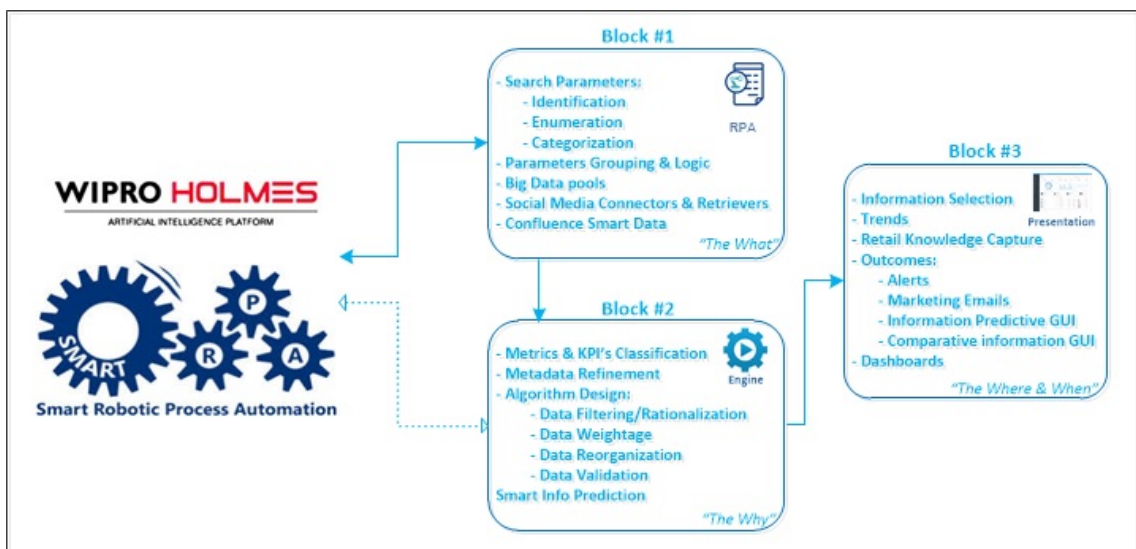


Figura A.1: Ideia inicial do projeto, utilizando o HOLMES

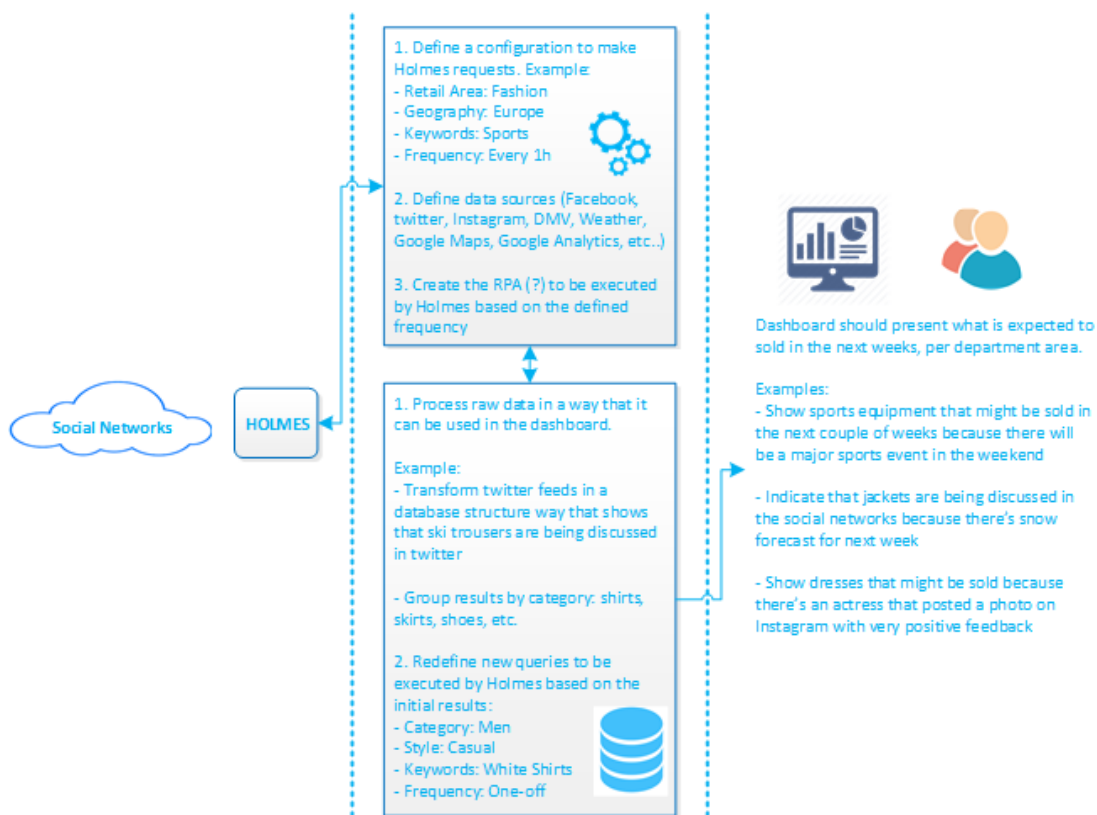


Figura A.2: Alteração da arquitetura inicial

## Anexo B

# Relação entre os Processos de negócio do *Retalhista A*

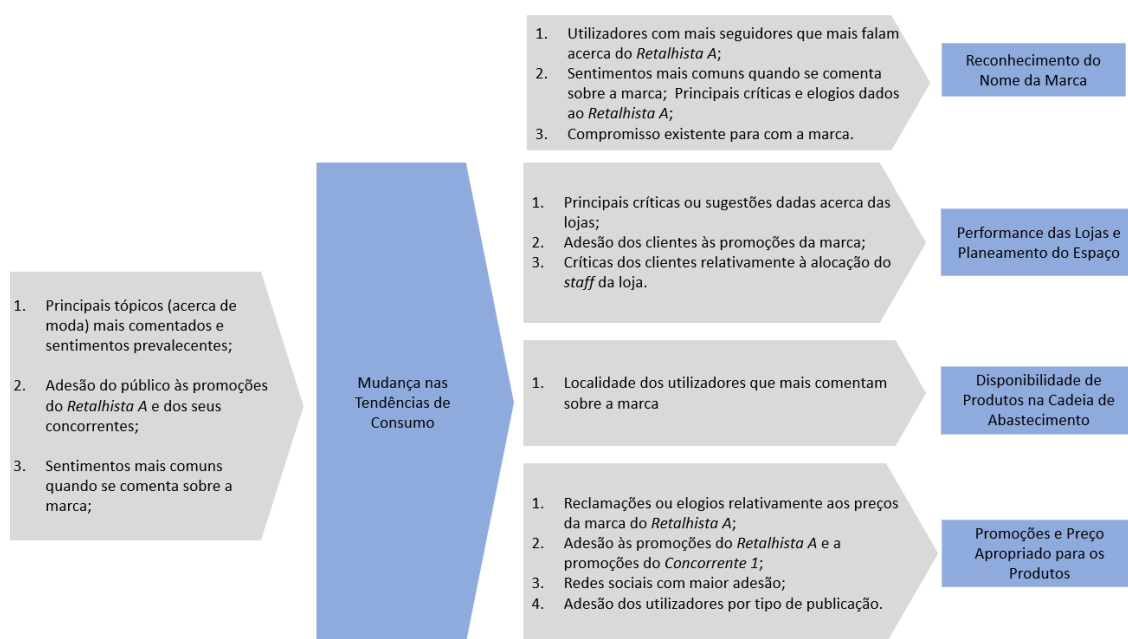


Figura B.1: Relação entre os processos de negócio do *Retalhista A* e dados a extrair para cada um dos processos de negócio



# Anexo C

## Resultados obtidos com o DDP

Dimension	Category	Media	Sentiment	Popularity	Date	Social Id	Subscriber Name	Subscriber Locatio	State	Latitude	Longitude	Comment	Spell Corrected Text (Twitter)	Bank Name
Helpdesk	Helpdesk	Twitter	Negative		07-May-2018 21:43:									
Shoes	Shoes	Twitter	Negative		07-May-2018 21:04:									
Website	Website	Twitter	Negative		07-May-2018 19:58:									
		Twitter	Neutral		07-May-2018 19:03:									
Grievances	Grievance	Twitter	Negative		07-May-2018 18:58:									
General	General	Twitter	Negative		07-May-2018 18:16:									
Store	Store	Twitter	Neutral		07-May-2018 18:13:									
		Twitter	Negative		07-May-2018 17:42:									
Social	Social	Twitter	Negative		07-May-2018 17:41:									
Delivery	Delivery	Twitter	Neutral		07-May-2018 17:36:									
Customer S	Customer	Twitter	Negative		07-May-2018 17:01:									
Commenda	Commenc	Twitter	Positive		07-May-2018 17:00:									
Commenda	Commenc	Twitter	Negative		07-May-2018 16:56:									
Store	Store	Twitter	Neutral		07-May-2018 16:27:									
Commenda	Commenc	Twitter	Neutral		07-May-2018 16:24:									
Customer S	Customer	Twitter	Neutral		07-May-2018 16:12:									
Customer S	Customer	Twitter	Neutral		07-May-2018 15:59:									
Customer S	Customer	Twitter	Negative		07-May-2018 15:58:									
General	General	Twitter	Negative		07-May-2018 15:57:									
		Twitter	Neutral		07-May-2018 15:49:									
Helpdesk	Helpdesk	Twitter	Neutral		07-May-2018 15:37:									
Message Source	Friends (Twitter)	Listed (Twitter)	Like Count	Share Count	No of Comments (Facebook)	Gender	Parenthood Status	Marital Status	Keywords	Handle	Date_Created			
Tweets	19	0	0	0	0	Male			waiting	(	07-May-2018 23:26:22			
Tweets	351	4	0	0	0	Male			shoes	(	07-May-2018 23:26:22			
Tweets	1311	48	0	0	0	Male			website	(	07-May-2018 23:26:22			
Tweets	633	1	0	0	0					(	07-May-2018 23:26:22			
Tweets	23	0	0	0	0				issue	(	07-May-2018 23:26:22			
Tweets	633	1	0	0	0				Insurance	(	07-May-2018 23:26:22			
Tweets	250	6	0	0	0	Male			store	(	07-May-2018 23:26:22			
Tweets	633	1	0	0	0					(	07-May-2018 23:26:22			
Tweets	432	0	0	0	0				twitter	(	07-May-2018 23:26:22			
Tweets	45	3	0	0	0				arrive	(	07-May-2018 23:26:22			
Tweets	32	0	0	0	0				contact	(	07-May-2018 23:26:22			
Tweets	16	1	0	0	0				Thank	(	07-May-2018 23:26:22			
Tweets	227	23	0	0	0	Male			glad	(	07-May-2018 23:26:22			
Tweets	16	1	0	0	0				store	(	07-May-2018 23:26:22			
Tweets	1192	27	0	0	0				fabulous	(	07-May-2018 23:26:22			
Tweets	32	0	0	0	0				number	(	07-May-2018 23:26:22			
Tweets	32	0	0	0	0				number	(	07-May-2018 23:26:22			
Tweets	32	0	0	0	0				called	(	07-May-2018 23:26:22			
Tweets	1372	29	0	0	0	Female			check	(	07-May-2018 23:26:22			
Tweets	32	0	0	0	0					(	07-May-2018 23:26:22			
Tweets	29	0	0	0	0				waiting	(	07-May-2018 23:26:22			

Figura C.1: Tipo de ficheiros recebidos pela equipa de DDP







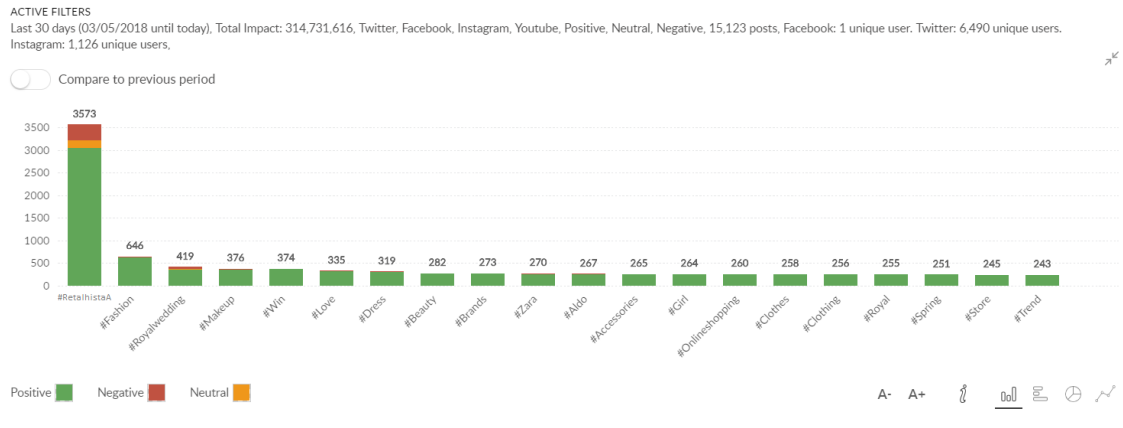


Figura D.2: Hashtags mais utilizadas pelos utilizadores

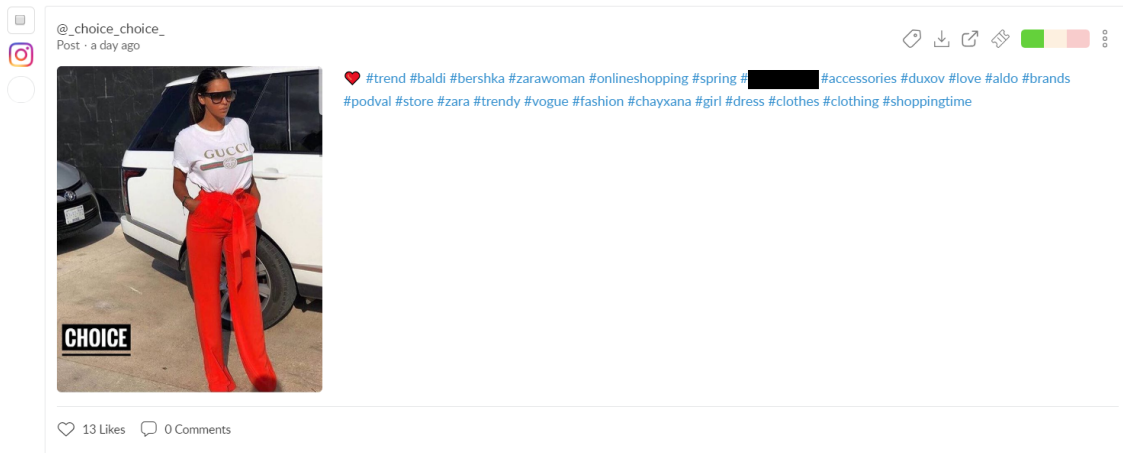


Figura D.3: Exemplo de comentários obtidos quando se clica na barra de #fashion do gráfico da Figura D.2 da página 70

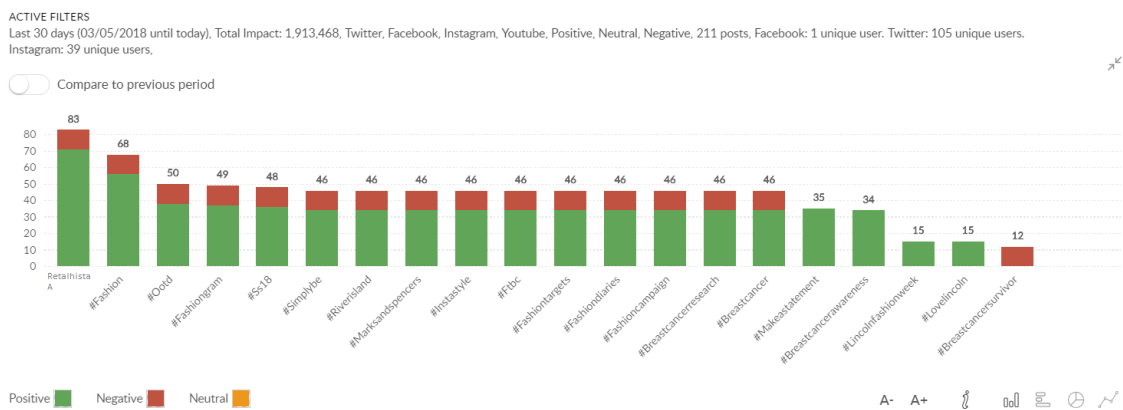


Figura D.4: Hashtags acerca do tema moda, mais usados pelos utilizadores

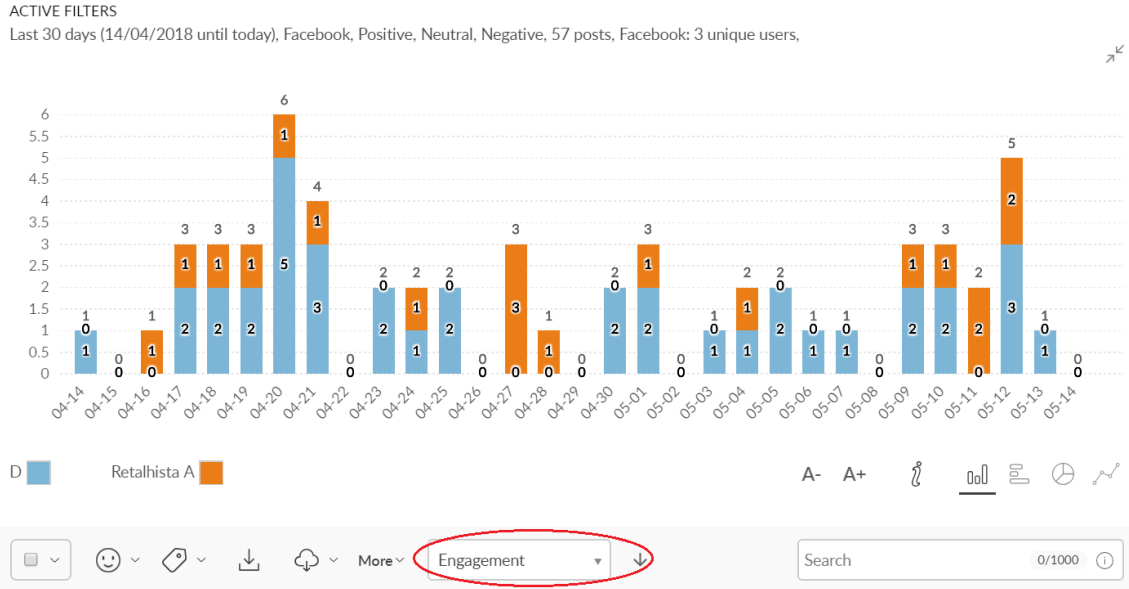


Figura D.5: Número de publicações, por dia, do *Retailista A* (a laranja) e do *Concorrente 1* (a azul)

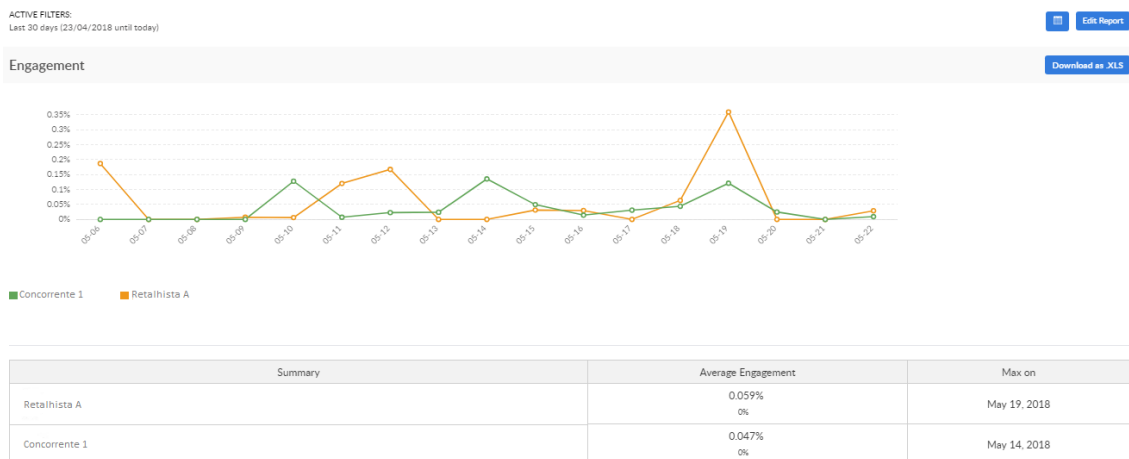


Figura D.6: *Engagement* dos utilizadores para com o *Retailista A* e o *Concorrente 1*

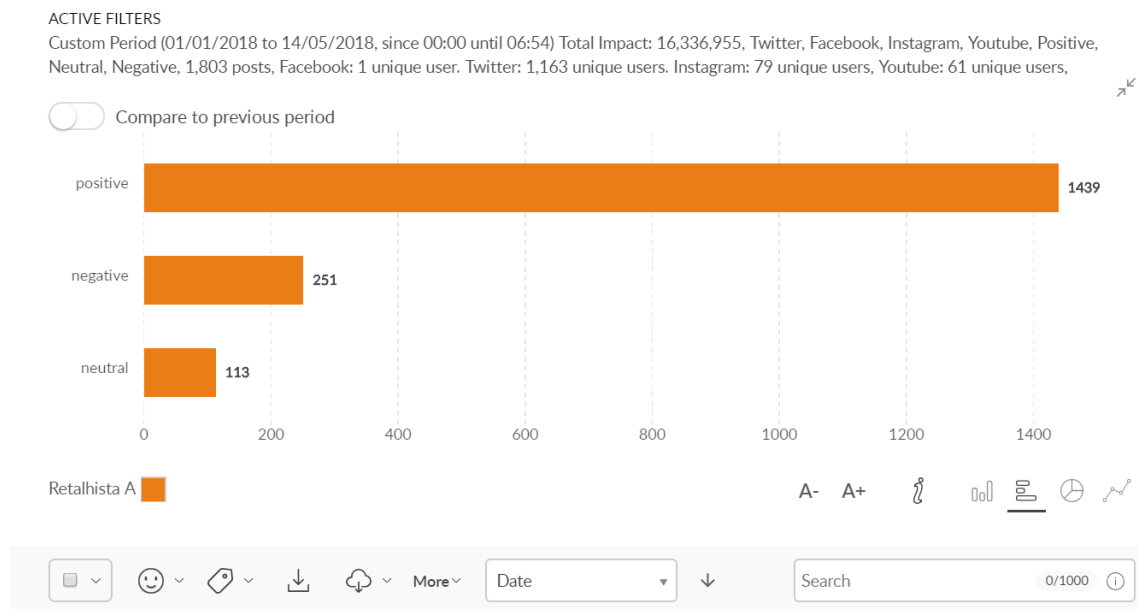


Figura D.7: Análise de sentimentos expressos pelos utilizadores acerca do *Retalhista A*

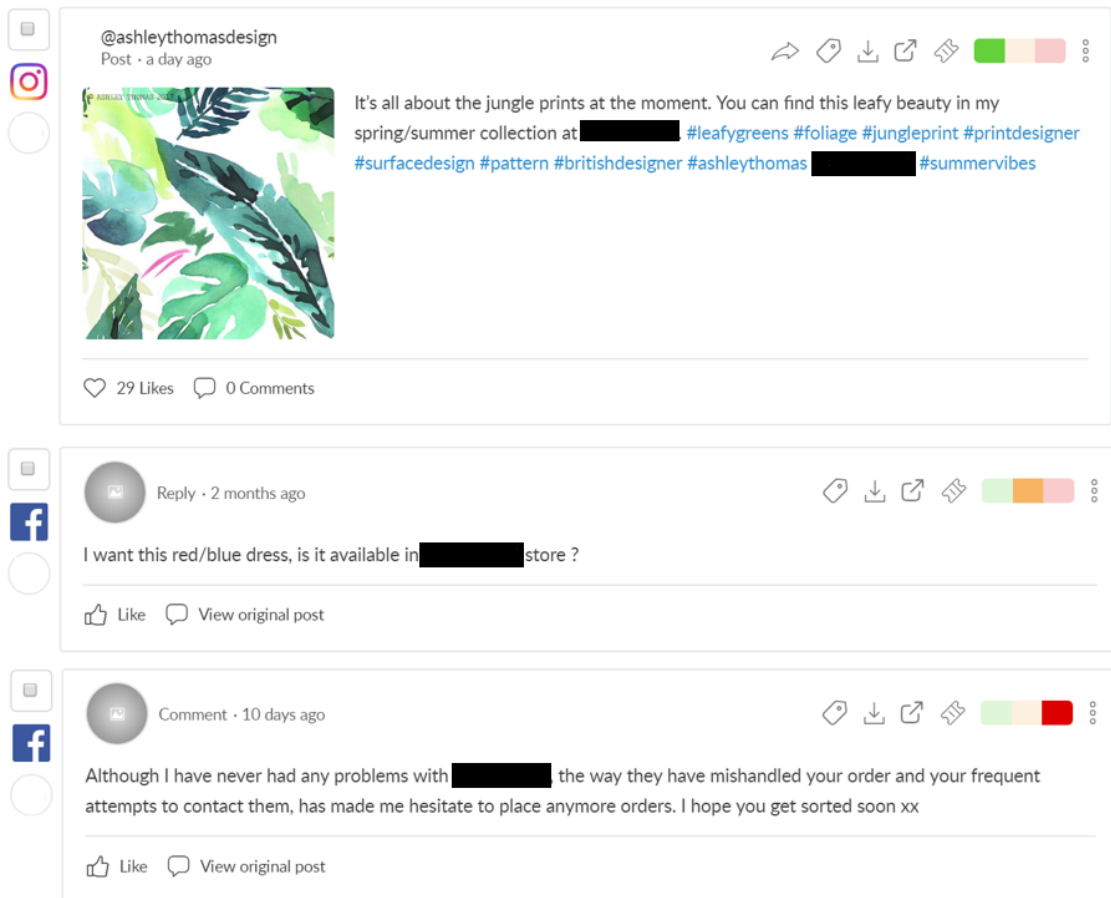


Figura D.8: Exemplos de comentários acerca do *Retalhista A*

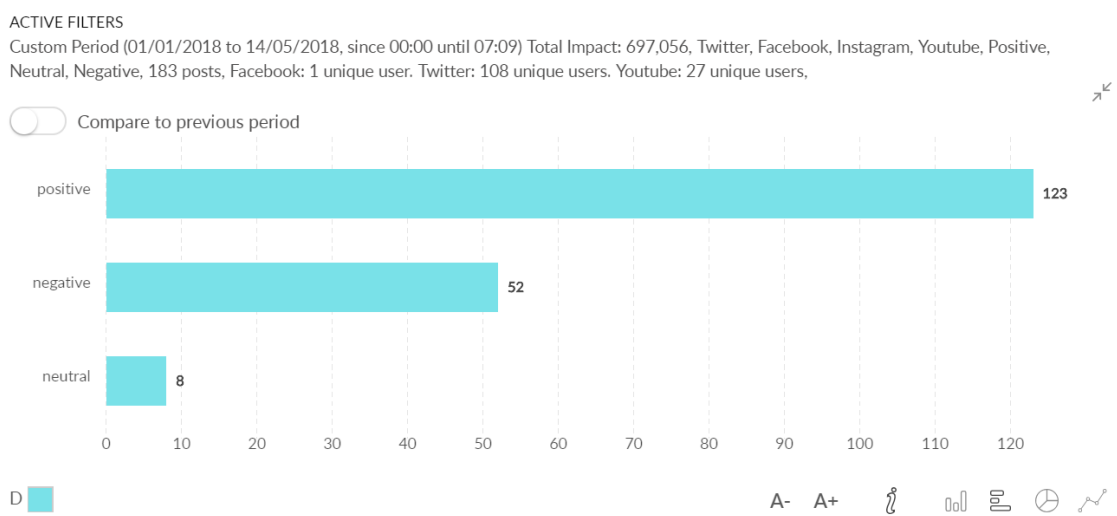


Figura D.9: Análise de sentimentos expressos pelos utilizadores acerca do *Concorrente 1*

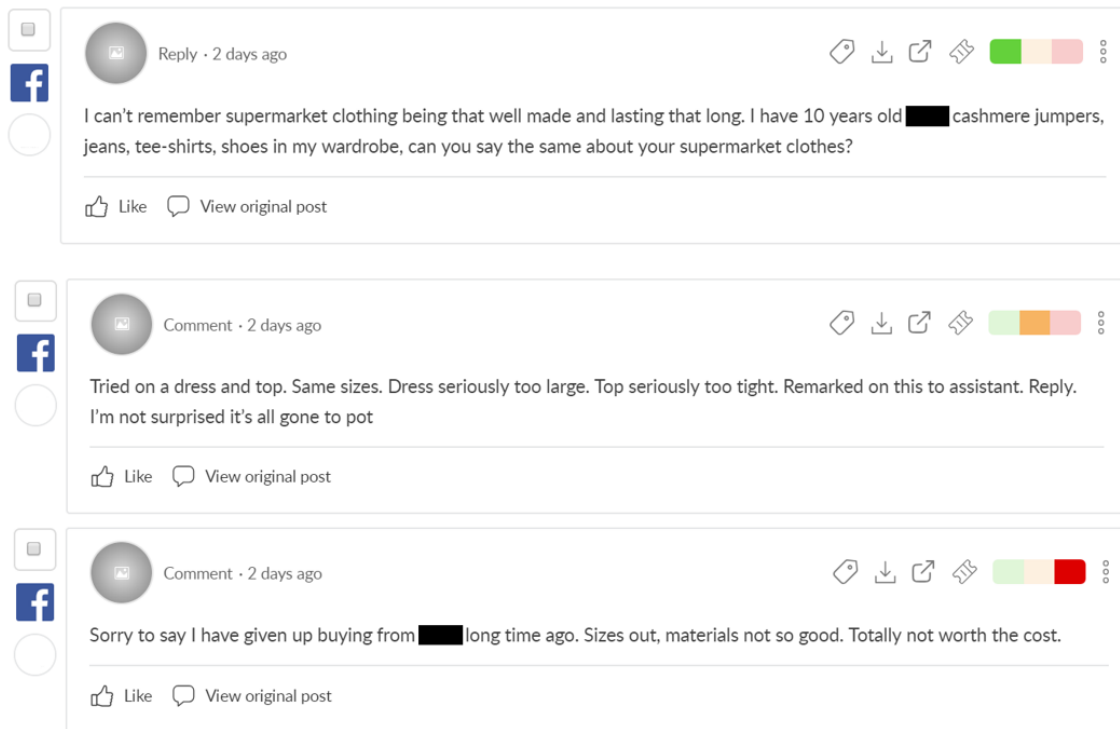


Figura D.10: Exemplos de comentários acerca do *Concorrente 1*

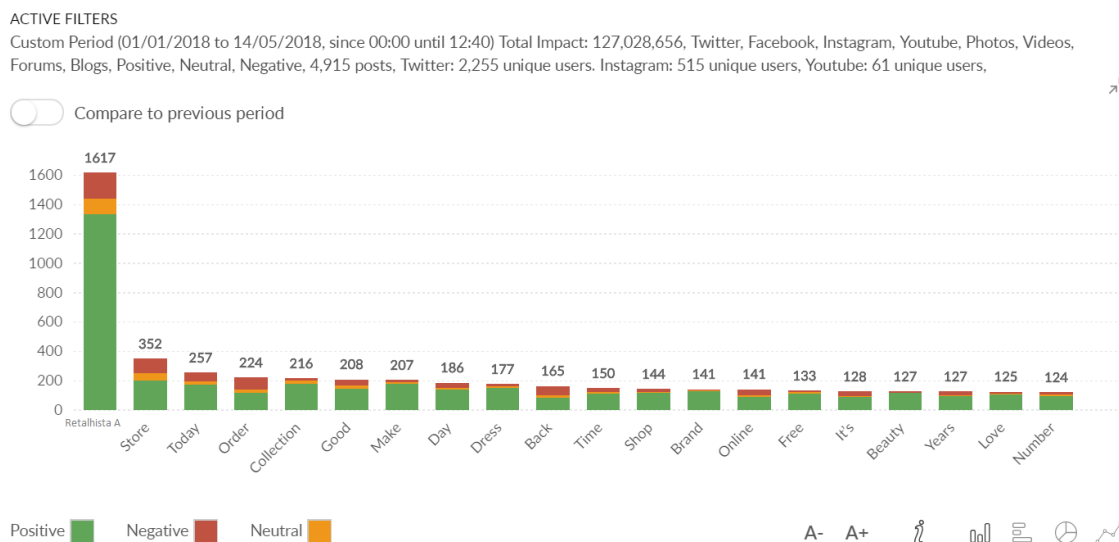


Figura D.11: Assuntos mais comentados acerca do *Retailista A*, desde 1 de janeiro de 2018 até 14 de maio de 2018

ACTIVE FILTERS

Custom Period (01/01/2018 to 21/05/2018, since 00:00 until 13:05) Total Impact: 28,056,410, Twitter, Facebook, Instagram, Youtube, Photos, Videos, Forums, Blogs, Positive, Neutral, Negative, 506 posts, Facebook: 1 unique user. Twitter: 233 unique users. Instagram: 37 unique users, Youtube: 8 unique users,

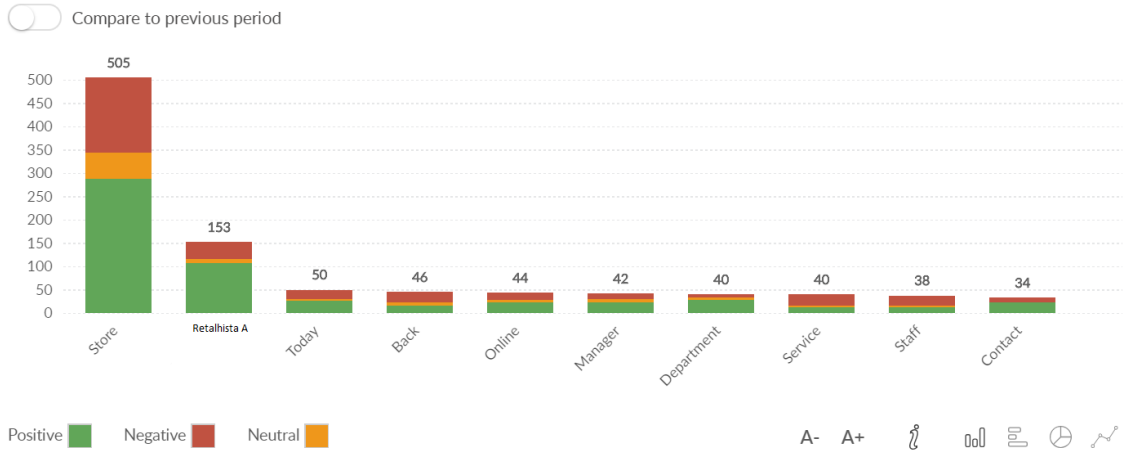


Figura D.12: Assuntos mais comentados acerca das lojas do *Retailista A*, desde 1 de janeiro de 2018 até 21 de maio de 2018

Brunswick Shopping @shopbrunswick  
Tweet · 3 days ago  
Craghoppers have landed in [redacted]. And there's amazing deals in store on both menswear and ladieswear. # [redacted] Scarborough #Craghoppers [https:// \[redacted\]](#)

erin @erintuffx\_  
Tweet · 11 days ago  
@[redacted] I bought it online, would I still take it into store? :)

Claire @willendintears  
Tweet · 2 days ago  
No wonder [redacted] is losing money. The toilets in the #Norwich shop are shut, no others in the store apparently, so I have walked out.

Figura D.13: Comentários exemplo acerca das lojas do *Retailista A*

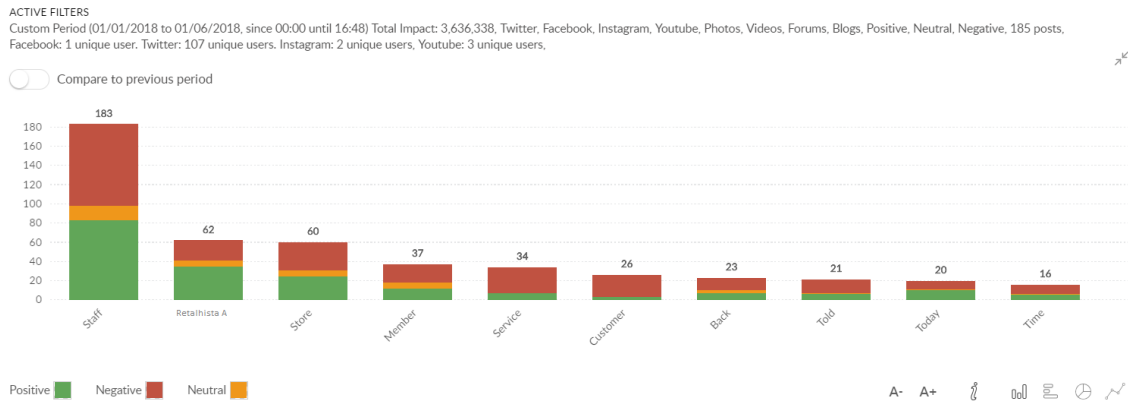


Figura D.14: Análise geral acerca dos comentários relativos ao *staff* das lojas do *Retailista A*

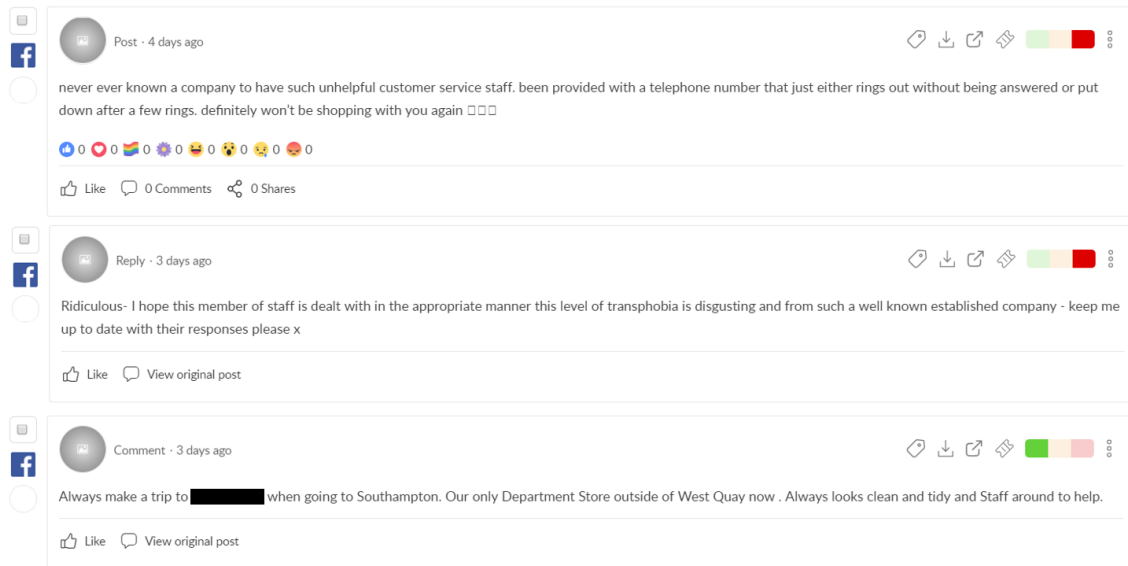


Figura D.15: Pequena amostra de comentários que é possível visualizar a partir da Figura D.14

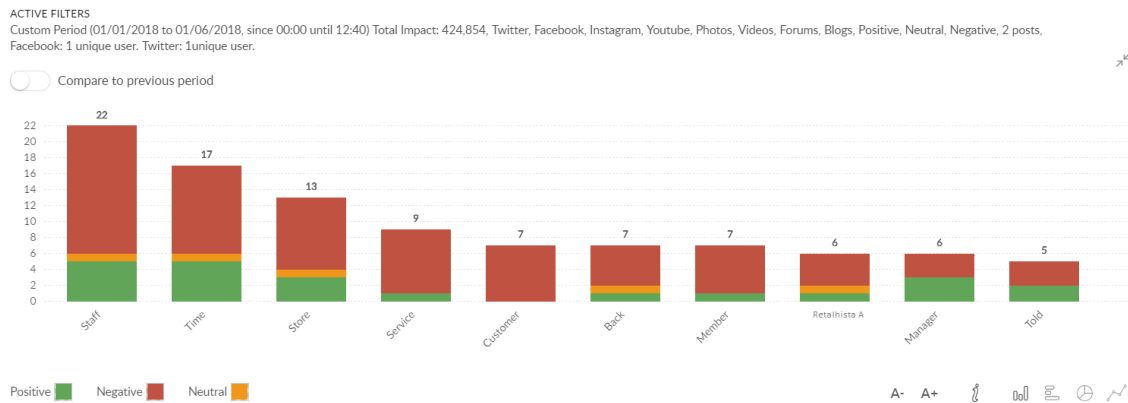


Figura D.16: Análise mais restrita acerca das críticas relativas ao *staff* das lojas do *Retailista A*



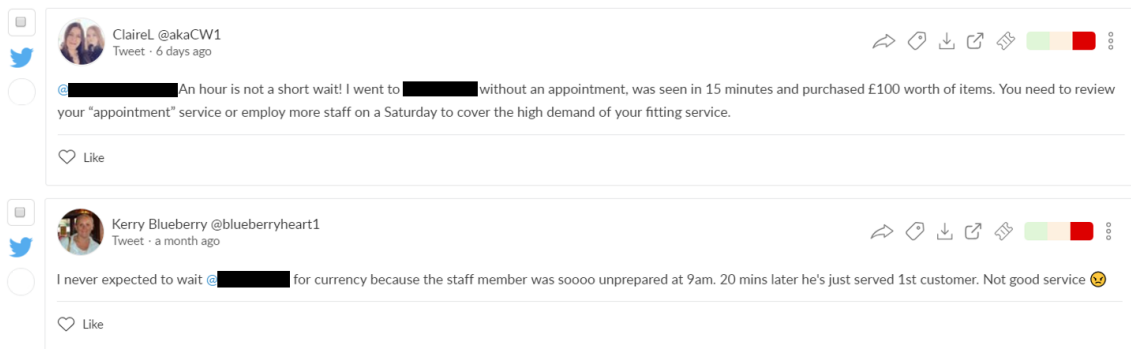


Figura D.17: Pequena amostra de comentários que é possível visualizar a partir da Figura D.16

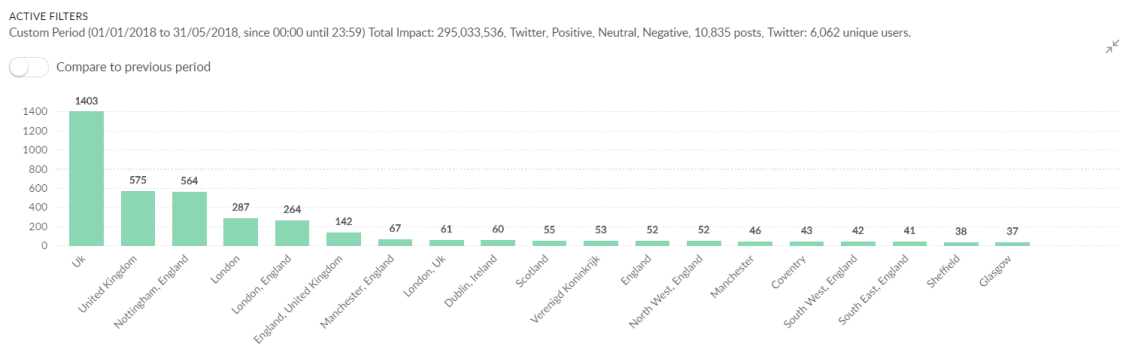


Figura D.18: Localização dos utilizadores que mais comentam sobre a marca do *Retalhista A*

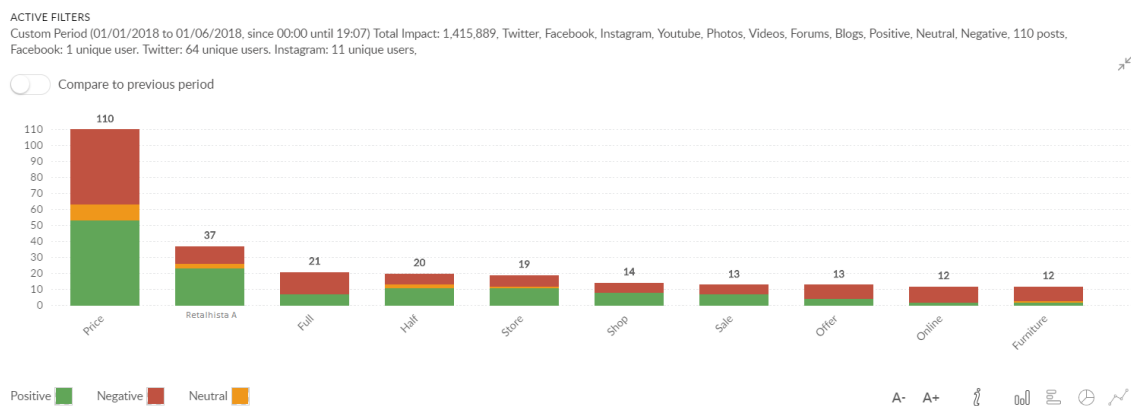


Figura D.19: Reclamações ou elogios relativamente aos preços da marca do *Retalhista A*

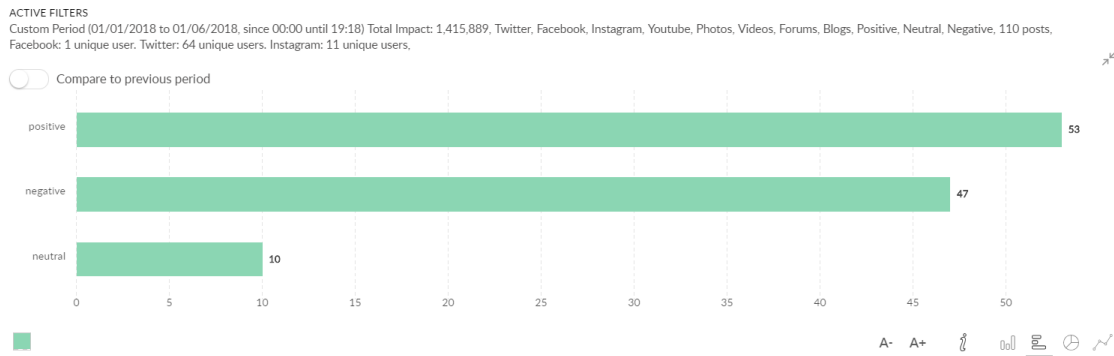


Figura D.20: Sentimentos expressos pelos utilizadores relativamente aos preços da marca do *Retalhista A*

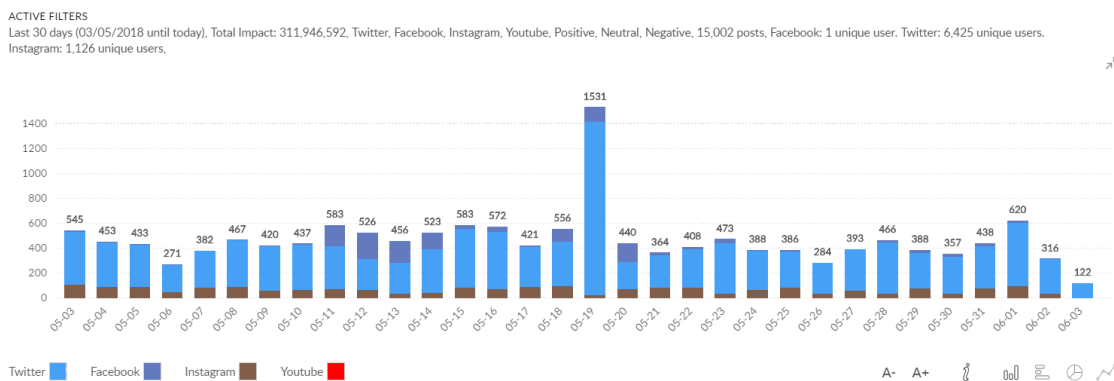


Figura D.21: Redes sociais com maior adesão dos utilizadores para o caso do *Retalhista A*

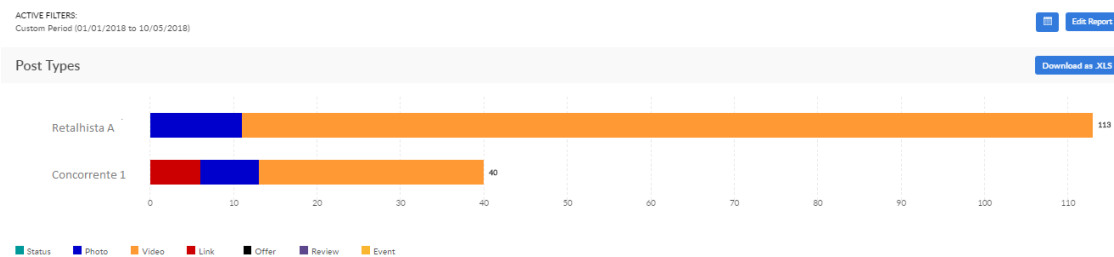


Figura D.22: Adesão por tipo de publicação

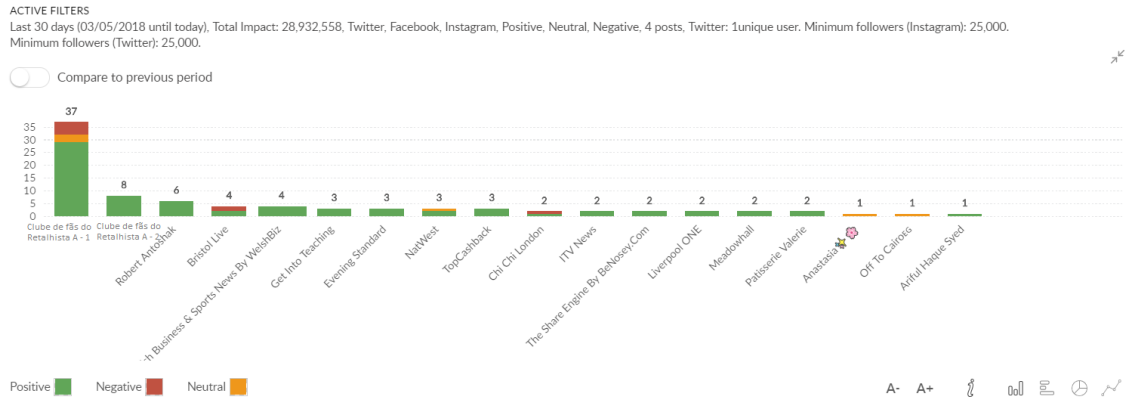


Figura D.23: Influenciadores que mais comentam acerca da marca *Retalhista A*, nas redes sociais

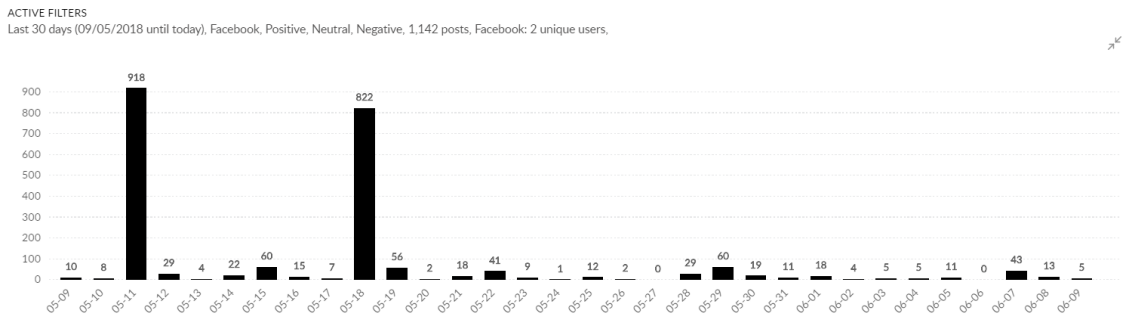


Figura D.24: Número de comentários, por dia, na página do *Facebook* do *Retalhista A*

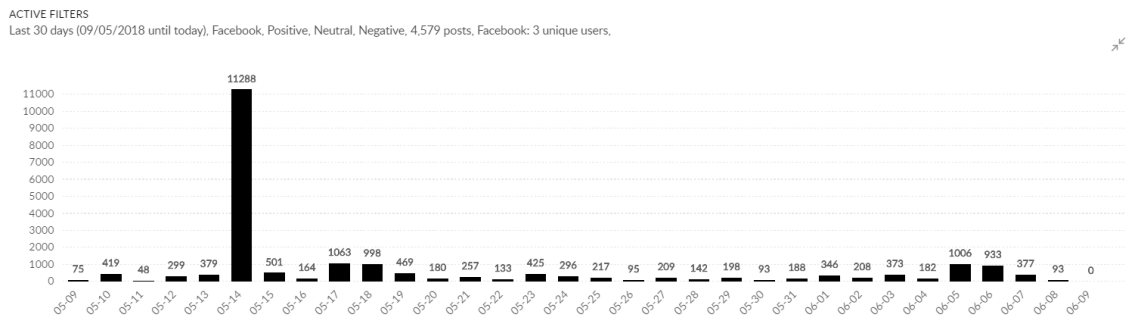


Figura D.25: Número de comentários, por dia, na página do *Facebook* do *Concorrente 1*

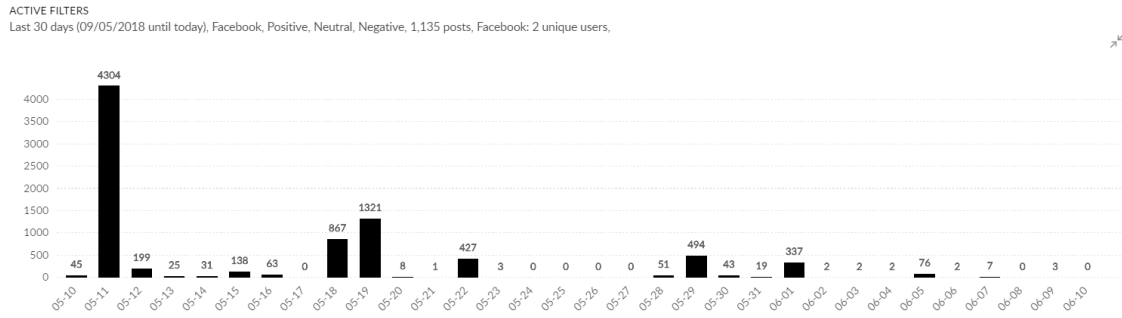


Figura D.26: Número de likes, por dia, na página do Facebook do Retalhista A

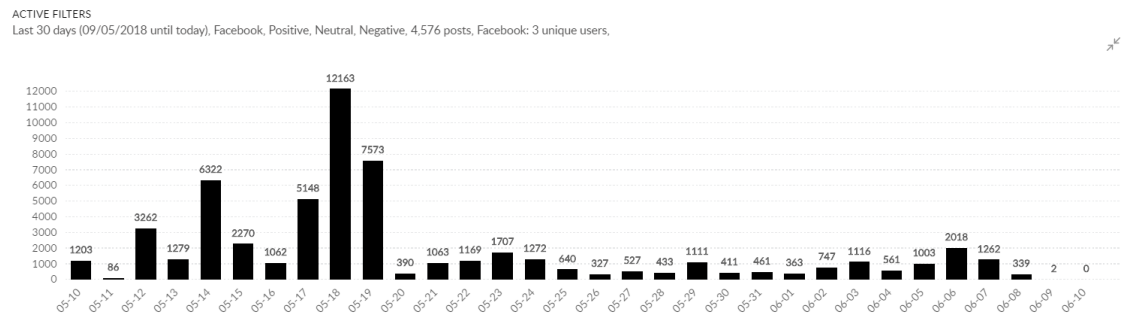


Figura D.27: Número de likes, por dia, na página do Facebook do Concorrente 1

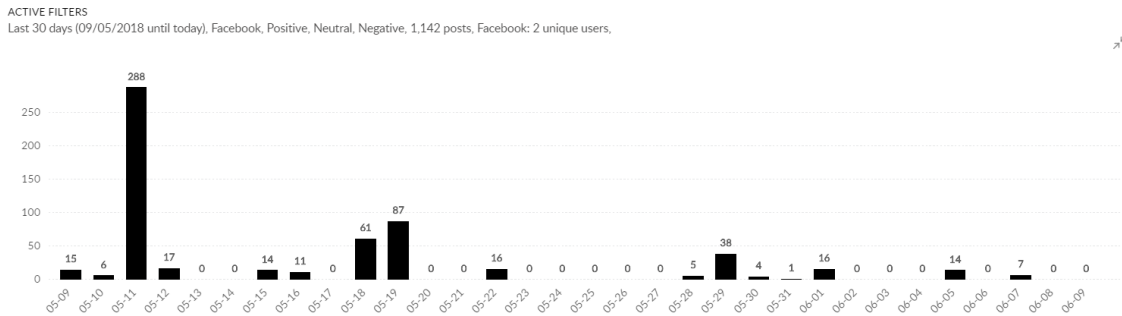


Figura D.28: Número de partilhas das publicações do Retalhista A, por dia

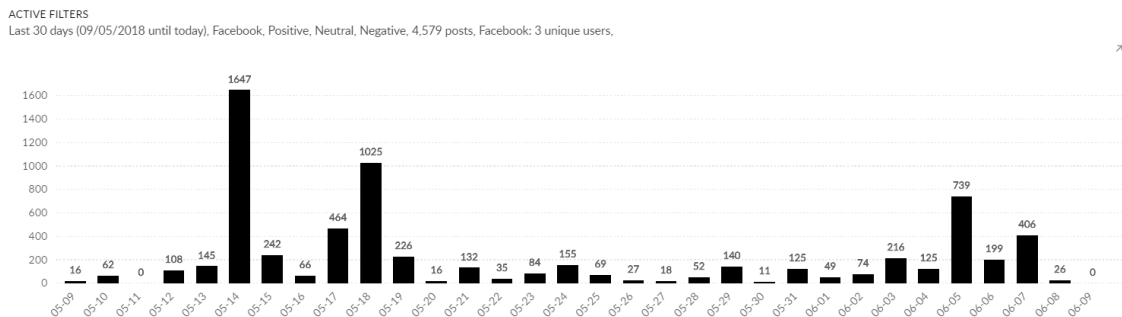


Figura D.29: Número de partilhas das publicações do Concorrente 1, por dia

# Referências

- [1] *2018 Retail Industry Trends & Predictions | Future of Retail | Vend*. Último acesso em 21 de março de 2018. URL: <https://www.vendhq.com/us/2018-retail-trends-predictions>.
- [2] José Soeiro. «3\_SWOT». Em: (2015).
- [3] Mr S M Vohra e Prof J B Teraiya. «A COMPARATIVE STUDY OF SENTIMENT ANALYSIS TECHNIQUES». Em: 12.02 (2013), pp. 975–6760. URL: <https://pdfs.semanticscholar.org/3f10/b006bab60c7f363bc03e72ad405d264b8d42.pdf>.
- [4] Ch Srinivasa Rao, G Satyanarayana Prasad e DrVedula Venkateswara Rao. «A Survey on Opinion Mining on Twitter Data: Tasks, Approaches, Applications and Challenges for Sentimental Analysis». Em: *IJCSN -International Journal of Computer Science and Network* 7.1 (2018), pp. 2277–5420. URL: <http://ijcsn.org/IJCSN-2018/7-1/A-Survey-on-Opinion-Mining-on-Twitter-Data-Tasks-Approaches-Applications-and-Challenges-for-Sentimental-Analysis.pdf>.
- [5] *Amazon Go. Abriu o primeiro supermercado onde não tem de passar pela caixa*. Último acesso em 12 de abril de 2018. URL: <https://www.dinheirovivo.pt/empresas/galeria/amazon-go-abriu-o-primeiro-supermercado-onde-nao-tem-de-passar-pela-caixa/#>.
- [6] Tiago C De França e Jonice Oliveira. «Análise de Sentimento de Tweets Relacionados aos Protestos que ocorreram no Brasil entre Junho e Agosto de 2013». Em: (2013). URL: <http://www.each.usp.br/digiampietri/BraSNAM/2014/p11.pdf>.
- [7] *Análise S.W.O.T.: O Guia para tomada de decisões*. Último acesso em 17 de maio de 2018. URL: <http://www.fcapjr.com.br/analiseswot/>.
- [8] *Blog | 12 métricas essenciais para gerenciar sua marca no Facebook*. Último acesso em 3 de junho de 2018. URL: <https://www.buzzmonitor.com.br/blog/12-metricas-essenciais-para-gerenciar-sua-marca-no-facebook>.
- [9] *Blog | Análise de Sentimento: como funciona?* Último acesso em 22 de maio de 2018. URL: <https://www.buzzmonitor.com.br/blog/analise-de-sentimento-como-funciona>.

- [10] *Comércio | Entrar, escolher e sair: Amazon abre loja sem empregados | PÚBLICO*. Último acesso em 12 de abril de 2018. URL: <https://www.publico.pt/2018/01/22/economia/noticia/entrar-escolher-e-sair-amazon-abre-loja-sem-empregados-1800235>.
- [11] koti. «DATA MINING TECHNIQUES: A SURVEY PAPER». Em: *IJRET: International Journal of Research in Engineering and Technology* eISSN pISSN (2013), pp. 2319–1163.
- [12] Hussain Ahmad Madni, Zahid Anwar e Munam Ali Shah. «Data mining techniques and applications — A decade review». Em: *2017 23rd International Conference on Automation and Computing (ICAC)*. 2017. ISBN: 978-0-7017-0260-1. DOI: 10.23919/ICAC.2017.8082090.
- [13] Chi-In Wong *et al.* «Design of a Crawler for Online Social Networks Analysis». Em: (2014). URL: <http://www.wseas.org/multimedia/journals/communications/2014/a165704-469.pdf>.
- [14] *DFS on Binary Tree Array*. Último acesso em 26 de abril de 2018. URL: <http://mishadoff.com/blog/dfs-on-binary-tree-array/>.
- [15] *How Google Search Works - Search Console Help*. Último acesso em 26 de abril de 2018. URL: <https://support.google.com/webmasters/answer/70897?hl=en>.
- [16] *How Search Works - YouTube*. Último acesso em 26 de abril de 2018. URL: <https://www.youtube.com/watch?v=BNHR6IQJGZs>.
- [17] *Inteligência Artificial e o futuro do trabalho – eg!creative – Medium*. Último acesso em 17 de abril de 2018. URL: <https://medium.com/eg-creative-blog/intelig%C3%A2ncia-artificial-e-o-futuro-do-trabalho-e811b6f45574>.
- [18] Jorge Caiado. «Métodos de Previsão em Gestão». Em: (2016). URL: [http://www.silabo.pt/Conteudos/8575\\_PDF.pdf](http://www.silabo.pt/Conteudos/8575_PDF.pdf).
- [19] Sunita Ghotiya, Juhi Mandal e Saravanakumar Kandasamy. «Migration from relational to NoSQL database». Em: (2017). DOI: 10.1088/1757-899X/263/4/042055. URL: <http://iopscience.iop.org/article/10.1088/1757-899X/263/4/042055/pdf>.
- [20] *O que é Data Mining: Conceitos e Técnicas sobre Data Mining*. Último acesso em 18 de junho de 2018. URL: <https://www.devmedia.com.br/conceitos-e-tecnicas-sobre-data-mining/19342>.
- [21] Nunes Dias *et al.* «Previsão de Demanda: Uma Análise quantitativa baseada em séries temporais de uma empresa fabricante de portas». Em: (2015). URL: [http://www.fecilcam.br/anais/ix\\_eepa/data/uploads/1-engenharia-de-operacao-e-processos-da-producao/1-01.pdf](http://www.fecilcam.br/anais/ix_eepa/data/uploads/1-engenharia-de-operacao-e-processos-da-producao/1-01.pdf).

- [22] S. Brusque e L.C. Zucatto. «Previsão de Vendas para Empresa Varejista de Confecções Adulto Feminino e Masculino». Em: *Revista de Administração e Negócios da Amazônia* (2015). ISSN: 21768366. DOI: 10.18361/2176-8366/rara.v7n2p88-111.
- [23] *Quase metade das empresas usa redes sociais para chegar aos clientes – Observador*. Último acesso em 21 de março de 2018. URL: <https://observador.pt/2017/11/21/quase-metade-das-empresas-usa-redes-sociais-para-chegar-aos-clientes/>.
- [24] *Redes sociais ganham 11 novos utilizadores por segundo - Renascença*. Último acesso em 21 de março de 2018. URL: <http://rr.sapo.pt/artigo/104619/redes-sociais-ganham-11-novos-utilizadores-por-segundo>.
- [25] Rahul Rajput e Arun Kumar Solanki. «Review of Sentimental Analysis Methods using Lexicon Based Approach». Em: *International Journal of Computer Science and Mobile Computing* 52.2 (2016), pp. 159–166. URL: <https://pdfs.semanticscholar.org/240c/586df4d4963a3e17f7fa35bbadf10276654a.pdf>.
- [26] Walaa Medhat, Ahmed Hassan e Hoda Korashy. «Sentiment analysis algorithms and applications: A survey». Em: (2014). DOI: 10.1016/j.asej.2014.04.011. URL: [http://kt.ijs.si/markodebeljak/Lectures/Seminar\\_MPS/2012\\_on/Seminars\\_2015\\_16/Simon%20Brmez/Bibliography/%5B5%5D%20Sentiment%20analysis%20algorithms%20and%20applications%20A%20survey.pdf](http://kt.ijs.si/markodebeljak/Lectures/Seminar_MPS/2012_on/Seminars_2015_16/Simon%20Brmez/Bibliography/%5B5%5D%20Sentiment%20analysis%20algorithms%20and%20applications%20A%20survey.pdf).
- [27] *Social Business Para Todos | A E.Life e o Buzzmonitor - YouTube*. Último acesso em 29 de maio de 2018. URL: <https://www.youtube.com/watch?v=kbdY7DsPnn8>.
- [28] Stefan Stieglitz *et al.* «Social media analytics – Challenges in topic discovery, data collection, and data preparation». Em: (2018). DOI: 10.1016/j.ijinfomgt.2017.12.002. URL: [https://ac.els-cdn.com/S0268401217308526/1-s2.0-S0268401217308526-main.pdf?\\_tid=f1c0d08c-3414-42cf-b471-fbc22587705e&acdnt=1524213001\\_9777814dc421937ff741ff7fd044aa0f](https://ac.els-cdn.com/S0268401217308526/1-s2.0-S0268401217308526-main.pdf?_tid=f1c0d08c-3414-42cf-b471-fbc22587705e&acdnt=1524213001_9777814dc421937ff741ff7fd044aa0f).
- [29] Wu He, Shenghua Zha e Ling Li. «Social media competitive analysis and text mining: A case study in the pizza industry». Em: *International Journal of Information Management* (2013). ISSN: 02684012. DOI: 10.1016/j.ijinfomgt.2013.01.001.
- [30] Touhid Bhuiyan, Yue Xu e Audun Jøsang. «State-of-the-Art Review on Opinion Mining from Online Customers' Feedback». Em: (2009), pp. 4–7. URL: <http://eprints.qut.edu.au/>.
- [31] Ramzan Talib *et al.* «Text Mining: Techniques, Applications and Issues». Em: *IJACSA International Journal of Advanced Computer Science and Applications* 7.11 (2016). URL: [https://thesai.org/Downloads/Volume7No11/Paper\\_53-Text\\_Mining\\_Techniques\\_Applications\\_and\\_Issues.pdf](https://thesai.org/Downloads/Volume7No11/Paper_53-Text_Mining_Techniques_Applications_and_Issues.pdf).

- [32] *Text mining vs data mining: discover the differences* -. Último acesso em 2 de abril de 2018. URL: <http://www.expertsystem.com/text-mining-vs-data-mining-differences/>.
- [33] Dipali V Talele MECSE e Ghriem D Chandrashekhar Badgujar. «The Art of Opinion Mining and Its Application Domains: - A Survey». Em: (2012), pp. 975–8887. URL: <https://pdfs.semanticscholar.org/47cf/93f7044b26458987c682e44dd413e4fbdd78.pdf>.
- [34] *The future of retail: How to make your bricks click* | McKinsey & Company. Último acesso em 21 de março de 2018. URL: <https://www.mckinsey.com/business-functions/marketing-and-sales/our-insights/the-future-of-retail-how-to-make-your-bricks-click>.
- [35] Roberto Braga, Fernandes Filho e Porto Alegre. «UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL ESCOLA DE ENGENHARIA PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO INTEGRAÇÃO DE MODELOS DE PREVISÃO DE DEMANDA QUALITATIVOS E QUANTITATIVOS E COMPARAÇÃO COM SEUS DESEMPENHOS INDIVIDUAIS». Em: (2015). URL: <https://www.lume.ufrgs.br/bitstream/handle/10183/117896/000968719.pdf?sequence=1>.
- [36] *Web Crawlers — Everything You Need to Know – Cabot Technology Solution – Medium*. Último acesso em 26 de abril de 2018. URL: [https://medium.com/@cabot\\_solutions/web-crawlers-everything-you-need-to-know-6dce26ee8ad8](https://medium.com/@cabot_solutions/web-crawlers-everything-you-need-to-know-6dce26ee8ad8).
- [37] *What is AI (Artificial Intelligence)? - Definition from WhatIs.com*. Último acesso em 22 de março de 2018. URL: <http://searchcio.techtarget.com/definition/AI>.
- [38] *What is Machine Learning? - YouTube*. Último acesso em 8 de março de 2018. URL: <https://www.youtube.com/watch?v=HcqpanDadyQ>.
- [39] *What is opinion mining - Next level in sentiment analytics* | BrandsEye. Último acesso em 2 de abril de 2018. URL: <https://www.brandseye.com/news/what-is-opinion-mining-next-level-sentiment-analytics/>.
- [40] *Wipro Named as a 2018 World's Most Ethical Company by the Ethisphere Institute for the 7th Successive Year - Wipro*. Último acesso em 18 de março de 2018. URL: <https://www.wipro.com/newsroom/press-releases/2018/wipro-named-as-a-2018-worlds-most-ethical-company-by-the-ethisphere-institute-for-the-7th-successive-year/>.