

Aplicações da *Sentiment Analysis* na Gestão de Empresas

Ray Generoso de Almeida Bragança

Dissertação para obtenção do Grau de Mestre em
Gestão
(2.º ciclo de estudos)

Orientador: Prof. Doutor Francisco José Peixeiro Antunes

dezembro de 2022

Declaração de Integridade

Eu, Ray Generoso de Almeida Bragança, que abaixo assino, estudante com o número de inscrição M9677 de Mestrado em Gestão da Faculdade de Ciências Sociais e Humanas, declaro ter desenvolvido o presente trabalho e elaborado o presente texto em total consonância com o **Código de Integridades da Universidade da Beira Interior**.

Mais concretamente afirmo não ter incorrido em qualquer das variedades de Fraude Académica, e que aqui declaro conhecer, que em particular atendi à exigida referenciação de frases, extratos, imagens e outras formas de trabalho intelectual, e assumindo assim na íntegra as responsabilidades da autoria.

Universidade da Beira Interior, Covilhã 27 /12/2022

Ray G. de A. Bragança

Agradecimentos

Antes de mais, gostava de agradecer a Deus-Pai Todo-Poderoso pelo dom da vida, pela maturidade, por me ter guiado ao longo desta longa e difícil caminhada e por me ter concedido forças para superar este desafio que culminou com a elaboração deste trabalho dissertativo.

Aos meus pais, Cristóvão Bragança e Justina de Almeida Bragança, agradeço pela disciplina, ensinamentos, educação e valores transmitidos que me tornaram no homem que sou hoje. Em especial, agradeço à minha mãe pelas palavras de conforto que transmitiam sempre segurança e confiança que tudo ficaria bem. Aos meus queridos irmãos, Danila Patrícia e Juvão Tolstoy, agradeço pelo incentivo, pela motivação, pelo apoio e pela amizade que não tem preço.

À minha companheira, agradeço pelo amor, amizade, suporte, palavras de encorajamento, momentos de descontração, e sobretudo, por fazer sentir a sua presença mesmo quando estivéssemos distantes geograficamente.

Ao meu orientador, Professor Doutor Francisco Antunes, agradeço muito pela orientação, paciência, compreensão, simpatia, acessibilidade, total disponibilidade, profissionalismo e confiança depositada.

A todos os amigos e restantes familiares que direta ou indiretamente contribuíram para a materialização deste trabalho, expresso o meu mais profundo e sincero agradecimento.

Muito obrigado a todos!

Resumo

O crescimento do digital vai-se evidenciando cada vez mais como sendo um processo irreversível e como tal, os gestores têm de ter capacidade de adaptação e saber identificar oportunidades para tirarem melhor proveito destes recursos tecnológicos. Neste sentido, a *sentiment analysis* é uma técnica analítica que permite analisar e classificar, através de corpos de textos, a polaridade (positiva, neutra ou negativa) dum determinado assunto.

À medida que aumenta o volume de informação textual disponível online, esta técnica tem um grande potencial de aplicação. No entanto, na literatura não se encontram estudos anteriores para determinar quais são as funções de negócios reais que usam a análise de sentimentos (SA), nem em que medida/propósitos as fontes de dados são usadas para tal empreendimento e, portanto, é necessário conhecer as formas de uso da SA na gestão de negócios, para maximizar o seu uso potencial e os benefícios que vêm com ele.

Este trabalho tem por objetivo explorar e identificar as temáticas em que são aplicadas a *sentiment analysis* na gestão de empresas, bem como identificar as principais fontes de dados para fazer este tipo de análise. Estas temáticas foram enquadradas nas cinco funções básicas da gestão: administrativa, financeira, de *marketing*, de produção e de recursos humanos. Para alcançar isto, o estudo foi conduzido através de revisão sistemática da literatura, conjuntamente com uma análise bibliométrica e a elaboração de uma proposta de taxonomia. Para a realização do estudo, foram extraídos 1.151 artigos da *Web of Science*, provenientes de periódicos e conferências.

Os resultados sugerem que a SA é maioritariamente utilizada nas funções de *marketing* e na financeira, embora também se verifiquem aplicações nas funções administrativa, produção e recursos humanos, mas de forma residual. Concluiu-se ainda que existem 4 tipos de fontes de informação: documentação interna, documentação financeira, redes sociais/publica e académica.

Palavras-chave

Análise; Sentimentos; Gestão; Polaridade; Bibliometria; Taxonomia

Abstract

It is increasingly evident that the growth of digital technology is becoming an irreversible process, implying that managers must have the ability to adapt and to identify opportunities to make the most of these technological resources. Therefore, sentiment analysis is an analytical technique that allows the analysis and classification of the polarity (positive, neutral, or negative) of a given subject embedded within a text.

As the volume of textual information available online increases, this technique has great potential for application. However, no previous studies were found in the literature to determine which actual business functions are using sentiment analysis (SA), nor to what extent or purpose distinct data sources are being used for such an endeavor. Thus, it is necessary to know how SA is being used in business management, to maximize its potential use and the benefits that come along with it.

This paper aims to explore and identify the themes for which sentiment analysis is applied in business management, as well as to identify the main data sources for doing this type of analysis. These themes were framed within the five basic business functions: general management, finance, marketing, production and human resources. To achieve the proposed goals, this study underwent a systematic literature review, used a bibliometric analysis, and developed a taxonomy proposal. To conduct the study, 1.151 articles, whether from journals or conferences, were extracted from Web of Science.

The results suggest that the SA is mostly used in marketing and finance, there are also applications regarding the other functions, though in a residual way. It was also concluded that there are 4 types of information sources: academic, internal documentation, financial documentation, and social networks/public information.

Keywords

Sentiment;Analysis;Business;Management;Polarity;Bibliometry;Taxonomy

Índice

Agradecimentos.....	v
Resumo.....	vii
<i>Abstract</i>	ix
Índice.....	xi
Lista de Figuras.....	xiii
Lista de Tabelas.....	xv
Lista de Acrónimos.....	xvii
Capítulo 1.....	1
Introdução.....	1
Capítulo 2.....	3
Funções Empresariais.....	3
2.1. Função Administrativa.....	3
2.2. Função Financeira.....	3
2.3. Função de <i>Marketing</i>	4
2.4. Função de Produção.....	5
2.5. Função de Recursos Humanos.....	5
Capítulo 3.....	7
<i>Sentiment Analysis</i>	7
3.1. Níveis de análise.....	7
3.2. Etapas do processo.....	8
3.3. Classificação de sentimentos.....	10
3.3.1 <i>Machine Learning</i>	11
3.3.1.1. Aprendizagem supervisionada.....	11
3.3.1.2. Aprendizagem não supervisionada.....	12
3.3.2 Método baseado no Léxico.....	12
3.3.2.1. Abordagem de dicionário.....	13
3.3.2.2. Abordagem de <i>corpus</i>	13
Capítulo 4.....	15
Recolha, apresentação e análise dos dados.....	15
4.1. Análise Bibliométrica.....	15
4.1.1 Principais <i>Clusters</i>	17
4.1.2 Principais Autores.....	19
4.1.3 Principais Publicações e Conferências.....	20
4.2. Temáticas por função.....	21
4.2.1. Função Administrativa.....	24

4.2.2. Função Financeira	25
4.2.3. Função de <i>Marketing</i>	25
4.2.4. Função de Produção	26
4.2.5. Função de Recursos Humanos	27
4.3. Proposta de Taxonomia.....	27
4.4. Origem dos dados por função empresarial	31
4.4.1 Função Administrativa.....	32
4.4.2 Função Financeira	33
4.4.3 Função de <i>Marketing</i>	34
4.4.4 Função de Produção	35
4.4.5 Função de Recursos Humanos	35
Capítulo 5	37
Conclusões, limitações e trabalhos futuros	37
Referências Bibliográficas	41

Lista de Figuras

Figura 1. Métodos e técnicas de SA.....	10
Figura 2. Esquema de pesquisa.....	16
Figura 3. Rede de palavras com coocorrência	17
Figura 4. Número de publicações por anos	22
Figura 5. Temáticas e número de artigos associados (cada artigo pode ter mais que uma)	23
Figura 6. Distribuição de temáticas de aplicação da SA na função administrativa.....	24
Figura 7. Distribuição de temáticas de aplicação da SA na função financeira.....	25
Figura 8. Distribuição de temáticas de aplicação da SA na função de <i>marketing</i>	26
Figura 9. Distribuição de temáticas de aplicação da SA na função de produção.....	26
Figura 10. Distribuição de temáticas de aplicação da SA na função de recursos humanos	27

Lista de Tabelas

Tabela 1. Clusters correspondentes às funções de gestão	18
Tabela 2. Os 20 autores com mais artigos publicados sobre a SA na gestão	20
Tabela 3. Top 20 de periódicos com publicações sobre a SA na gestão	20
Tabela 4. As 20 principais conferências com publicações sobre a SA na gestão	21
Tabela 5. Associação dos artigos incluídos por função empresarial	22
Tabela 6. Proposta de taxonomia.....	28
Tabela 7. Repartição dos grupos de fontes de informação pelas várias funções empresariais. ...	32
Tabela 8. Origem de dados da função administrativa.....	33
Tabela 9. Origem de dados da função financeira	33
Tabela 10. Origem de dados da função de <i>marketing</i>	34
Tabela 11. Origem de dados da função de produção.....	35
Tabela 12. Origem de dados da função de recursos humanos.....	35

Lista de Acrónimos

API	<i>Application Programming Interface</i>
CART	<i>Classification and Regression Trees</i>
eWOM	<i>Electronic Word-Of-Mouth</i>
ML	<i>Machine Learning</i>
OIT	Organização Internacional do Trabalho
POS	<i>Part-of-speech</i>
RF	<i>Random Forest</i>
SA	<i>Sentiment Analysis</i>
SVM	<i>Support Vector Machines</i>
WOM	<i>Word-of-Mouth</i>
WOS	<i>Web of Science</i>

Capítulo 1

Introdução

Viver num mundo em que o consumo tecnológico é cada vez maior, certamente implica que mudanças e ajustes têm e terão de ser feitos para que se consiga acompanhar o ritmo desta nova realidade. Portanto, é de extrema importância que as empresas ou organizações, através dos gestores, se encaixem rapidamente para que possam explorar novas possibilidades de aplicação destas tecnologias com o objetivo de identificarem novas oportunidades de negócio e delinearem novas estratégias para se destacarem (Hess *et al.*, 2016).

Vivemos numa era onde grande parte da interação com os clientes e a divulgação de conteúdos são feitas utilizando plataformas digitais, podendo a *sentiment analysis* (SA) ter um papel essencial na captação, extração e interpretação da informação (por exemplo as opiniões de clientes/colaboradores), na medição de aceitação de um dado produto/serviço, nas previsões de dados financeiros, entre outros, segundo Biswas *et al.* (2020) e Wankhade, Rao & Kulkarni (2022), constituindo um campo de investigação relativamente novo, mas que tem observado um crescimento nos últimos anos (Mantyla, Graziotin & Kuutila, 2018; Silva *et al.*, 2018).

De forma resumida e objetiva, a SA determina se o excerto dum texto expressa uma opinião positiva, negativa ou neutra sobre um determinado assunto (Micu *et al.*, 2017). Deste modo, toda a informação textual disponibilizada *online* é passível de ser analisada utilizando a SA. Esta técnica analítica possui um grande potencial de aplicação e a sua utilização é gradualmente mais visível (Benamara, Taboada & Mathieu, 2017). Em vista disso, torna-se imperativo saber de que formas é que se pode aproveitar a SA na gestão de empresas para maximizar o potencial de utilização e os benefícios que vêm acompanhados.

Desta feita, havendo interesse em saber de que formas é que esta técnica pode ser utilizada pelos gestores, definem-se como questões de pesquisa:

1. *De que modo é que se tem estado a aplicar a SA no estudo da gestão das empresas?*
2. *Quais são as funções empresariais que mais aplicam a SA?*
3. *É possível construir uma taxonomia da utilização da SA dentro das funções empresariais?*
4. *Quais são as fontes de informação mais utilizadas pelos estudos para a obtenção de informação para a SA, dentro de cada função empresarial?*

Para atingir os objetivos, usou-se uma metodologia qualitativa com uma abordagem de análise documental, através de uma ampla revisão sistemática da literatura de artigos retirados da *Web of Science* (WOS). Após a definição dos parâmetros de pesquisa, foram filtrados os resultados por áreas de interesse e logo de seguida avançou-se para a seleção manual dos artigos. Para dar resposta às questões de pesquisa, conciliou-se ainda a seleção manual dos artigos com uma análise bibliométrica realizada com o software VOSviewer, como aprofundado no capítulo 4.

O segundo capítulo apresenta o trabalho de gestão estruturado em cinco funções principais (administrativa, financeira, de *marketing*, de produção e de recursos humanos) e potencialidades de utilização da *sentiment analysis* em cada uma.

No terceiro capítulo é feita a apresentação da SA e uma breve contextualização histórica. Logo de seguida, são explicados os níveis de classificação em que a SA pode ser implementada. Nos pontos subsequentes são discutidos o funcionamento da SA, o seu fluxo de trabalho, as principais abordagens e técnicas e os algoritmos mais utilizados.

O quarto capítulo apresenta todos os passos que foram dados para que fossem obtidas as respostas às questões de pesquisa. É explicado o processo de recolha, seleção e exclusão dos artigos. No mesmo capítulo, são descritas as definições feitas no *software* utilizado para realizar a análise bibliométrica, cuja discussão e interpretação também tomam lugar no mesmo capítulo. Contém, ainda, a fundamentação teórica para a construção da taxonomia proposta, que é complementada com uma análise das temáticas identificadas. Por fim, integra também uma análise das principais fontes de informação utilizadas, nos artigos selecionados, para extrair os dados para aplicação da SA.

Por último, o quinto capítulo integra as principais conclusões e as considerações finais do trabalho desenvolvido, bem como as limitações do mesmo.

Refere-se, adicionalmente, que os temas apresentados nesta dissertação estão incluídos num artigo que foi aceite para apresentação na *42nd Eurasian Business and Economics Society (EBES) Conference*, a decorrer em Lisboa de 12 a 14 de janeiro de 2023, intitulado “*A systematic review and taxonomy proposal of sentiment analysis use within business functions*”, cuja versão alargada será submetida a 16 de março de 2023, à *Eurasian Business Review*.

Capítulo 2

Funções Empresariais

A gestão de empresas é uma ciência com muita versatilidade e com muita abrangência. É facilmente conjugada e adaptada às outras ciências. Neste sentido, dentro dela, existem funções em que se subdivide. De forma geral, parece haver um consenso entre alguns autores (Drucker, 1986; Chiavenato, 2004; Lisboa et al., 2011; Almeida, 2016) no que tange à divisão em cinco funções elementares: administrativa, financeira, de *marketing*, de produção e de recursos humanos. Em função disto, os artigos analisados neste trabalho serão enquadrados nestas funções empresariais.

2.1. Função Administrativa

Esta função é responsável pela articulação estrutural e organização funcional da empresa. Esta função está relacionada com aspetos que afetam a empresa de forma geral (Barreto, 2017) como o planeamento estratégico; a tomada de decisão; a gestão do conhecimento e a gestão da informação. Engloba ainda outros temas como a aprendizagem organizacional, que é a forma como os membros de uma organização aprendem (Armstrong & Taylor, 2014), o desempenho social corporativo e a responsabilidade social corporativa, que tratam dos processos de resposta social, políticas, programas e resultados observáveis da empresa (Wartick & Cochran, 1985; Wood, 1991) e a imagem corporativa. Através da imagem corporativa, a empresa pode apresentar ao público todas as dimensões do seu negócio e atividades, como por exemplo o que é, como faz, o que representa no negócio, quais são os seus produtos, quem são os seus consumidores/clientes, quais são os seus indicadores económicos e qual é a sua reputação (Zerfass & Viertmann, 2017). O escrutínio destas dimensões da função administrativa, seja através das redes sociais, seja através de comentários sobre as comunicações ou documentação exarada para o público é passível de ser sujeita à *sentiment analysis*.

2.2. Função Financeira

De acordo com Duarte & Barbosa (2019), esta função está relacionada com o planeamento, a captação, o levantamento orçamental e a gestão dos recursos financeiros, envolvendo também os registos contabilísticos das operações realizadas pela empresa (Almeida, 2016).

É responsável por gerir o dinheiro da empresa, protegê-lo e promover uma utilização eficaz do mesmo (Trigueiro & Marques, 2014). Isto inclui a maximização do retorno dos investimentos e a manutenção de um certo grau de liquidez para o cumprimento das obrigações. Deste modo, a função financeira abrange tanto a área de contabilidade como a área de mercados de capital.

Na parte da contabilidade, são fornecidas, através das demonstrações financeiras, informações que refletem a posição financeira da empresa (Lisboa et al., 2011). A contabilidade financeira

produz informação uniformizada e da divulgação externa que transmite uma imagem da realidade económica, financeira e patrimonial da empresa (Almeida, 2016).

Nos dias de hoje, além de páginas de notícias, as opiniões e sentimentos de consumidores expressos em comentários ou avaliações *online* fornecem muita informação. Portanto, a informação partilhada *online* não só é utilizada para analisar empresas e fazer previsões de mercado (Anuratha et al., 2019; Tanulia & Girsang, 2019), como também é utilizada para ajudar a fixar preços de produtos (Archak, Ghose & Ipeiritis, 2011; Almeida, 2016). Da mesma forma, os investidores também levam em consideração a informação da internet e o *feedback online* do consumidor antes de tomarem uma decisão de investimento (Sun et al., 2020; Chung, Shin & Park, 2022). As diferentes opiniões e notícias *online* contribuem significativamente na tomada de decisão (Anuratha et al., 2019).

Ao contrário do que acontecia antigamente, os investidores atuais não dependem somente da lógica para tomarem decisões, agora também o fazem considerando o seu estado emocional (Yadav et al., 2020), isto é, o chamado *investor sentiment* ou sentimento do investidor, que pode afetar o comportamento de diversos tipos de ativos financeiros. Por exemplo Luo (2020) apresenta a previsão de preços e taxa de retorno das *bitcoins* através dos dados e sentimentos recolhidos a partir do *Twitter* ou a análise de riscos financeiros associados a qualquer forma de financiamento (Peng et al., 2011). A análise das experiências pessoais e/ou crenças influenciam no sentimento e na forma de pensar do investidor, podem assim influenciar a sua decisão final (Pompian, 2017), sendo passíveis de ser analisadas com *sentiment analysis* como fonte de informação para a tomada de decisão.

2.3. Função de Marketing

Para Almeida (2016, p. 286), o “*marketing* é a área funcional da gestão que se ocupa da relação da organização com o mercado, baseado no princípio de que a criação de valor (satisfação) para o cliente é a melhor forma de atingir os objetivos organizacionais”. Segundo a *American Marketing Association* (AMA) tal criação de valor “envolve processos para criar, comunicar, entregar e trocar ofertas (serviços/produtos) que tenham valor para os clientes, parceiros, e sociedade em geral”¹.

Com o comércio *online* em constante crescimento, tornou-se imperativo recolher, analisar e processar todos os dados de clientes que as empresas possam. Tal análise, tem como finalidade identificar o público-alvo, perceber quais as suas necessidades e tentar ajustar o produto para que satisfaça estas necessidades (Liu, Jiang & Zhao, 2019). Neste processo a *sentiment analysis* representa uma ferramenta de análise que pode permitir o escrutínio dos sentimentos dos clientes em relação a produtos/serviços existentes/disponibilizados, bem como servir de base ao desenvolvimento de novos produtos/serviços, bem como à fixação dos seus preços.

¹ <https://www.ama.org/the-definition-of-marketing-what-is-marketing/>

2.4. Função de Produção

Relativamente à função de produção, há quem opte por chamá-la de “operações” (Chiavenato, 2004), representando a reunião de recursos e processos destinados à produção de bens e serviços para serem colocados no mercado (Trigueiro & Marques, 2014; Duarte & Barbosa, 2019). De acordo com Smith (1992), o planeamento e controlo do processo de fabrico nas empresas era feito de forma que a produção fosse célere e rentável, ou seja, primava pela eficiência. Porém, com o aumento da concorrência resultante da globalização dos mercados, as empresas foram obrigadas a incorporar o conceito de eficácia na medição da sua performance produtiva (Almeida, 2016), porque por exemplo, atualmente não basta só fabricar grandes quantidades de produtos por unidade de tempo, é também essencial que estes produtos cheguem ao cliente no momento que a empresa se comprometeu a entregá-los quando recebeu o pedido de encomenda. Ainda dentro deste tema, é importante ressaltar um ponto muito importante: quando se fala em produto, não se fala necessariamente em algo material ou físico. Atualmente, os produtos podem assumir várias formas e formatos. Muitas são as empresas cujo negócio é prestar serviços.

Dentro da função produção, a logística assume as tarefas de planejar, implementar e controlar de forma eficiente e eficaz os transportes de receção e expedição de materiais, bem como as operações de armazenamento de bens, serviços e informação desde ponto de origem até ao local de consumo². Também aqui a recolha e análise através da *sentiment analysis* das opiniões dos intervenientes nos vários processos abrangidos pela gestão da cadeia de abastecimentos (por exemplo de trabalhadores e operadores logísticos) pode ser relevante na deteção de constrangimentos e implementação de melhorias de processos.

2.5. Função de Recursos Humanos

A noção de recursos humanos remete para os conhecimentos, capacidades, competências, talentos e aptidões, implícitos ou explícitos, de cada pessoa, que permitem atingir metas individuais e coletivas em contexto organizacional (Almeida, 2016). Trigueiro & Marques (2014), continuam por dizer que os gestores de recursos humanos têm a responsabilidade de recrutar pessoas para integrarem o quadro de trabalhadores e que esta responsabilidade varia desde o recrutamento até à rescisão de contrato.

Para que haja uma melhor dinâmica de equipa e de trabalho, os gestores de recursos humanos têm a responsabilidade de criar estruturas e ambientes em que os funcionários se sintam bem. Para lograr-se isto, é importante que a satisfação do funcionário e as condições de trabalho estejam alinhadas.

Alguns fatores que podem influenciar na satisfação do trabalhador são a remuneração, a (des)promoção de cargos, a supervisão, os benefícios, as recompensas, as condições de trabalho, os colegas de trabalho e a natureza do trabalho e comunicação (Bellani, Ramadhani & Tamar, 2017). Cumulativamente, a Organização Internacional do Trabalho (2020), OIT, refere que “as

² https://cscmp.org/CSCMP/Educate/SCM_Definitions_and_Glossary_of_Terms.aspx

condições de trabalho abrangem uma vasta gama de temas e questões, desde o tempo de trabalho (horas de trabalho, períodos de descanso e horários de trabalho), até à remuneração, bem como as condições físicas e as exigências mentais que existem no local de trabalho”.

Sabendo que a pandemia de SARS-COV2 obrigou muitas empresas a fazerem mudanças, nomeadamente a implementação/expansão do trabalho à distância ou trabalho remoto (Hamouche, 2021), a *sentiment analysis* aparenta ser uma ferramenta muito útil na gestão de recursos humanos, procurando explicitar a perceção dos empregados não só face à satisfação individual, como às eventuais alterações nas condições de trabalho.

Capítulo 3

Sentiment Analysis

Assim, como o seu nome implica, a análise de sentimentos tem como objetivo a determinação da polaridade de uma opinião, ou seja, para aferir se uma opinião sobre um determinado assunto é positiva, neutra ou negativa (Turney, 2002; Micu et al., 2017; Mantyla, Graziotin & Kuutila, 2018).

A SA, tal como a conhecemos hoje, teve o seu surgimento no início dos anos 2000 (Turney, 2002; Pang, Lee & Vaithyanathan, 2002; Dave, Lawrence & Pennock, 2003; Nasukawa & Yi, 2003). Embora fosse um conceito praticamente novo, a área de investigação continuou sem ter grande relevância por mais alguns anos. Por volta de 2006, houve um aumento de estudos publicados devido à evolução da internet e ao surgimento das redes sociais (Liu, 2015; Mantyla, Graziotin & Kuutila, 2018).

Como a SA está diretamente ligada a qualquer tipo de interação com texto, que pode estar em documentos nos formatos *PDF*, *HTML*, *XML*, *TXT*, *DOC*, entre outros (Feldman, 2013; Kaur & Bhatia, 2016), e face ao aumento da presença *online* (quer ao nível pessoal, quer empresarial), seja nas redes sociais, nos *blogs*, *websites* de notícias ou noutra tipo de plataforma, em que seja possível haver interação virtual entre os utilizadores, é expectável que a sua utilização por parte dos gestores tenha aumentado também ao longo do tempo. Contudo, nem a extensão das áreas de aplicação na gestão, nem as fontes textuais utilizadas para aplicação da SA são perfeitamente conhecidas, pelo que este trabalho procura preencher esta lacuna de conhecimento.

3.1. Níveis de análise

Segundo vários autores (Pang, Lee & Vaithyanathan, 2002; Liu, 2012; Feldman, 2013; D'Andrea et al., 2015; Hussein, 2016; Alqaryouti, 2017), os sentimentos podem ser analisados em três níveis diferentes, sendo eles: (1) o nível documental, (2) o nível da frase e (3) o nível de aspeto/entidade.

- 1. Nível documental** – Neste nível, o documento (texto) é analisado como um todo, ou seja, assume-se que o documento todo é sobre uma única entidade, isto é, um produto, serviço, assunto/tema, pessoa, empresa ou evento (Liu, 2012, p.11), e determina-se a polaridade (positivo, neutro, negativo) para o documento inteiro (Pang, Lee & Vaithyanathan, 2002; Turney, 2002). Isto faz com que este método de análise seja inadequado para documentos que avaliam ou comparam mais do que uma entidade (Liu, 2012; Hussein, 2016); Por exemplo, se o documento a analisar fosse sobre uma revisão de um filme, seria dada uma nota global para a revisão inteira, sem considerar outros aspetos como o género, a banda sonora, o desempenho dos atores, a qualidade de imagem, o enredo, etc.;
- 2. Nível da frase** – Neste nível, o documento é repartido em frases e cada frase é analisada separadamente (Feldman, 2013). Pressupõe-se que cada frase tem apenas uma opinião e

uma entidade, logo não é tido em conta qualquer outro tipo de informação sobre o alvo da opinião. No final, por exemplo, as polaridades de todas as frases podem ser somadas para classificar o documento inteiro, ou podem ser utilizadas de forma individual (Wankhade, Rao & Kulkarni, 2022). Cada frase recebe uma classificação, mas antes de analisar a polaridade da frase, primeiro determina-se se uma frase é subjetiva ou objetiva (Liu, 2012; Hussein, 2016). As frases subjetivas são aquelas que contêm sentimentos ou opiniões próprias como crenças, convicções, suposições, desejos, emoções, etc., enquanto uma frase objetiva é aquela que tem teor factual ou estatístico (Pang & Lee, 2008). Só as frases subjetivas é que são analisadas;

- 3. Nível de aspeto/entidade** – Os dois níveis anteriores só funcionam bem quando são avaliados documentos ou frases que se referem a uma única entidade, por isto considera-se que são insuficientes para aplicação (Alqaryouti, 2017). Neste nível, a opinião é decomposta em polaridade dos sentimentos (positivo, neutro, negativo) e é feita a identificação do alvo da opinião (entidade), ao contrário do que acontece nos outros níveis. O objetivo deste terceiro nível de análise é determinar a polaridade dos sentimentos das entidades e dos seus atributos, características ou aspetos (Liu, 2012, 2015; Feldman, 2013). O titular de uma opinião pode estar satisfeito com um dado produto, mas insatisfeito com alguns aspetos do produto. Liu (2012, p. 5), a frase “*A qualidade de chamada do iPhone é boa, mas a sua duração de bateria é curta*” avalia o *iPhone* (entidade) e dois aspetos: a qualidade de chamada e o tempo de vida da bateria. No que concerne à polaridade dos sentimentos no exemplo dado, o sentimento associado à qualidade de chamada é positivo, mas o que se refere ao tempo de vida da bateria é negativo.

3.2. Etapas do processo

O processo de análise de um corpo de texto passa por várias etapas, ocorrendo o mesmo em quatro fases: a (1) recolha de dados; o (2) pré-processamento de texto; a (3) extração de aspetos/caraterísticas e seleção; e a (4) análise dos sentimentos (Kaur, Mangat & Nidhi, 2017; Mukhopadhyay, 2018; Subbaraju et al., 2020).

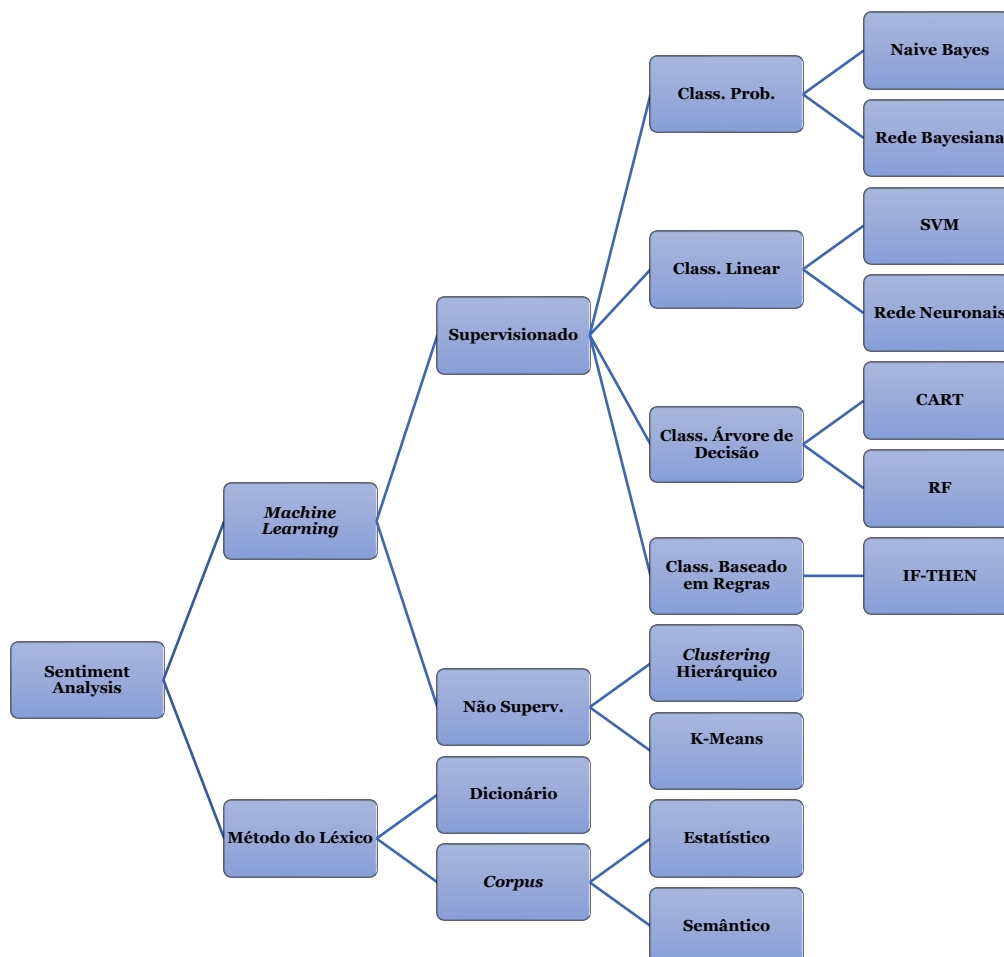
- 1. Recolha de dados** – Os dados de análise podem recolhidos a partir da internet, nomeadamente de *blogs*, fóruns de internet, e redes sociais como o *Twitter* e o *Facebook*, por exemplo. (Subbaraju et al., 2020). Uma das formas mais comuns para aceder aos dados textuais disponíveis na internet é através da utilização de *Application Programming Interfaces* (APIs) desde que se utilizem os protocolos de internet adequados, como explica Mukhopadhyay (2018). Contudo quaisquer dados podem ser utilizados desde que estejam armazenados em formato de texto (Feldman, 2013).
- 2. Pré-processamento de texto** – Neste passo, o texto tem de passar por um processo de limpeza, onde é removida toda a informação desnecessária para o processo de classificação (Alqaryouti, 2017; Kaur, Mangat & Nidhi, 2017), englobando os seguintes procedimentos:

- a. *Tokenization* – o texto é separado em segmentos chamados *tokens*. *Tokens* são unidades linguísticas como palavras, números, datas, pontuação, símbolos outras unidades significativas (Alqaryouti, 2017). Por norma ocorre ao nível da frase;
 - b. *Stemming* – processo que encurta ou reverte a palavra para sua forma de base (Kaur, Mangat & Nidhi, 2017). Por exemplo a palavra *caindo*, seria revertida para a sua forma de raiz *cair*;
 - c. *Stop words removal* – remoção de palavras que não transmitem qualquer sentimento à frase (Rajput, Haider & Ghani, 2016). Alguns exemplos deste tipo de palavra são os pronomes (eu/tu/aquele/contigo), artigos (a/o) e preposições (por/de/com);
 - d. *Part-of-speech (POS) tagging* – é o processo de classificar cada palavra consoante à sua classe gramatical (adjetivo, advérbio, verbo, substantivo, pronome, etc.), segundo Alqaryouti (2017). A *POS Tagging* é muito útil para o processamento em *Machine Learning*, sem esquecer que a identificação de adjetivos e advérbios também são úteis para determinar a polaridade e a subjetividade (Mukhopadhyay, 2018).
- 3. Extração de aspetos/caraterísticas e seleção** – depois de filtrar o texto, segue-se a extração de aspetos/caraterísticas. A análise de sentimentos tem como objetivo analisar a polaridade de uma opinião emitida sobre um(a) aspeto/caraterística de um determinado assunto (*e.g.*, um produto). Portanto, a extração de aspetos/caraterísticas tem como objetivo a sua identificação (Mukhopadhyay, 2018). No fundo, os aspetos ou as caraterísticas que se pretende extrair e selecionar são aqueles que tiveram algum sentimento expressado sobre eles. Também podemos chamar este passo de deteção de sentimentos (Subbaraju et al., 2020). Os elementos principais que se analisam para fazer a extração são:
- a. *N-Gram* – que determina as caraterísticas do conjunto de treino dos algoritmos de aprendizagem supervisionada de *Machine Learning* (Ahuja et al., 2019).
 - b. *Term Frequency-Inverse Document Frequency* – que é um método para avaliar a importância de uma dada palavra num corpo de texto (Ahuja et al., 2019).
 - c. *Negation handling* – palavras como *não, nunca, nem, jamais, nenhum, etc.* são palavras que podem alterar o sentido de uma frase, então este elemento visa medir o impacto da negação na frase e inverter a polaridade das palavras por ela(s) afetadas (D'Andrea et al., 2015; Farooq et al., 2017; Kaur, Mangat & Nidhi, 2017).
- 4. Análise de sentimentos** – é o último passo do processo e é onde acontece a determinação da polaridade do texto. As técnicas usadas para este processo constituem a próxima secção.

3.3. Classificação de sentimentos

Existem duas técnicas que são utilizadas para classificar os sentimentos: a *Machine Learning* (ML) ou o Método Baseado no Léxico³. Contudo, alguns autores advogam a possibilidade de utilizarem uma forma híbrida, combinando as anteriores (D’Andrea et al., 2015; Rajput, Haider & Ghani, 2016; Alqaryouti, 2017; Sadia, Khan & Bashir, 2018; Suryachandra & Reddy, 2019; Subbaraju et al., 2020; Mehta & Pandya, 2020; Gupta & Agrawal, 2020). D’Andrea et al. (2015) explicam que a principal vantagem da abordagem híbrida é a simbiose do método baseado no léxico com a ML, enquanto a principal limitação são os textos com muito ruído (palavras irrelevantes para o assunto do texto em análise), o que faz com que seja frequentemente atribuída uma pontuação neutra porque o método não deteta qualquer sentimento (a Figura 1, apresenta um resumo das técnicas principais).

Figura 1. Métodos e técnicas de SA



Fonte: Adaptado de Mehta & Pandya (2020)

³ Do inglês *lexicon-based method*

3.3.1 Machine Learning

O propósito desta técnica é treinar as máquinas (programas), para que os algoritmos que implementam aprendam por intermédio de repetições. Estas repetições representam a aprendizagem automática, porque a cada vez que o algoritmo percorre o conjunto de dados, ele ajusta o seu modelo de representação de conhecimento, adquirindo a capacidade de efetuar previsões com alguma exatidão em situações futuras com base na informação proveniente dos dados anteriormente analisados (Jordan & Mitchell, 2015). Este processo implica o estabelecimento de regras lógicas que são geradas de acordo com o reconhecimento de padrões existentes dentro do conjunto de dados (Mahesh, 2020).

Existem duas abordagens principais que são aplicadas na abordagem de ML: a Aprendizagem Supervisionada e a Aprendizagem Não Supervisionada (Stimpson & Cummings, 2014; Alqaryouti, 2017; T.K., Annavarapu & Bablani, 2021), embora alguns autores (Mahesh, 2020; Zhang, Li & Wu, 2020) defendam que exista ainda uma terceira abordagem que é a junção das outras duas (aprendizagem semi-supervisionada).

3.3.1.1. Aprendizagem supervisionada

Caraterizada pelo facto de serem conhecidos ou definidos antecipadamente os resultados corretos das previsões, isto é, rótulos, esta abordagem é bastante utilizada para treinar máquinas a partir de dados rotulados (Subbaraju et al., 2020; Mehta & Pandya, 2020). No início, são fornecidos dois conjuntos de dados: um conjunto de treino e um conjunto de saída. Para cada elemento do conjunto de treino, existe apenas um rótulo correspondente no conjunto de saída. Como a SA é feita através da análise textual, as variáveis do conjunto de treino seriam, por exemplo, palavras e o conjunto de saída seria constituído pelas opções possíveis de polaridade de sentimentos (positivo, neutro ou negativo). Este tipo de aprendizagem é utilizado principalmente para realizar tarefas relacionadas com previsão (Stimpson & Cummings, 2014).

É possível categorizar os algoritmos pelo tipo de classificadores envolvidos (Suryachandra & Reddy, 2019; Subbaraju et al., 2020; Mehta & Pandya, 2020): (1) classificadores probabilísticos, (2) classificadores lineares, (3) classificadores de árvores de decisão, e (4) classificadores baseados em regras.

1. **Classificadores probabilísticos** – aqueles que preveem a probabilidade de dados pertencerem a certas categorias (Mehta & Pandya, 2020), como por exemplo o *Naive Bayes* (Subbaraju et al., 2020) e a rede Bayesiana;
2. **Classificadores lineares** – os que determinam em que tipo de grupo/categoria é que um tipo de dados se encaixa, através da utilização de equações lineares para analisar as características (atributos) dos dados do conjunto representados em vetores (Suryachandra & Reddy, 2019), como o *Support Vector Machine (SVM)* (Mountrakis, Im & Ogole, 2011; Addan, 2019) ou as redes neuronais (Mehta & Pandya, 2020).
3. **Classificadores de árvores de decisão** – que são os modelos gráficos que descrevem decisões e os seus possíveis resultados através da tecnologia de ramificação (Mehta & Pandya, 2020; T.K., Annavarapu & Bablani, 2021), por exemplo usando a *Classification*

and *Regression Tree* (CART) ou o *Random Forest* (RF) (Murphy, 2012, p. 270; T.K., Annavarapu & Bablani, 2021).

- 4. Classificadores baseados em regras** – são sistemas concebidos para seguir regras pré-estabelecidas para armazenar, ordenar e manipular dados. Ao fazer isto, imitam inteligência humana. Por norma regem-se por regras baseadas em condições do tipo: SE acontece X, ENTÃO faz Y (Mehta & Pandya, 2020). O nome original deste algoritmo em inglês é IF-THEN.

3.3.1.2. Aprendizagem não supervisionada

No método de aprendizagem não supervisionada não existe um conjunto de saída pré-definido com resultados certos (Stimpson & Cummings, 2014; Alqaryouti, 2017; Subbaraju et al., 2020), pelo que o algoritmo não possui um conjunto de resultados corretos para utilizar como referência para aprender, mas apenas o conjunto de entrada (treino).

A aprendizagem não supervisionada pressupõe a descoberta de padrões dentro de um conjunto de dados (Stimpson & Cummings, 2014; Serras, 2015). A forma mais popular de aplicar este tipo de aprendizagem é através de algoritmos de *clustering* (Subbaraju et al., 2020; T.K., Annavarapu & Bablani, 2021), cujo objetivo é medir o que diferencia os objetos uns dos outros para os agrupar posteriormente com os seus semelhantes. Ao fazer isto, o algoritmo ganha a capacidade de aprender sem ajuda e de identificar padrões de similaridade e de diferenças. A implementação da análise de *clusters* é muitas vezes feita com os seguintes algoritmos de clustering: K-Means e *Clustering* Hierárquico (Subbaraju et al., 2020; T.K., Annavarapu & Bablani, 2021).

3.3.2 Método baseado no Léxico

O método baseado no léxico é uma técnica que consiste em analisar um corpo de texto com a finalidade de o classificar através de uma lista pré-definida de palavras, cujas polaridades também já tenham sido pré-determinadas (D’Andrea et al., 2015), chamada léxico de sentimento⁴ (Kaity & Balakrishnan, 2020). Tal lista é construída pela análise dos adjetivos, dos nomes e dos verbos (Taboada, Anthony & Voll, 2006; Silva et al., 2018; Gupta & Agrawal, 2020) a fim de determinar a sua polaridade. No entanto, uma das maiores complexidades deste processo reside no facto de que o significado base das palavras pode não ser suficientemente claro/óbvio, porque pode variar consoante o contexto em que se encontra (D’Andrea et al., 2015).

A conceção de um léxico de sentimentos pode ser feita de forma manual ou de forma automática com a ajuda de computadores (Kim & Hovy, 2006). A criação de um léxico requer muito tempo e uma quantia considerável de recursos (Lo et al., 2017; Kaity & Balakrishnan, 2020). Por esta razão é que muitas das análises de sentimento noutras línguas (Mihalcea, Banea & Wiebe, 2007) são feitas com recurso aos léxicos de sentimentos existentes na língua inglesa, como por exemplo o *SentiWordNet* e o *WordNet* (Denecke, 2008; Gupta & Agrawal, 2020), embora muitos outros existam (por exemplo o *Harvard General Inquirer* (Stone et al., 1966); o *SenticNet* (Cambria et al. 2010), o *Valence Aware Dictionary for Sentiment Reasoning* (Hutto & Gilbert, 2014), o

⁴ Do inglês *sentiment lexicon*.

EmoLex (Mohammad & Turney, 2010), o *Affective Norms for English Words* (Bradley & Lang 1999). Apesar de serem poucos, também existem léxicos noutras línguas como árabe (Al-Twairesh, Al-Khalifa & Al-Salman, 2016), chinês/coreano/japonês (Kim, Li & Lee, 2010), espanhol (Perez-Rosas, Banea & Mihalcea, 2012), hindi (Hassan et al., 2011), português (Pereira, 2021), entre outras (Kaity & Balakrishnan, 2020).

O método baseado no léxico pode ser aplicado com duas abordagens: a de dicionário e a de *corpus* (D'Andrea et al., 2015; Rajput, Haider & Ghani, 2016; Sadia, Khan & Bashir, 2018; Suryachandra & Reddy, 2019; Mehta & Pandya, 2020; Gupta & Agrawal, 2020).

3.3.2.1. Abordagem de dicionário

A Abordagem de dicionário começa com uma lista de palavras recolhidas manualmente, designada por *seed words* (Rajput, Haider & Ghani, 2016).

As *seed words* (palavras de raiz) constituem um conjunto de palavras com fortes conotações positivas ou negativas na sua polaridade (Taboada et al., 2011), como por exemplo *excelente* ou *detestável*. O objetivo desta abordagem é expandir a lista inicial através da procura recursiva de todos os sinónimos e antónimos das *seed words* nos dicionários de sentimentos/léxico de sentimento (Mukhopadhyay, 2018). Este processo continua até que não se consiga encontrar mais nenhum sinónimo ou antónimo para adicionar à lista das *seed words*.

A fraqueza desta abordagem está na sua incapacidade para encontrar palavras de opinião específicas de domínio e contexto (Rajput, Haider & Ghani, 2016; Mukhopadhyay, 2018). Isto é, às vezes esta abordagem não consegue distinguir corretamente o sentido da palavra, porque existem palavras que podem ter uma conotação positiva ou negativa, dependendo do contexto em que forem utilizadas.

3.3.2.2. Abordagem de corpus

A Abordagem de *corpus* vem dar resposta à esta limitação da Abordagem de dicionário ao tentar descobrir a orientação semântica das palavras específicas do contexto através da identificação do padrão sintático ou de coocorrências no texto em relação às *seed words*, utilizando restrições linguísticas (Rajput, Haider & Ghani, 2016; Alqaryouti, 2017; Gupta & Agrawal, 2020). Geralmente, para determinar a orientação semântica das palavras utilizam-se algoritmos de aprendizagem sequencial como o *Conditional Random Field* (Lafferty, McCallum & Pereira, 2001). A Abordagem de *corpus* é adequada para a criação e visualização de mapas de relações comparativas, que são frequentemente utilizados como um instrumento importante na área da gestão de riscos empresariais e da tomada de decisões (Mukhopadhyay, 2018) Esta abordagem subdivide-se em dois métodos: (1) o estatístico e (2) o semântico.

1. Método Estatístico – neste método, a frequência com que uma dada palavra aparece no corpo de texto é um forte indicador da polaridade, ou seja, se o texto for maioritariamente positivo, a palavra será positiva. Se a palavra aparecer várias vezes repetidas num texto maioritariamente negativo, então a sua polaridade também será negativa (Kaur, Mangat

& Nidhi, 2017). Se o número de vezes (frequência) que a palavra ocorre em contexto positivo for igual ao que aparece em texto negativo, então a palavra será neutra (Gupta & Agrawal, 2020);

2. Método Semântico – neste método, atribuem-se polaridades semelhantes às palavras que são semanticamente parecidas (Mukhopadhyay, 2018) e pode ser feito encontrando os sinónimos/antónimos da palavra que se pretende classificar (Gupta & Agrawal, 2020).

Capítulo 4

Recolha, apresentação e análise dos dados

Para que fossem atingidos os objetivos deste estudo, empregou-se o método de revisão sistemática. Recorreu-se a este método por acreditar-se que é direto, justo, transparente e facilmente reproduzido (Cook, Mulrow & Haynes, 1997; Tranfield, Denyer & Smart, 2003; Crossan & Apaydin, 2010). Para tal efeito, realizou-se uma pesquisa documental na *Web of Science* (WOS), com pesquisa inicial por *Topic*, com o objetivo de filtrar os resultados inserindo a expressão “*sentiment analysis*” (no *Article title*, *Abstract* e *Keywords*). De modo a restringir os resultados ao âmbito aos objetivos deste estudo, limitou-se a pesquisa às áreas de *Operations Research*, *Management Science*, *Business*, *Management*, *Economics* e *Business Finance*.

Após a limitação dos resultados, a próxima fase foi a seleção dos artigos que seriam analisados pela leitura do *abstract* e, caso este não fosse suficientemente esclarecedor, procedia-se a uma leitura do artigo. Procurou-se que todos os artigos tivessem uma aplicação concreta da *sentiment analysis* em tarefas características das funções empresariais, ao invés de uma abordagem meramente contextual (por exemplo os que apresentavam temas apenas com a mera indicação que poderiam ser aplicados pela gestão, mas sem implementação efetiva de tais considerações).

A Figura 2 esquematiza todo o processo descrito.

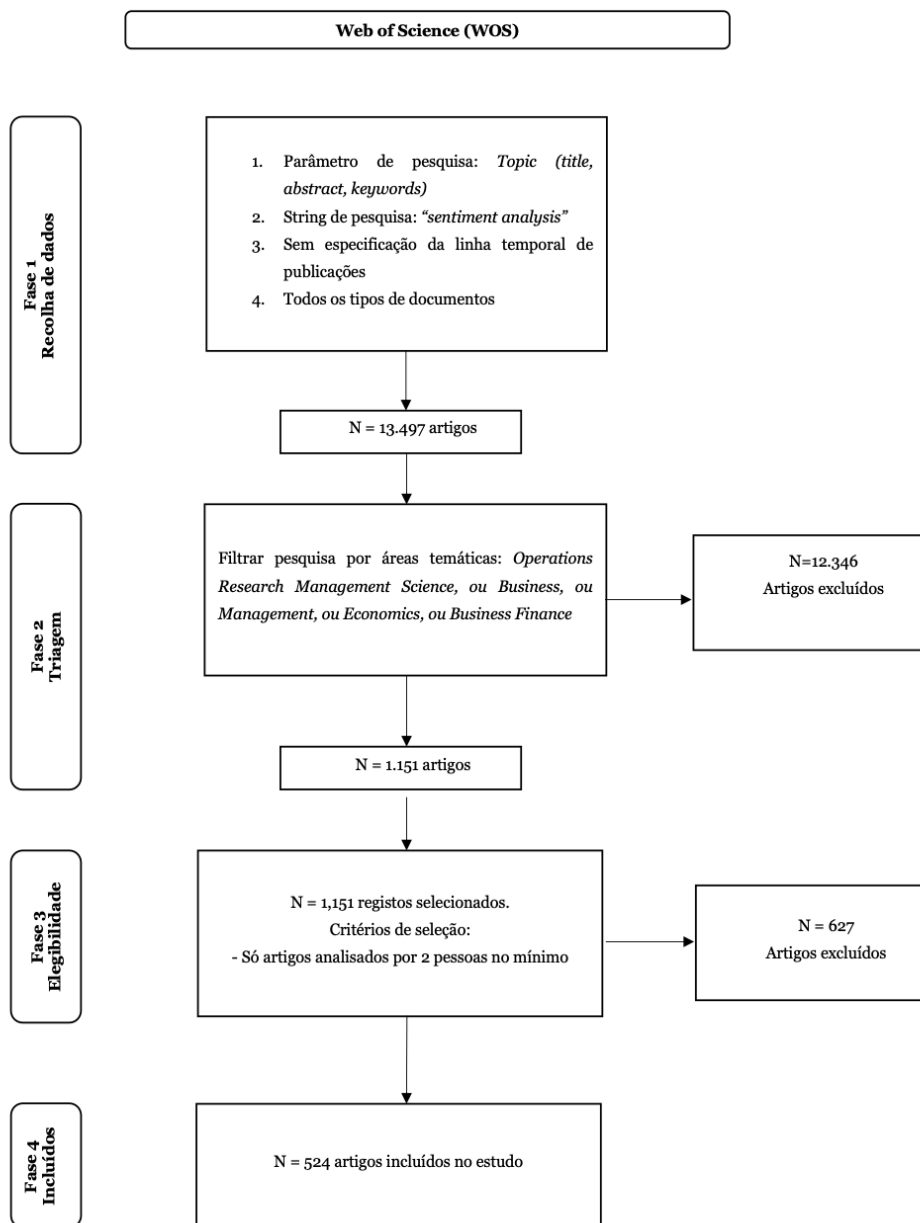
4.1. Análise Bibliométrica

De modo a perceber o panorama geral de aplicação da SA nas funções empresariais, utilizou-se a versão 1.6.18. do VOSViewer (Van Eck & Waltman, 2010) para realizar uma análise bibliométrica.

A noção de análise bibliométrica, inventada e utilizada pela primeira vez por Pritchard (1969), denota a evolução ou o estado da literatura de uma determinada área de investigação. A bibliometria tem sido utilizada para medir a qualidade, o impacto e a influência de autores, revistas e instituições com base em publicações (Lowry, Romans & Curtis, 2004; Lowry et al., 2013; Hassan & Loebbecke, 2017). À vista disso, neste estudo pretendeu-se fazer as seguintes análises:

- Análise de coocorrência – para identificar os principais *clusters* de funções empresariais ligadas à aplicação da SA, pelos termos mais frequentes;
- Análise dos autores mais ativos – para medir a produtividade de autores, pelo número de artigos;
- Análise dos periódicos e das conferências com mais publicações – para medir a relevância das conferências e periódicos que impulsionam a divulgação do tema, com base no número de obras publicadas;
- Produtividade anual – para avaliar o crescimento da aplicação da SA na gestão, através da quantificação de artigos publicados anualmente sobre o tema em estudo.

Figura 2. Esquema de pesquisa



Fonte: Elaboração própria

Atendendo a que o VOSviewer suporta a leitura dos ficheiros extraídos da WOS, criou-se um ficheiro com os artigos selecionados para o estudo e foi escolhida a opção de criar um mapa de termos coocorrentes baseado em dados de texto, selecionando-se a opção "title and abstract fields". Atendendo a que o objetivo desta análise é determinar a frequência com que uma palavra aparece simultaneamente em corpos de texto diferentes (*e.g.*, artigos) optou-se pelo método de contagem binária, uma vez que no *binary counting* apenas é registado o número de documentos em que um termo ocorre, independentemente de quantas vezes o termo aparece em cada documento (Ibekwe, Bochi & Martínez-Ávila, 2021).

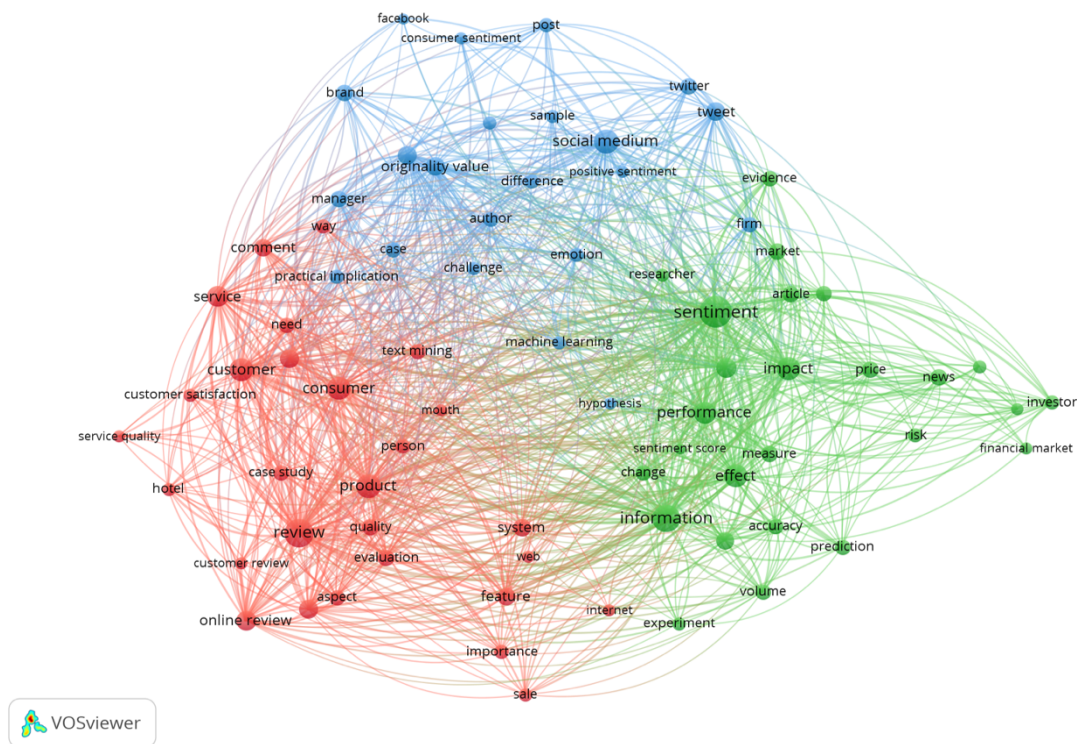
4.1.1 Principais *Clusters*

Do registo amostral constavam 11.157 termos, dos quais 76 termos com 19 ocorrências no mínimo integraram o estudo e foram divididos em *clusters* que representam as áreas funcionais em que a SA é aplicada.

Os dez termos mais frequentes foram: *sentiment* (214), *review* (165), *social medium* (115), *product* (111), *consumer* (105), *impact* (102), *customer* (98), *effect* (97), *performance* (88) e *service* (82). Estes termos ajudaram a indicar os aspetos mais importantes em que a SA tem sido estudada/aplicada na gestão.

A Figura 3 ilustra a rede das palavras mais frequentes. No mapa, o tamanho dos nós é proporcional ao número de repetições do termo em causa, enquanto as linhas representam as ligações entre palavras que ocorrem simultaneamente e a proximidade dos nós, representa a força da ligação entre elas (Van Eck & Waltman, 2022).

Figura 3. Rede de palavras com coocorrência



Fonte: Elaboração própria

Utilizando o método de análise de coocorrência, o resultado da análise feita pelo VOSviewer identificou três grupos distintos de palavras. A análise destes *clusters* está representada na Tabela 1.

Tabela 1. Clusters correspondentes às funções de gestão

Cor	Itens [N]	Palavras com coocorrência	Nome do cluster
Vermelho	30	<i>review (165), product (111), consumer (105), customer (98), service (82), online review (74), experience (63), rating (58), comment (54), system (52), industry (45), evaluation (43), need (42), quality (42), case study (40), aspect (39), text mining (38), way (37), importance (32), person (32), attribute (31), hotel (30), sale (27), customer satisfaction (25), internet (25), mouth (25), service quality (23), web (23), customer review (22), tourist (20)</i>	Função de marketing
Verde	27	<i>sentiment (214), impact (102), effect (97), performance (88), relationship (74), influence (50), market (49), news (49), price (47), return (46), measure (45), article (44), evidence (42), accuracy (41), period (40), volume (40), prediction (39), firm (39), investor (35), risk (33), experiment (31), researcher (30), stock (28), stock market (25), hypothesis (24), financial market (23), sentiment score (21)</i>	Função financeira
Azul	19	<i>social medium (115), tweet (63), originality value (56), design methodology approach (54), manager (51), twitter (50), brand (47), case (37), post (35), author (33), difference (30), practical implication (30), sample (29), negative sentiment (26), consumer sentiment (23), machine learning (23), positive sentiment (23), facebook (19), covid (16)</i>	Fontes de informação

Fonte: Elaboração própria

O **primeiro cluster** é o que detém o maior número de itens. Os termos *review, online review, customer satisfaction, customer review* e *rating* por chamam-nos a atenção para uma visão de avaliação e classificação *online* sobre algum aspeto de um determinado serviço, que são essencialmente características do marketing digital (Chung & Tseng, 2012). King, Racherla & Bush (2014) explicam que as *online reviews* são exemplo de um tipo de *electronic word-of-mouth* (eWOM). Westbrook (1987, p.261) define *word-of-mouth* (WOM) como “comunicações informais direcionadas a outros consumidores sobre a propriedade, utilização ou características de determinados bens e serviços e/ou dos seus vendedores”. Portanto, o eWOM é nada mais e nada menos do que a versão WOM da internet (Hennig-Thurau et al., 2004). Ao mesmo tempo, os *ratings* (avaliações/classificações) atribuídas *online* refletem o nível e o grau de satisfação do cliente ou consumidor em relação a algum serviço oferecido.

Atualmente, existem variadas formas de demonstrar este tipo de sentimento. Alguns dos métodos mais conhecidos são: o sistema de classificação de cinco estrelas (Hogenboom, Boon & Frasinicar, 2012), dos emoticons (Hogenboom et al., 2013; D’Andrea et al., 2015) e dos emojis (Yoo & Rayz, 2021). Simultaneamente, a análise de mercado é outra área em que a SA pode aproveitada na gestão de empresas, porque ela está centrada na análise das avaliações *online* sobre produtos ou serviços para identificar potenciais concorrentes, clientes valiosos e avaliar a imagem da marca (Kumar, Kar & Ilavarasan, 2021).

O **segundo cluster** é constituído maioritariamente por termos relacionados com aspetos de natureza financeira. Sabe-se que os investidores são entidades racionais, usualmente não emocionais e que os preços representam o equilíbrio dos retornos esperados e dos possíveis riscos (Yadav et al., 2020). Aparentemente, isto faria com que não houvesse margem para que os sentimentos do investidor desempenhassem um papel na determinação de preços. Contudo, ainda segundo Yadav et al. (2020), as finanças contemporâneas, reconhecem tanto os investidores sentimentais, como os investidores racionais. Esta ideia reforça a afirmação de De Long et al. (1990) que diz que o conceito de sentimento de mercado assenta no pressuposto

fundamental de que as decisões dos investidores são impulsionadas pelos seus sentimentos, pelo que a aplicação da SA faz todo o sentido no domínio financeiro, mais propriamente na análise de notícias sobre ele (Yadav et al., 2020; Carosia, Coelho & Silva, 2021). Assim, a presença do termo *news* neste *cluster* faz todo o sentido, uma vez que os analistas dos mercados financeiros fazem previsões com base em informações e opiniões provenientes de notícias (Biswas et al., 2020). De forma semelhante, os programas inteligentes podem extrair a informação textual a partir da internet, em fóruns de notícias financeiras, e proceder à classificação do texto, revelando o sentimento que expressam.

O **terceiro cluster** não revela funções empresariais, apresentando antes um agrupamento relacionado com as fontes de informação utilizadas para realizar sobre elas a SA, indicando, nomeadamente, a importância das redes sociais para este processo (neste *cluster*, o termo que mais se destaca é o termo *social medium* com 112 coocorrências).

No âmbito da gestão de empresas, as redes sociais têm um papel fundamental na análise e extração de opiniões expressas em textos, porque elas permitem estabelecer uma ligação direta com os clientes, e oferecem às empresas a possibilidade de fortalecerem mais ainda esta ligação (Chatterjee & Kar, 2020). Por essa razão o termo *social medium* é representado pelo maior nó da rede azul no mapa de palavras. O *Twitter*, o *Facebook*, o *YouTube*, os *blogs* e *podcasts*, por exemplo, são meios bastante utilizados para este fim (Li et al., 2018; Koukaras, Tjortjis & Rousidis, 2020; Yao et al., 2021). Do mesmo modo, verifica-se a presença de termos como *consumer sentiment*, *brand*, *positive sentiment* e *negative sentiment* neste aglomerado. Numa primeira instância, ao olhar para estes termos percebe-se que se pode fazer um cruzamento com as redes sociais, pois muitas empresas servem-se das redes sociais para aferir o sentimento dos consumidores associado à alguma característica de um produto antes e/ou depois do lançamento do mesmo (Rathore & Ilavarasan, 2020). Por exemplo, Feldman et al. (2008) empregam a SA a fim de tentar perceber como é que os consumidores encaram o mercado de produtos, através de um caso de estudo em que comparavam as opiniões, satisfações ou insatisfações dos clientes no tocante às características do produto.

4.1.2 Principais Autores

A Tabela 2 destaca os autores mais proeminentes em função do número de obras publicadas sobre implementação da SA na gestão com base nos dados recolhidos da WOS. Ao todo, foram identificados 513 autores⁵.

⁵ Atendendo às restrições de espaço optou-se por uma apresentação mais condensada. A lista completa pode ser visualizada em <https://ibb.co/YyPGRk0>.

Tabela 2. Os 20 autores com mais artigos publicados sobre a SA na gestão

Posição	Autor(es)	Artigos [N]
1	Fernandez-Gavilanes, M., Juncal-Martinez, J., Garcia-Mendez, S., Costa-Montenegro, E. & Gonzalez-Castano, F. J.	4
2	Mostafa, M. M.	3
3	Chen, W. H., Lai & K.K., Cai, Y.	2
4	Kazmaier, J. & van Vuuren, J. H.	2
5	Nakayama, M. & Wan, Y.	2
6	Nemzer, L. R. & Neymotin, F.	2
7	Singh, A., Jenamani, M., Thakkar, J. J. & Rana, N. P.	2
8	Tafesse, W.	2
9	Aakash, A. & Aggarwal, A. G.	1
10	Abbasi-Moud, Z., Hosseinabadi, S., Kelarestaghi, M. & Eshghi, F.	1
11	Abbasi-Moud, Z., Vahdat-Nejad, H. & Sadri, J.	1
12	Abdi, A., Shamsuddin, S. M., Hasan, S. & Piran, M. J.	1
13	Abrahams, A.S., Jiao, J., Wang, G. A. & Fan, W. G.	1
14	Afanasyev, D.O., Fedorova, E. & Ledyeva, S.	1
15	Afego, P. & Alagidede, I.	1
16	Agarwal, S.	1
17	Agarwal, S., Kumar, S. & Goel, U.	1
18	Agarwal, S., Wang, L. & Yang, Y.	1
19	Agoraki, M. E. K., Aslanidis, N. & Kouretas, G. P.	1
20	Agrawal, S.R. & Mittal, D.	1

Fonte: Elaboração própria

4.1.3 Principais Publicações e Conferências

Os artigos incluídos no estudo foram publicados em periódicos e conferências. Dos 524 artigos, 445 provêm de periódicos e 79 saem de conferências. A Tabela 3 e a Tabela 4 enumeram em ordem decrescente, nomeadamente, os principais 20 periódicos e as principais 20 conferências com publicações relevantes ao tema do presente trabalho⁶.

Tabela 3. Top 20 de periódicos com publicações sobre a SA na gestão

Posição	Periódicos	Publicações [N]
1	Decision Support Systems	24
2	Expert Systems with Application	20
3	Journal of Business Research	14
4	Tourism Management	13
5	Electronic Commerce Research	10
6	Electronic Commerce Research and Applications	10
7	Journal of Retailing and Consumer Services	10
8	Technological Forecasting and Social Change	10
9	International Journal of Contemporary Hospitality Management	9
10	Information & Management	8
11	Computational Economics	7
12	Information Systems and E-Business Management	7
13	Internet Research	6
14	Journal of Organizational and End User Computing	6
15	Tourism Management Perspectives	6
16	Journal of Hospitality and Tourism Management	5
17	Annals of Operations Research	4
18	International Journal of Bank Marketing	4
19	International Journal of Information Technology & Decision Making	4
20	International Journal of Production Research	4

Fonte: Elaboração própria

Examinando a Tabela 3, observa-se que os periódicos que comandam a lista são o *Decision Support Systems* (24), o *Expert Systems with Application* (20) e o *Journal of Business Research*

⁶ Também aqui, pela exiguidade de espaço, optou-se por uma apresentação mais condensada. A lista completa de publicações pode ser visualizada em <https://ibb.co/wNJKPtB> e a das conferências em <https://ibb.co/DbB14tP>.

(14). Os dois primeiros são periódicos pluridisciplinares, ao passo que o outro é exclusivamente sobre temas afetos à gestão de empresas.

Tabela 4. As 20 principais conferências com publicações sobre a SA na gestão

Posição	Conferências	Publicações [N]
1	2015 48th Hawaii International Conference On System Sciences (HICSS)	5
2	Marketing And Smart Technologies, Vol 1	4
3	Enterprise Applications, Markets And Services In The Finance Industry, Financecom 2016	3
4	Tourism Management	13
5	5th International Conference On Information Technology And Quantitative Management, Itqm 2017	2
6	Data Science And Knowledge Engineering For Sensing Decision Support	2
7	Education Excellence And Innovation Management: A 2025 Vision To Sustain Economic Development During Global Challenges	2
8	Proceedings Of Eighteenth Wuhan International Conference On E-Business	2
9	Proceedings Of The 2021 Ieee 24th International Conference On Computer Supported Cooperative Work In Design (CSCWD)	2
10	2008 5th International Conference On Service Systems And Service Management, Vols 1 And 2	1
11	2013 Ieee International Conference On Industrial Engineering And Engineering Management (IEEM 2013)	1
12	2016 23rd Annual International Conference On Management Science & Engineering, Vols. I And II	1
13	2016 International Conference On Management, Economics And Social Development (ICMESD 2016)	1
14	2017 Ieee 2nd International Conference On Big Data Analysis (ICBDA)	1
15	2017 International Conference On Ict For Smart Society (ICISS)	1
16	2018 Ieee International Conference On Industrial Engineering And Engineering Management (IEEE IEEM)	1
17	2018 Portland International Conference On Management Of Engineering And Technology (PICMET '18)	1
18	2019 Basiq International Conference: New Trends In Sustainable Business And Consumption	1
19	2019 IEEE Conference On Computational Intelligence For Financial Engineering & Economics (CIFER 2019)	1
20	2020 IEEE Congress On Evolutionary Computation (CEC)	1

Fonte: Elaboração própria

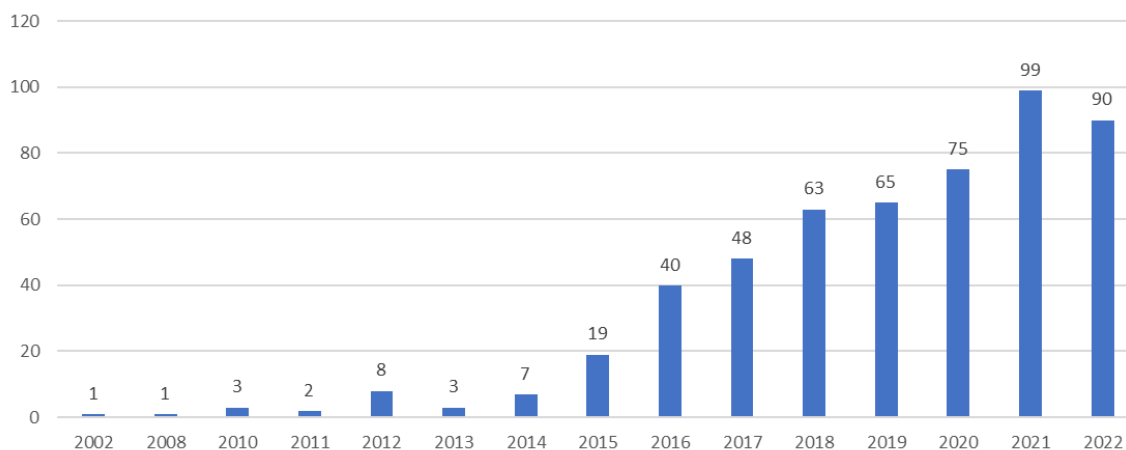
4.2. Temáticas por função

Como indicado no Capítulo 2, parece existir algum consenso em relação à divisão das funções de gestão em cinco funções elementares: administrativa, financeira, de *marketing*, de produção e de recursos humanos. Contudo, tal consenso parece ainda faltar no atinente às temáticas abordadas por cada uma das funções, pelo que importava construir uma taxonomia da utilização/estudo da SA dentro das funções empresariais, através de uma análise sistemática das temáticas abordadas pelos artigos selecionados e, posteriormente, a construção de uma proposta de taxonomia, seguindo o método intuitivo, que assenta na perceção e entendimento do investigador sobre o objeto de estudo, particularmente quando não exista um modelo de análise (Nickerson, Varshney & Muntermann, 2013).

O primeiro passo foi ler cada artigo e extrair as temáticas neles abordados. O segundo passo foi organizar as temáticas identificadas de forma que elas se encaixassem numa das cinco áreas funcionais da gestão de empresas, conforme estabelecido no capítulo 2, gerando a proposta de taxonomia apresentada mais adiante.

Na Figura 4 apresenta-se a evolução do número de artigos publicados, a partir dos artigos incluídos neste estudo.

Figura 4. Número de publicações por anos



Fonte: Elaboração própria (a série de 2022 engloba dados apenas até junho)

Atendendo a que um artigo pode abordar mais do que uma função empresarial, por exemplo um artigo que abrange simultaneamente as funções de *marketing* e de produção, tal artigo não foi contado como “um artigo geral”, mas sim como dois artigos diferentes, em que um seria de *marketing* e o outro seria sobre produção. Portanto, contrariamente ao que acontece na Figura 4, em que os artigos são vistos de forma generalizada, na Tabela 5 as funções empresariais são contabilizadas individualmente, independentemente de aparecerem juntas no mesmo artigo.

Tabela 5. Associação dos artigos incluídos por função empresarial

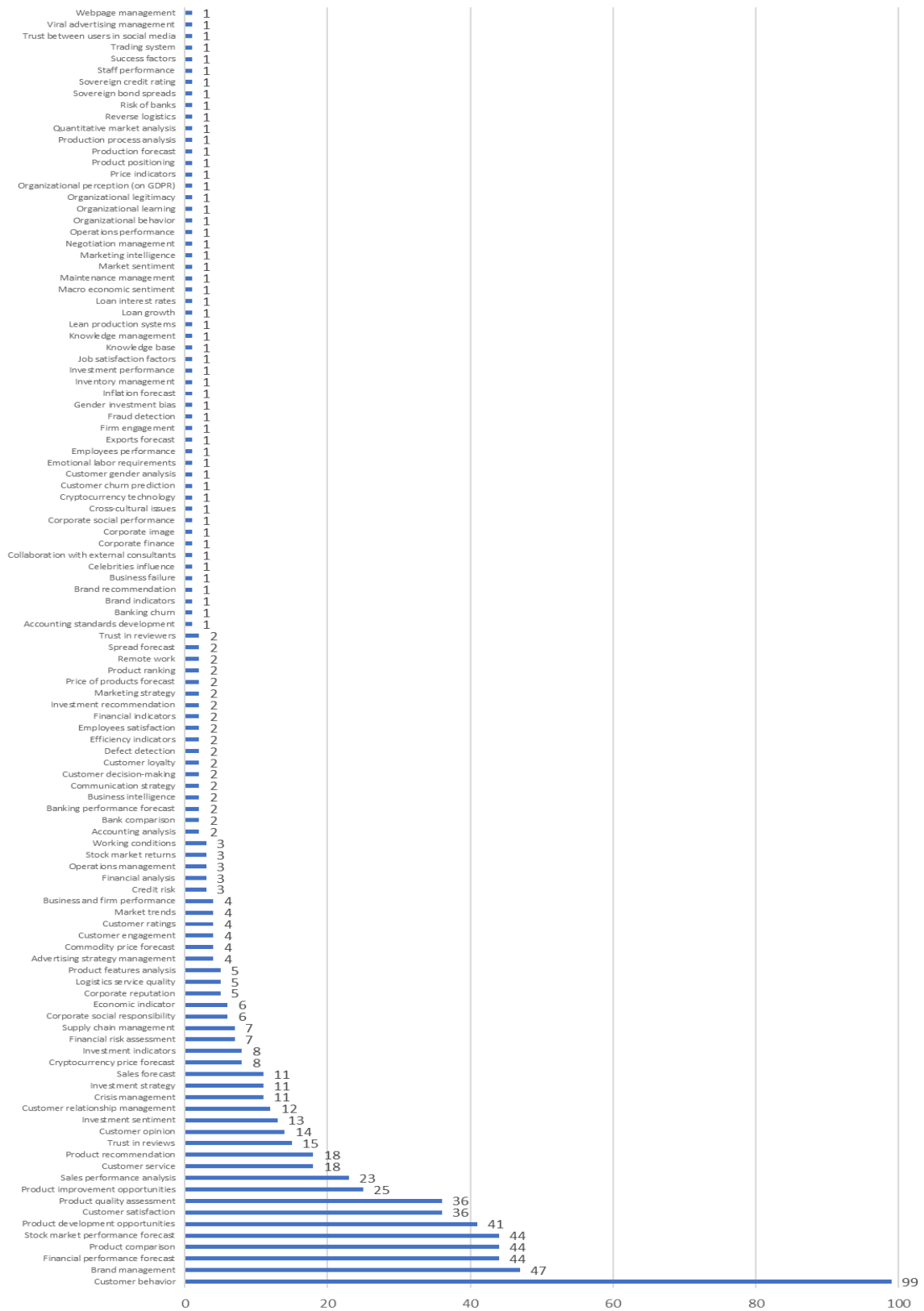
	Administrativa	Financeira	Marketing	Produção	Recursos Humanos	Total
2002	0	1	0	0	0	1
2008	0	0	1	0	0	1
2010	0	0	3	0	0	3
2011	0	0	2	1	0	3
2012	2	1	10	0	0	13
2013	0	2	2	0	0	4
2014	1	5	4	0	0	10
2015	2	6	16	1	0	25
2016	4	13	35	3	0	55
2017	4	18	46	2	1	71
2018	3	24	57	2	2	88
2019	6	22	65	1	2	96
2020	10	21	76	3	0	110
2021	8	35	91	5	5	134
2022	7	32	80	4	1	94*
Σ	47	180	488	22	10	748
%	6,28	24,06	65,24	2,94	1,47	100

Fonte: Elaboração própria (*dados recolhidos até junho 2022)

Na Tabela 5 observa-se que os artigos estão maioritariamente distribuídos entre as funções de *marketing* e financeira. Estes resultados estão em consonância com os resultados da análise

bibliométrica, que também indicam que estas são as duas maiores áreas em que é aplicada a técnica de SA.

Figura 5. Temáticas e número de artigos associados (cada artigo pode ter mais que uma)



Fonte: Elaboração própria

Evidencia-se que o *marketing* é a função empresarial com maior relevância, seguido pela financeira e ambas evidenciam também investigação há mais tempo que as outras. As restantes funções, nomeadamente, a administrativa, a de produção e a de recursos humanos demonstram uma evolução positiva e mais recente, o que parece indicar um estágio inicial de interesse pela aplicação da SA nestas áreas.

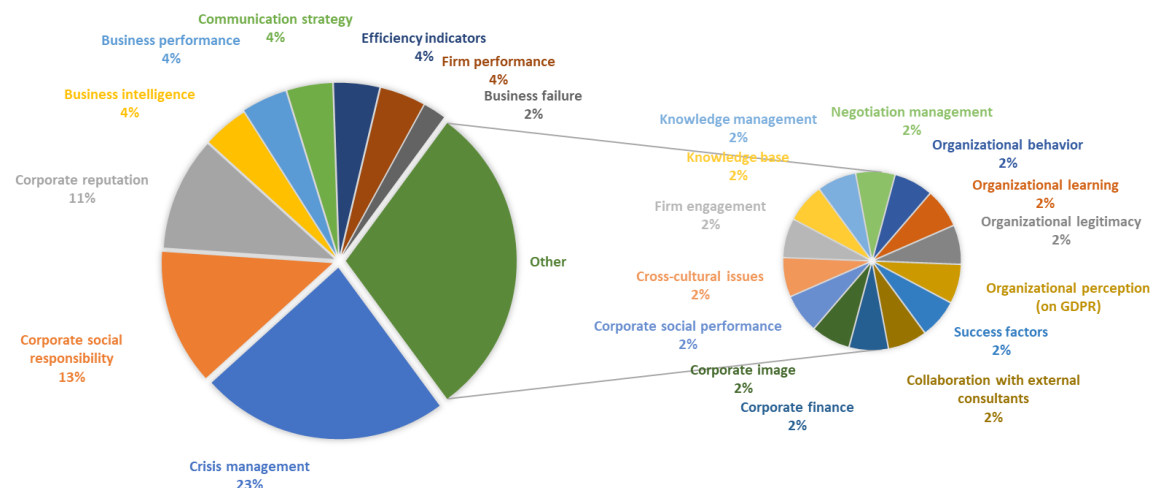
Analisando o gráfico na

Figura 5, observa-se, mais uma vez, que as temáticas identificadas estão maioritariamente distribuídas entre as funções de *marketing* e financeira (a análise mais detalhada é efetuada nas próximas subsecções). Das 748 aplicações (contando com as repetições de temáticas) da SA nas funções empresariais, 488 pertencem à função de *marketing*, 180 para a financeira, 47 para a administrativa, 22 para a de produção e 11 para a de recursos humanos. Estes resultados corroboram novamente os resultados da análise bibliométrica.

4.2.1. Função Administrativa

Na parte da gestão administrativa (Figura 6), verifica-se que o maior ponto aplicação é o da gestão de crises empresas, porque exige uma resposta rápida e assertiva. Para isto, é importante que existam estratégias montadas para qualquer eventualidade. Também serve para mitigar riscos para reduzir a possibilidade de surgir alguma crise. A criação de estratégias também faz parte da gestão administrativa. Ainda dentro desta função, a questão de responsabilidade social corporativa influencia diretamente a reputação da empresa. Uma vez que estes aspetos dependem da interação ou da opinião com os consumidores faz todo o sentido que seja aqui aplicada a SA.

Figura 6. Distribuição de temáticas de aplicação da SA na função administrativa



Fonte: Elaboração própria

Estes são pontos que podem ter algum efeito no sucesso da empresa e, portanto, é necessário que haja sempre um bom relacionamento entre os colaboradores para que se continue a trabalhar de forma eficiente e/ou eficaz de modo a providenciar um bom serviço aos consumidores. Comparativamente às outras temáticas, constata-se que a atenção dada na vertente das relações

interpessoais nas empresas, bem como às questões relacionadas com o conhecimento, tem sido pouca. Ironicamente, estas temáticas estão representadas pela maior fatia (*other*).

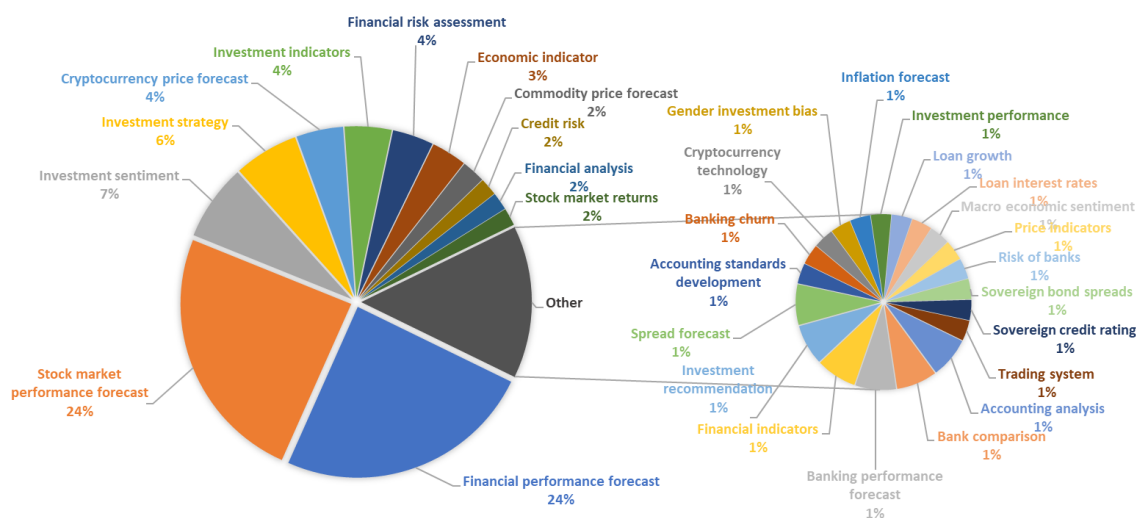
O que se pode concluir a partir daqui é que desde 2019 (Tabela 5), as aplicações da SA na função administrativa têm vindo a crescer e novas formas de as aplicar têm sido descobertas.

4.2.2. Função Financeira

Na função financeira (Figura 7), a SA é predominantemente utilizada nas temáticas de previsão de preços de ações e de desempenho financeiro. Observa-se que seguidamente está a temática de investimentos e a previsão de mercados de criptomoedas são outras áreas que também têm tido utilização de SA, principalmente explorando os sentimentos expressos por investidores. No outro lado do gráfico, percebe-se que estão as temáticas com pouca expressão (*e.g.* conceção de créditos, previsão de taxa de inflação, temas associados à banca, etc.).

Note-se que apesar de representar uma maior porção, os temas em *other* não têm uma grande representatividade (menos de 2%, para cada um). É um indicativo de que estejam a começar a ser explorados.

Figura 7. Distribuição de temáticas de aplicação da SA na função financeira



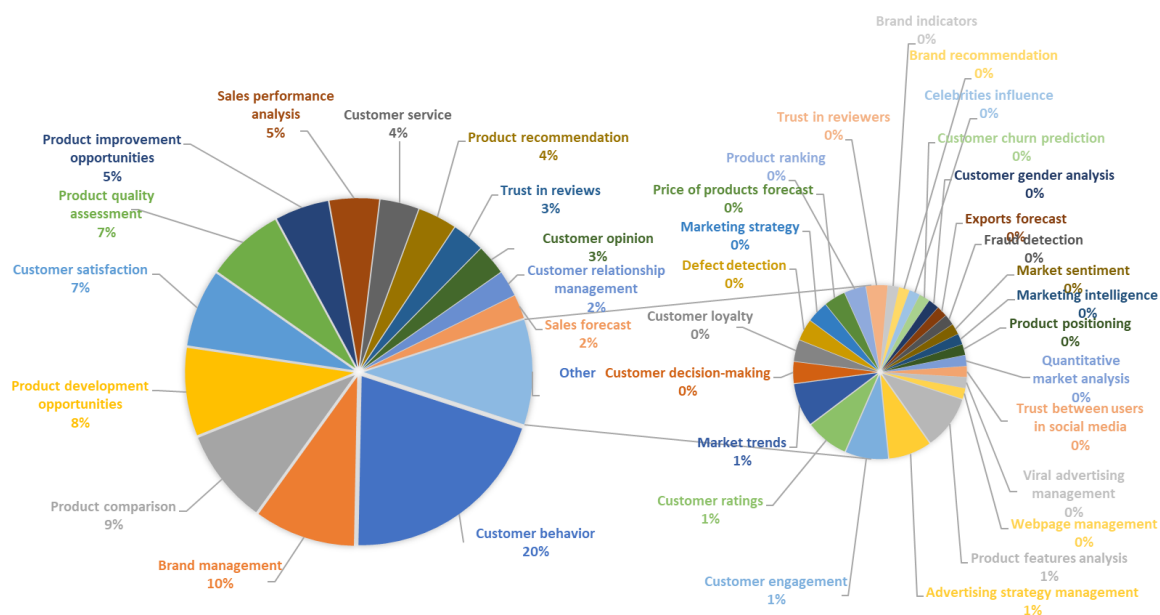
Fonte: Elaboração própria

4.2.3. Função de Marketing

No *marketing* (Figura 8), as temáticas mais populares são as que têm a ver com o comportamento do consumidor, a imagem da empresa e com as revisões de produtos e serviços. Outro aspeto muito importante é que uma parte da avaliação do mercado e da concorrência é feita utilizando o *feedback* recebido nas redes sociais, embora, como será analisado na secção 4.4.3, o recurso a conjuntos de dados (*datasets*) já processados seja a principal fonte de informação para o efeito. O *feedback* pode ser dado diretamente do cliente para a empresa (na página *web* da empresa) ou entre consumidores *online*. Neste sentido, muitos consumidores confiam mais nas opiniões *online* que leem de outros consumidores.

A seguir, repara-se que surgem os temas relacionados com a marca da empresa e os seus produtos, porque um influencia o outro. Note-se também que o *marketing* também possui uma porção *other*, mas não é tão grande quanto às outras (administrativa e financeira). No entanto, a parte curiosa é que o *other* do *marketing* integra uma maior lista de temáticas. Ou seja, é aquela que tem maior diversidade de todas as funções empresariais. Faz todo sentido que assim seja, porque afinal de contas, o *marketing* é a função empresarial com maior aplicação de SA, tal como foi evidenciado pela análise bibliométrica e como será evidenciado na taxonomia proposta no ponto 4.3.

Figura 8. Distribuição de temáticas de aplicação da SA na função de *marketing*

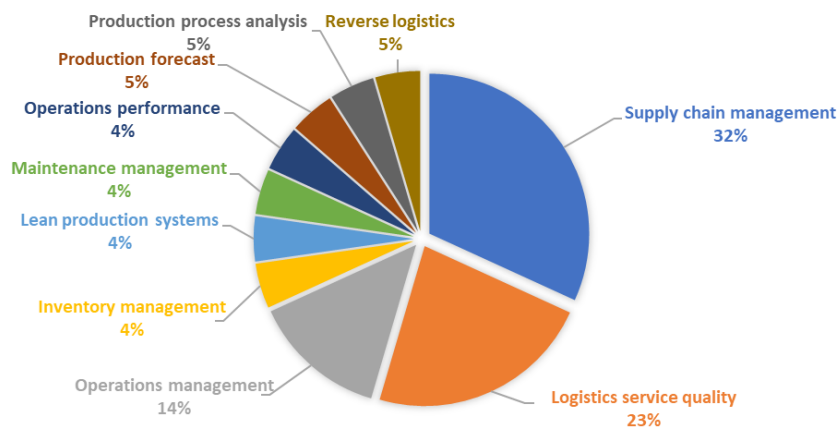


Fonte: Elaboração própria

4.2.4. Função de Produção

No concerne à área de produção (Figura 9), o foco reside principalmente na cadeia de abastecimento, porque a mesma engloba os processos conceção do produto até chegar ao consumidor final.

Figura 9. Distribuição de temáticas de aplicação da SA na função de produção



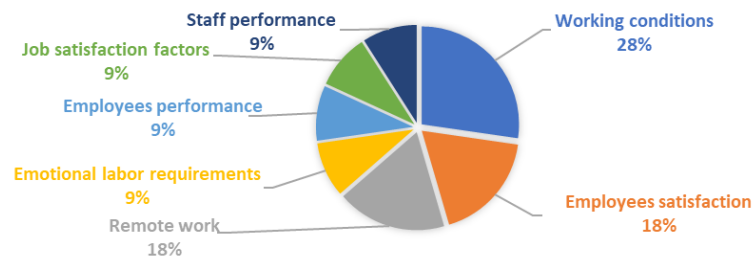
Fonte: Elaboração própria

Nela encontramos a logística, bem como a gestão de *stocks*, que se não for bem feita, pode aumentar custos para empresa devido ao produto extra que vai perdendo qualidade por ficar mais tempo armazenado. Ao contrário das funções empresariais anteriormente discutidas, a produção não tem uma porção *other*. Isto indica que atualmente a aplicação na produção é limitada às temáticas identificadas.

4.2.5. Função de Recursos Humanos

Nos recursos humanos (Figura 10), vê-se que existe claramente uma maior ênfase da utilização da SA para estudar as condições de trabalho. Não é de todo surpreendente, uma vez que nos últimos anos a crise pandémica obrigou as empresas a reestruturarem as condições de trabalho com a inserção do modelo de trabalho à distância. Este ponto tem uma relação direta com os fatores que influenciam na satisfação dos trabalhadores e no desempenho geral da empresa. De ponto de vista analítico, a função de recursos humanos é visivelmente aquela que menos aplicações tem tido. Apesar de serem conhecidas todas as temáticas de aplicação, a utilização da SA no estudo desta função continua a ser pouca ou quase inexistente.

Figura 10. Distribuição de temáticas de aplicação da SA na função de recursos humanos



Fonte: Elaboração própria

4.3. Proposta de Taxonomia

Após a análise individual de cada artigo incluído no estudo, foi possível esquematizar e classificar cada artigo (524 no total). A taxonomia desenvolvida e apresentada na Tabela 6 representa o atual estado do quadro dos campos e formas de aplicação da SA na gestão de empresas. Foram identificadas 15 subfunções e 111 de temáticas de aplicação (contando apenas uma vez cada temática). Esta proposta de taxonomia contém três níveis.

No nível 1 estão as cinco funções empresariais. O nível 2 contém as subfunções de cada função empresarial onde a SA é aplicada. O nível dois foi criado para agrupar temáticas de lógica comum (*e.g.*, não misturando temáticas de cliente, com temáticas de marca). O nível 3 integra as temáticas de cada subfunção.

Esta proposta segue de perto o proposto por Nickerson, Varshney & Muntermann (2013), que referem que uma taxonomia tem de ser obrigatoriamente concisa (1), robusta (2), compreensível (3), extensível (4) e explicativa (5):

1. Concisão – uma taxonomia deve conter um número adequado de dimensões e um número de características em cada dimensão;

2. Robustez – uma taxonomia deve conter dimensões e características suficientes que permitam diferenciar os objetos uns dos outros;
3. Compreensão – uma taxonomia útil deve conseguir classificar todos os objetos do estudo, ou seja, tem de conseguir descrever cada um;
4. Extensibilidade – uma taxonomia deve permitir que sejam adicionadas novas dimensões e características quando surgirem novos tipos de objetos;
5. Poder Explicativo – uma taxonomia útil não é aquela que contém dimensões e características que descrevem todos os detalhes possíveis dos objetos, pelo contrário, fornecem informações úteis sobre a natureza dos objetos em estudo ou de objetos futuros para nos ajudar a compreender os objetos. Ou seja, a informação esquematizada deve permitir sair do geral para o particular.

Por uma questão de clareza, “dimensão” refere-se, neste caso, às temáticas identificadas, enquanto as “caraterísticas” resultam da combinação da função e da subfunção empresarial. Os objetos de estudo são os artigos analisados.

A taxonomia proposta respeita cada um dos 5 objetivos propostos por Nickerson, Varshney & Muntermann (2013), no sentido de que o número de dimensões surge em função dos artigos analisados. Uma vez que as subfunções representam grupos de temáticas assentes na mesma lógica, a distinção entre as diferentes temáticas é facilmente identificada. Por outro lado, na eventualidade de surgir uma temática que não se enquadre em nenhuma das características, automaticamente haverá uma extensão da taxonomia, originando novas dimensões e características. Por conseguinte, obtém-se uma taxonomia que distingue e agrupa temáticas parecidas, mas que deixa claro que não são a mesma coisa (compreensão e poder explicativo). No entanto, é importante salientar que podem existir várias temáticas num só artigo.

Tabela 6. Proposta de taxonomia

Nível 1: Função empresarial	Nível 2: Subfunção	Nível 3: Temática
Função administrativa <i>General management</i>	Negócio e organização <i>Business and organization</i>	Fracasso empresarial <i>Business failure</i>
		Inteligência empresarial <i>Business intelligence</i>
		Desempenho empresarial <i>Business and firm performance</i>
		Colaboração com consultores externos <i>Collaboration with external consultants</i>
		Estratégia de comunicação <i>Communication strategy</i>
		Gestão de crises <i>Crisis management</i>
		Questões interculturais <i>Cross-cultural issues</i>
		Indicadores de eficiência <i>Efficiency indicators</i>
		Finanças empresariais <i>Corporate finance</i>
		Imagem corporativa <i>Corporate image</i>
		Reputação corporativa <i>Corporate reputation</i>
		Desempenho social das empresas <i>Corporate social performance</i>
		Responsabilidade social das empresas <i>Corporate social responsibility</i>

Nível 1: Função empresarial	Nível 2: Subfunção	Nível 3: Temática
		<i>Envolvimento empresarial</i> <i>Firm engagement</i>
		Gestão de negociações <i>Negotiation management</i>
		Comportamento organizacional <i>Organizational behavior</i>
		Legitimidade organizacional <i>Organizational legitimacy</i>
		Aprendizagem organizacional <i>Organizational learning</i>
		Perceção organizacional (sobre o RGPD) <i>Organizational perception (on GDPR)</i>
		Fatores de sucesso <i>Success factors</i>
		Conhecimento <i>Knowledge</i>
		Base de conhecimento <i>Knowledge base</i>
		Gestão de conhecimento <i>Knowledge management</i>
Função financeira <i>Finance and economics</i>	Contabilidade <i>Accounting</i>	Análise contabilística <i>Accounting analysis</i>
		Desenvolvimento de normas contabilísticas <i>Accounting standards development</i>
	Banca <i>Banking</i>	Comparação bancária <i>Bank comparison</i>
		Rotatividade bancária <i>Banking churn</i>
		Previsão do desempenho bancário <i>Banking performance forecast</i>
		Risco dos bancos <i>Risk of banks</i>
	Crédito <i>Credit</i>	Risco de crédito <i>Credit risk</i>
		Variação do nível de empréstimos <i>Loan growth</i>
		Taxas de juro de empréstimos <i>Loan interest rates</i>
		Classificação de dívida soberana <i>Sovereign credit rating</i>
		Spreads de títulos de dívida soberana <i>Sovereign bond spreads</i>
		Previsão da evolução dos spreads <i>Spread forecast</i>
	Criptomoeda <i>Cryptocurrency</i>	Previsão do preço de moeda criptográfica <i>Cryptocurrency price forecast</i>
		Tecnologia de criptomoedas <i>Cryptocurrency technology</i>
	Mercados financeiros <i>Financial markets</i>	Previsão dos preços de bens <i>Commodity price forecast</i>
		Previsão do desempenho financeiro <i>Financial performance forecast</i>
		Análise financeira <i>Financial analysis</i>
		Avaliação do risco financeiro <i>Financial risk assessment</i>
		Indicador económico <i>Economic indicator</i>
		Indicadores financeiros <i>Financial indicators</i>
		Indicadores de preços <i>Price indicators</i>
		Previsão de inflação <i>Inflation forecast</i>
		Enviesamento de género nos investimentos <i>Gender investment bias</i>
		Indicadores de investimento <i>Investment indicators</i>
		Desempenho do investimento <i>Investment performance</i>
		Recomendação de investimento <i>Investment recommendation</i>
		Sentimento de investimento <i>Investment sentiment</i>
Estratégia de investimento		

Nível 1: Função empresarial	Nível 2: Subfunção	Nível 3: Temática	
		<i>Investment strategy</i>	
		Sentimento macroeconómico <i>Macro economic sentiment</i>	
		Previsão do desempenho da bolsa de valores <i>Stock market performance forecast</i>	
		Retornos de bolsas de valores <i>Stock market returns</i>	
		Sistema de negociação <i>Trading system</i>	
Função de Marketing	Publicidade <i>Advertising</i>	Influência de celebridades <i>Celebrities influence</i>	
		Deteção de fraudes <i>Fraud detection</i>	
		Gestão da estratégia publicitária <i>Advertising strategy management</i>	
		Gestão de publicidade viral <i>Viral advertising management</i>	
		Indicadores de marca <i>Brand indicators</i>	
	Marca <i>Brand</i>	Gestão de marcas <i>Brand management</i>	
		Recomendação de marcas <i>Brand recommendation</i>	
		Comportamento do cliente <i>Customer behavior</i>	
	Cliente <i>Customer</i>	Previsão da rotatividade do cliente <i>Customer churn prediction</i>	
		Tomada de decisão do cliente <i>Customer decision-making</i>	
		Envolvimento do cliente <i>Customer engagement</i>	
		Análise do género do cliente <i>Customer gender analysis</i>	
		Fidelidade do cliente <i>Customer loyalty</i>	
		Opinião do cliente <i>Customer opinion</i>	
		Avaliações de clientes <i>Customer ratings</i>	
		Gestão da relação com o cliente <i>Customer relationship management</i>	
		Satisfação do cliente <i>Customer satisfaction</i>	
		Serviço ao cliente <i>Customer service</i>	
		Confiança nos revisores <i>Trust in reviewers</i>	
		Confiança nas revisões <i>Trust in reviews</i>	
		Confiança entre utilizadores nas redes sociais <i>Trust between users in social media</i>	
		Mercado <i>Market</i>	Sentimento de mercado <i>Market sentiment</i>
			Tendências de mercado <i>Market trends</i>
	Inteligência de marketing <i>Marketing intelligence</i>		
	Estratégia de marketing <i>Marketing strategy</i>		
	Análise quantitativa do mercado <i>Quantitative market analysis</i>		
	Produto (serviço) <i>Product (service)</i>	Deteção de defeitos <i>Defect detection</i>	
		Previsão do preço dos produtos <i>Price of products forecast</i>	
		Comparação de produtos <i>Product comparison</i>	
		Oportunidades de desenvolvimento de produtos <i>Product development opportunities</i>	
Análise das características do produto <i>Product features analysis</i>			
Oportunidades de melhoria dos produtos <i>Product improvement opportunities</i>			

Nível 1: Função empresarial	Nível 2: Subfunção	Nível 3: Temática		
		Posicionamento do produto <i>Product positioning</i>		
		Recomendação de produtos <i>Product recommendation</i>		
		Classificação de produtos <i>Product ranking</i>		
		Avaliação da qualidade do produto <i>Product quality assessment</i>		
	Vendas <i>Sales</i>	Previsão de exportação <i>Exports forecast</i>		
		Previsão de vendas <i>Sales forecast</i>		
		Análise do desempenho das vendas <i>Sales performance analysis</i>		
		Gestão de páginas web <i>Webpage management</i>		
		Função de produção <i>Production</i>	Logística <i>Logistics (operations)</i>	Gestão de operações <i>Operations management</i>
				Desempenho das operações <i>Operations performance</i>
Gestão de inventário <i>Inventory management</i>				
Sistemas de produção sem desperdício <i>Lean production systems</i>				
Qualidade do serviço logístico <i>Logistics service quality</i>				
Gestão de manutenção <i>Maintenance management</i>				
Previsão da produção <i>Production forecast</i>				
Análise do processo de produção <i>Production process analysis</i>				
Logística inversa <i>Reverse logistics</i>				
Gestão da cadeia de fornecimento <i>Supply chain management</i>				
Função de recursos humanos <i>Human Resources</i>	Questões laborais <i>Labor issues</i>	Requisitos emocionais do trabalho <i>Emotional labor requirements</i>		
		Desempenho dos empregados <i>Employees performance</i>		
		Satisfação dos empregados <i>Employees satisfaction</i>		
		Fatores de satisfação profissional <i>Job satisfaction factors</i>		
		Trabalho à distância <i>Remote work</i>		
		Desempenho do pessoal <i>Staff performance</i>		
		Condições de trabalho <i>Working conditions</i>		

Fonte: Elaboração própria

4.4. Origem dos dados por função empresarial

Inicialmente, este trabalho continha apenas as três primeiras questões de pesquisa, no entanto, ao longo do desenvolvimento do mesmo, a análise de *clusters* revelou a existência de um aglomerado (terceiro *cluster*) que sublinha a importância das fontes de recolha de dados para a realização da SA. Deste modo, a relevância da identificação deste *cluster* foi tão significativa que levou à expansão dos contributos inicialmente pensados para esta dissertação. Consequentemente, um dos objetivos para este trabalho passou, também, a ser a identificação da origem dos dados envolvidos nos artigos identificados. Os pontos subsequentes analisam, então as fontes de dados para as diferentes funções de gestão.

Foram determinados quatro grupos de origem de dados: informação interna, redes sociais/informação pública, documentação académica e documentação/informação financeira. A Tabela 7 descreve resumidamente os resultados obtidos, que serão analisados nos próximos subpontos.

Tabela 7. Repartição dos grupos de fontes de informação pelas várias funções empresariais.

Função Administrativa		Função Financeira		Função de Recursos Humanos	
Grupo	%	Grupo	%	Grupo	%
Documentação académica	2,04	Documentação académica	2,99	Documentação académica	0,00
Documentação/informação financeira	0,00	Documentação/informação financeira	44,78	Documentação/informação financeira	0,00
Informação interna	36,73	Informação interna	7,46	Informação interna	81,82
Redes sociais/informação pública	61,22	Redes sociais/informação pública	44,78	Redes sociais/informação pública	18,18
	Σ 100		Σ 100		Σ 100
Função de Marketing		Função de Produção			
Grupo	%	Grupo	%		
Documentação académica	1,41	Documentação académica	12,00		
Documentação/informação financeira	0,60	Documentação/informação financeira	0,00		
Informação interna	68,47	Informação interna	56,00		
Redes sociais/informação pública	29,52	Redes sociais/informação pública	32,00		
	Σ 100		Σ 100		

Fonte: Elaboração própria

4.4.1 Função Administrativa

Para a função administrativa, os dados são maioritariamente recolhidos através de informação disponibilizada publicamente. Olhando para o gráfico da distribuição das temáticas na função administrativa (Figura 6), a gestão de crises é a temática mais explorada e por sinal, é resultante de situações inesperadas que não dependem das empresas (*e.g.*, crise económica e Covid-19). Para acompanhar e monitorizar a evolução de crises vigentes, a utilização de informação pública é claramente apropriada. Adicionalmente, questões relacionadas com a responsabilidade social das corporativa, que é a segunda temática mais explorada, estão, também, intimamente ligadas à fatores externos. As empresas servem-se de informação pública/redes sociais para perceberem quais são as questões sociais que carecem de maior atenção, por exemplo.

Por outro lado, uma parte do processo de tratamento da gestão de crise necessita de avaliar dados internos, para avaliar o impacto da crise dentro da empresa. Para tal, os dados que fazem referência a situação inesperada na empresa, refletem-se nos relatórios de desempenho empresarial, bem como reclamações de clientes que não de ficar nos registos internos.

Portanto, pelas razões enunciadas, é seguro dizer que o método de recolha mais utilizado na função administrativa é por meio de informação pública/redes sociais, seguido pelo uso de documentação interna (ver Tabela 8).

Tabela 8. Origem de dados da função administrativa

Grupo	Qtd.	Fonte de dados	Qtd.
Documentação acadêmica	1	<i>Financial literature</i>	1
Informação interna	18	<i>Customer review dataset</i>	7
		<i>Administration letters to shareholders</i>	2
		<i>Financial report</i>	2
		<i>Negotiation support system</i>	1
		<i>Customer calls</i>	1
		<i>Employee messaging</i>	1
		<i>Employee review dataset</i>	1
		<i>Interviews</i>	1
		<i>Prototype system</i>	1
		<i>Assembly process plans</i>	1
Redes sociais/informação pública	30	<i>Twitter</i>	13
		<i>Facebook</i>	5
		<i>Public news</i>	4
		<i>Social media dataset (generic)</i>	3
		<i>Online news</i>	2
		<i>Online public firm information</i>	1
		<i>Sina Weibo</i>	1
		<i>Google searches</i>	1
Σ 49		Σ 49	

Fonte: Elaboração própria

4.4.2 Função Financeira

A função financeira (Tabela 9) é a segunda função cujos estudos mais recorrem à aplicação da SA. Para melhor tomar uma decisão financeira, convém que os dados sejam oficiais e de fontes financeiras fidedignas. Este tipo de dados permite, por exemplo, perceber a situação financeira de uma empresa e agir em conformidade. Assim, metade dos estudos usou SA sobre documentação financeira. Estes dados, por norma, já contêm dados estruturados e são mais fáceis de serem utilizados imediatamente. No entanto, a outra metade dos estudos usa dados provenientes de redes sociais/informação pública.

Tabela 9. Origem de dados da função financeira

Grupo	Qtd.	Fonte de dados	Qtd.
Documentação acadêmica	6	<i>Financial literature</i>	3
Documentação/informação financeira	90	<i>Research literature</i>	3
		<i>Accounting report</i>	3
		<i>Central bank communication</i>	1
		<i>Conference call transcriptions</i>	1
		<i>Financial dataset</i>	32
		<i>Financial news</i>	32
		<i>Financial report</i>	20
		<i>Macroeconomic indicators</i>	1
Informação interna	15	<i>Comment letters</i>	1
		<i>Customer review dataset</i>	14
Redes sociais/informação pública	90	<i>Facebook</i>	1
		<i>Financial blog</i>	3
		<i>Google searches</i>	1
		<i>Google Trends</i>	1
		<i>Online news</i>	3
		<i>Public news</i>	29
		<i>Sina Weibo</i>	4
		<i>Social media dataset (generic)</i>	6
<i>Twitter</i>	42		
Σ 201		Σ 201	

Fonte: Elaboração própria

As redes sociais permitem ter informação mais atualizada, mas o processo de extração da mesma é mais complexo, porque os dados são geralmente não estruturados. O processo de organização, quer por linha temporal ou por outro critério, leva mais tempo. Contudo, a importância da informação disponibilizada *online* atualmente é um dos meios que mais impacta o mercado financeiro devido à própria natureza volátil mesmo. A utilização da SA sobre outras fontes de dados apresenta-se meramente residual.

4.4.3 Função de *Marketing*

Segundo os estudos analisados, no *marketing* (Tabela 10) constata-se que as informações internas são o principal meio de recolha de dados para aplicação da SA, uma vez que representam mais de metade das fontes de informação.

Nos estudos analisados, a maior parte dos autores aparente ter utilizado um conjunto de dados (*datasets*) já processados pelas empresas, porque são mais fáceis de serem obtidos. Contudo, o segundo tipo de fonte de dados preferido é definitivamente através do uso das redes sociais. As duas modalidades de recolha têm as suas vantagens e desvantagens.

Como foi referido, os *datasets* já processados e disponíveis em bases de dados são de mais fácil acesso, mas podem conter informações menos recentes. As redes sociais permitem obter informação mais atualizada ou mesmo em tempo real, em contrapartida, o processo para a sua recolha é mais lento e mais difícil, carecendo de meios técnicos e conhecimento especializado (por exemplo ao nível da programação de aplicações informáticas), nem sempre dominado pelos investigadores. Contudo, a escolha do método de recolha poderá variar de acordo com a natureza do próprio estudo.

Tabela 10. Origem de dados da função de *marketing*

Grupo	Qtd.	Fonte de dados	Qtd.
Documentação académica	7	<i>Research literature</i>	2
		<i>Research questionnaire</i>	5
Documentação/informação financeira	3	<i>Financial news</i>	3
Informação interna	341	<i>Customer calls</i>	2
		<i>Customer review dataset</i>	336
		<i>Employee review dataset</i>	1
		<i>Interviews</i>	2
Redes sociais/informação pública	147	<i>Blogs/forums</i>	5
		<i>Facebook</i>	18
		<i>Flickr</i>	2
		<i>Google searches</i>	2
		<i>Instagram</i>	8
		<i>Mobile advertising service platform</i>	1
		<i>Online news</i>	6
		<i>Online public firm information</i>	1
		<i>Online questionnaire</i>	1
		<i>Public news</i>	4
		<i>Sina Weibo</i>	2
		<i>Social media dataset (generic)</i>	8
		<i>Twitter</i>	84
<i>YouTube</i>	5		
	Σ 498		Σ 498

Fonte: Elaboração própria

4.4.4 Função de Produção

Quanto à função de produção (Tabela 11), os dados utilizados nos estudos são, na sua maioria, provenientes de informações internas. Faz sentido que assim seja, porque a função de produção está relacionada maioritariamente com atividades que envolvem mais processos internos como a gestão de *stocks* e inventário, por exemplo. Para tal, a melhor fonte de informação são os dados já compilados e processados internamente.

Não obstante, a informação disponível publicamente (redes sociais ou outro fórum), também tem um papel essencial na recolha de dados para alguns aspetos desta função. Um exemplo seria a comunicação com parceiros (*e.g.*, fornecedores e distribuidores). As empresas podem utilizar estes meios para investigarem, através da SA, a reputação dos seus potenciais parceiros antes de se aliarem.

Tabela 11. Origem de dados da função de produção

Grupo	Qtd.	Fonte de dados	Qtd.
Documentação académica	3	<i>Research literature</i>	3
Informação interna	14	<i>Customer review dataset</i>	7
		<i>Interviews</i>	4
		<i>Assembly process plans</i>	1
		<i>Employee messaging</i>	1
		<i>Internal communications</i>	1
Redes sociais/informação pública	8	<i>Twitter</i>	6
		<i>Public news</i>	1
		<i>YouTube</i>	1
Σ 25		Σ 25	

Fonte: Elaboração própria

Em suma, sendo a produção uma função cujas operações sejam de cariz interno, faz sentido que os autores decidam usar dados produzidos pelas próprias empresas.

4.4.5 Função de Recursos Humanos

Assim como a função de produção, a de recursos humanos (Tabela 12), privilegia também os dados extraídos a partir de informações internas, para aplicação da SA.

A Figura 10, no ponto 4.2.5, já indicava que os principais aspetos analisados nos recursos humanos são as condições de trabalho, o trabalho remoto e a satisfação do trabalhador. O que se evidenciou nos estudos, é que maior parte da aquisição destes dados é feita a partir de questionários/inquéritos e avaliação aos trabalhadores das empresas. Logo, se os dados são gerados pelas próprias empresas, a forma mais fácil de conseguir acesso é consultando a documentação/informação interna.

Tabela 12. Origem de dados da função de recursos humanos

Grupo	Qtd.	Fonte de dados	Qtd.
Informação interna	9	<i>Customer review dataset</i>	2
		<i>Employee review dataset</i>	2
		<i>Employer review dataset</i>	2
		<i>Customer calls</i>	1
		<i>Email repository</i>	1
		<i>Employee questionnaire</i>	1
Redes sociais/informação pública	2	<i>Twitter</i>	2
Σ 11		Σ 11	

Fonte: Elaboração própria

Além do mais, o trabalho remoto ganhou grande destaque devido à pandemia. Para quem trabalhasse com o atendimento ao cliente, a sua avaliação de desempenho era feita através da auscultação das chamadas com clientes, e esta informação ficava guardada nos registos internos da empresa. Ou seja, para ter acesso a estes dados, é quase obrigatório ter de consultar documentação interna.

Entretanto, há lugares na internet onde colaboradores de várias empresas expressam opiniões (positivas, negativas ou neutras) de forma anónima ou não. Mesmo assim, a utilização das redes sociais como fontes de dados para esta função é residual.

Capítulo 5

Conclusões, limitações e trabalhos futuros

Este estudo incidiu sobre a aplicação e a evolução dos estudos sobre SA na gestão de empresas, uma vez que era uma área de investigação pouco explorada, mas que tem vindo a ganhar destaque nos últimos anos. Para mensurar esta evolução, foi feita uma análise da literatura existente até ao ano corrente (2022), por meio da combinação de uma revisão sistemática, de uma análise bibliométrica e de uma proposta de taxonomia. A análise bibliométrica visava identificar os autores mais proeminentes neste campo, os periódicos/conferências que mais contribuem para a difusão do tema, a frequência dos termos mais populares e crescimento da relevância destes mesmos termos ao longo dos anos.

Os resultados demonstram que a SA na gestão enfoca nos seguintes termos: *sentiment, review, social medium, product, consumer, impact, customer, performance e service*. Estes são os 10 mais frequentes em toda a análise de agrupamento por funções empresariais. Isto sugere-nos que a SA na gestão tem como foco o sentimento do emissor da opinião, os comentários do mesmo sobre uma entidade (*e.g.*, produto, serviço pessoal), o meio utilizado para expressar a opinião, a medição da repercussão da opinião, a satisfação do cliente e o desempenho empresarial.

Os resultados da análise bibliométrica expressam claramente uma tendência significativa para a transição digital. Dos três *clusters*, dois correspondem a funções empresariais e o outro está relacionado com as principais fontes de informação.

As duas funções reconhecidas foram, designadamente, a função de *marketing* e a função financeira. Das duas, a mais corrente é a de *marketing*. Dez anos depois, os resultados deste estudo ainda seguem a tendência apresentada por Feldman (2013) que já indicava que a SA era mais utilizada no *marketing*, na temática do comportamento do consumidor, seguida da função financeira. De facto, constatou-se mesmo que a SA é utilizada maioritariamente no *marketing* para avaliar o sentimento do consumidor em relação aos aspetos anteriormente referidos, acrescido de sugestões para melhorias de algum aspeto.

No caso da função financeira, a aplicação da SA é estudada na tentativa de ajudar a tomar melhores decisões de investimentos mediante a previsão de preços da bolsa de valores e dos valores de retorno da mesma, com base em documentação financeira, e notícias públicas de carácter financeiro (principalmente as publicadas na internet).

O terceiro *cluster* realça o papel das tecnologias de informação e comunicação na colheita de dados, sublinhando a importância das redes sociais.

É importante frisar que o resultado da análise de *clusters* apenas indica as temáticas mais populares em termos de utilização da SA na gestão. Durante a seleção dos artigos, reparou-se que existia, embora de forma residual, a aplicação da SA noutras funções da gestão como por exemplo na administrativa, na produção e nos recursos humanos. Por aparecerem em quantidades muito

reduzidas, o VOSviewer considerou-as insuficientes para as agrupar em *clusters* independentes. A confirmação destes resultados foi reafirmada pela análise detalhada dos artigos e que foi a base para a proposta da taxonomia.

Esta proposta esquematiza todos os domínios (subfunções e temáticas) de estudo da aplicação da SA na gestão, independentemente de serem residuais ou não. A classificação taxonómica revelou resultados semelhantes à análise bibliométrica, reafirmando o potencial de aplicação na função de *marketing* e na função financeira, representado pela diversidade de temáticas e subfunções comparativamente às outras funções. Na gestão administrativa é mais utilizada para questões relacionadas com a gestão de crises, responsabilidade social e reputação da empresa. Na função financeira é maioritariamente aplicada na previsão de mercados financeiros e decisões de investimentos.

Enquanto isto, no *marketing* é predominantemente usada para avaliar questões relacionadas com a satisfação do cliente/consumidor, marca e produtos. Já na produção, o foco na logística (cadeia de abastecimento e gestão de *stocks*). Por último, nos recursos humanos serve para avaliar questões de trabalho, onde o maior foco está relacionado com as condições de serviço, a satisfação do trabalhador e a performance do mesmo.

Tal como indicado pela análise, as fontes de informação são de grande importância para que se consiga implementar a SA, porque tudo começa pela informação. É importante ter dados para que se consiga tomar uma decisão minimamente informada sobre qualquer assunto. Desta forma, a reposta à 4.^a questão de pesquisa passou pelo levantamento da origem de dados em cada um dos artigos incluídos neste estudo, que resultou na agregação das fontes de dados em quatro grandes grupos: informação interna, redes sociais/informação pública, documentação académica e documentação/informação financeira.

Na função administrativa, a fonte preferencial é a informação interna. Na função financeira, a documentação/informação financeira apresenta igual importância à obtida através das redes sociais/informação pública. No *marketing* e na produção, a maior parte dos dados são provenientes de informação interna, seguido por dados obtidos nas redes sociais/informação pública. Já nos recursos humanos, quase toda a informação origina-se quase exclusivamente a partir de informação interna.

Findo o trabalho, conclui-se, pela classificação taxonómica, que a SA na gestão de empresas tem potencial para mais aplicações além das temáticas mencionadas, porque os resultados sugerem que se tem investigado mais sobre o assunto. O surgimento da COVID 19 em 2019/2020 veio demonstrar e realçar a importância da tecnologia através da exigência do trabalho à distância.

Embora a utilização de dados provenientes de uma única base de dados, a WOS, possa ser apontada como uma limitação deste trabalho há que apontar que na mesma altura que se recolheram os artigos na WOS, da forma descrita no capítulo 4, foram recolhidos também de modo semelhante os artigos na SCOPUS (pesquisa do termo “*Sentiment Analysis*”, restringidas às áreas de gestão, nomeadamente *Business, Management and Accounting, Economics e Econometrics and Finance* e para o mesmo período temporal).

Dessa pesquisa foram recolhidos 1.996 artigos. Após interseção, através dos títulos dos artigos, com os 1.151 artigos da WOS, concluiu-se que 818 artigos (40,98%) eram comuns a ambas as bases de dados e que desses 818 artigos tinham sido selecionados 474 para a elaboração deste estudo. Recordando que no total foram usados 524 usados, 90,65% desses artigos são comuns a ambas base de dados.

Embora não pareça expectável que haja grandes variações aos resultados aqui apresentados, seria interessante, em trabalho futuro, terminar a análise dos 1.178 artigos que aparecem na SCOPUS sem correspondência na WOS, mas que por restrições de tempo para a elaboração desta dissertação não puderam ser incorporados na análise.

Referências Bibliográficas

- Addan, D. (2019). *Support Vector Machines*. <https://www.inf.ufpr.br/dagoncalves/IA07.pdf>
- Ahuja, R., Chug, A., Kohli, S., Gupta, S. & Ahuja, P. (2019). The impact of features extraction on the sentiment analysis. *Procedia Computer Science*, 152, 341–348. <https://doi.org/10.1016/j.procs.2019.05.008>
- Al-Twairesh, N., Al-Khalifa, H. & Al-Salman, A. (2016). AraSenTi: Large-scale twitter-specific Arabic sentiment lexicons. *54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 697–705. <https://doi.org/10.18653/v1/P16-1066>
- Almeida, F. (2016). *Introdução à Gestão de Organizações* (4ª Edição). Escolar Editora.
- Alqaryouti, O. H. (2017). Aspect-Based Sentiment Analysis for Government Smart Applications Customers' Reviews [The British University in Dubai]. In *The British University in Dubai*. <https://bspace.buid.ac.ae/handle/1234/1056>
- Anuratha, K., Parvathy, M., Sujeetha, S. & Ghayathri, J. (2019). Role of social sentiment analysis in stock trends forecasting. *International Journal of Recent Technology and Engineering*, 8(2S5), 144–147. <https://doi.org/10.35940/ijrte.B1030.0782S519>
- Archak, N., Ghose, A. & Ipeirotis, P. G. (2011). Deriving the pricing power of product features by mining consumer reviews. *Management Science*, 57(8), 1485–1509. <https://doi.org/10.1287/mnsc.1110.1370>
- Armstrong, M. & Taylor, S. (2014). Armstrong's handbook of human resource management practice. In *Kogan Page* (13th Ed., p. 880).
- Barreto, J. M. (2017). *Introdução à Administração*. UFBA, Faculdade de Ciências Contábeis, Superintendência de Educação a Distância. https://repositorio.ufba.br/bitstream/ri/24417/1/eBook_Introducao_a_Administra%c3%a7%c3%a3o-Ciencias_Contabeis_UFBA.pdf
- Bellani, E., Ramadhani, S. R. & Tamar, M. (2017). Job Satisfaction as Predictor of employee engagement. *8th International Conference of Asian Association of Indigenous and Cultural Psychology (ICAAIP 2017)*, 15–19. <https://doi.org/10.2991/icaaip-17.2018.4>
- Benamara, F., Taboada, M. & Mathieu, Y. (2017). Evaluative Language Beyond Bags of Words: Linguistic Insights and Computational Applications. *Computational Linguistics*, 43(1), 201–264. https://doi.org/10.1162/COLI_a_00278
- Biswas, S., Ghosh, A., Chakraborty, S., Roy, S. & Bose, R. (2020). Scope of Sentiment Analysis on News Articles Regarding Stock Market and GDP in Struggling Economic Condition. *International Journal of Emerging Trends in Engineering Research*, 8(7), 3594–3609. <https://doi.org/10.30534/ijeter/2020/117872020>
- Bradley, M. M. & Lang, P. J. (1999). Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings. In *Technical Report C-1, The Center for Research in Psychophysiology*.
- Cambria, E., Speer, R., Havasi, C. & Hussain, A. (2010). SenticNet: A publicly available semantic resource for opinion mining. *Commonsense Knowledge: AAAI Fall Symposium*, 14–18.
- Carosia, A. E., Coelho, G. P. & Silva, A. E. (2021). Investment strategies applied to the Brazilian stock market: A methodology based on Sentiment Analysis with deep learning. *Expert Systems with Applications*, 184, 1–12. <https://doi.org/10.1016/j.eswa.2021.115470>

- Chatterjee, S. & Kar, A. K. (2020). Why do small and medium enterprises use social media marketing and what is the impact: Empirical insights from India. *International Journal of Information Management*, 53, 1–13. <https://doi.org/10.1016/j.ijinfomgt.2020.102103>
- Chiavenato, I. (2004). *Introdução à Teoria Geral da Administração* (7ª Edição). Elsevier.
- Chung, S., Shin, D. & Park, J. (2022). Predicting Firm Market Performance Using the Social Media Promoter Score. *Marketing Letters*, 33, 545–561. <https://doi.org/10.1007/s11002-022-09615-w>
- Chung, W. & Tseng, T. L. (2012). Discovering business intelligence from online product reviews: A rule-induction framework. *Expert Systems with Applications*, 39(15), 11870–11879. <http://dx.doi.org/10.1016/j.eswa.2012.02.059>
- Cook, D. J., Mulrow, C. D. & Haynes, R. B. (1997). Systematic Reviews: Synthesis of Best Evidence for Clinical Decisions. *Annals of Internal Medicine*, 126, 376–380. <https://doi.org/10.7326/0003-4819-126-5-199703010-00006>
- Crossan, M. M. & Apaydin, M. (2010). A multi-dimensional framework of organizational innovation: A systematic review of the literature. *Journal of Management Studies*, 47(6), 1154–1191. <https://doi.org/10.1111/j.1467-6486.2009.00880.x>
- D’Andrea, A., Ferri, F., Grifoni, P. & Guzzo, T. (2015). Approaches, Tools and Applications for Sentiment Analysis Implementation. *International Journal of Computer Applications*, 125(3), 26–33. <https://doi.org/10.5120/ijca2015905866>
- Dave, K., Lawrence, S., Pennock, D. M. (2003). Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews. *Twelfth International Conference on World Wide Web - WWW ’03*, 519–528. <https://doi.org/https://doi.org/10.1145/775152.775226>
- De Long, J. B., Shleifer, A., Summers, L. H. & Waldmann, R. J. (1990). Positive feedback investment strategies and destabilizing rational speculation. *The Journal of Science*, 45(2), 379–395. <https://doi.org/10.1111/j.1540-6261.1990.tb03695.x>
- Denecke, K. (2008). Using SentiWordNet for multilingual sentiment analysis. *International Conference on Data Engineering*, 507–512. <https://doi.org/10.1109/ICDEW.2008.4498370>
- Drucker, P. F. (1986). *Management: Tasks, Responsibilities, Practices*. Truman Talley Books.
- Duarte, A.P.A. & Barbosa, A. F. O. (2019). Teorias da Administração. In *Educação a Distância (EAD)* (3ª Edição).
- Farooq, U., Mansoor, H., Nongaillard, A., Ouzrout, Y. & Qadir, M. A. (2017). Negation Handling in Sentiment Analysis at Sentence Level. *Journal of Computers*, 12(5), 470–478. <https://doi.org/10.17706/jcp.12.5.470-478>
- Feldman, R., Fresko, M., Goldenberg, J., Netzer, O. & Ungar, L. (2008). Using text mining to analyze user forums. *2008 International Conference on Service Systems and Service Management*, 1–5. <https://doi.org/10.1109/ICSSSM.2008.4598504>
- Feldman, R. (2013). Techniques and applications for sentiment analysis: The main applications and challenges of one of the hottest research areas in computer science. *Communications of the ACM*, 56(4), 82–89. <https://doi.org/10.1145/2436256.2436274>
- Gupta, N. & Agrawal, R. (2020). Chapter 1- Application and Techniques of Opinion Mining. *Hybrid Computational Intelligence*, 1–23. <https://doi.org/10.1016/b978-0-12-818699-2.00001-9>

- Hamouche, S. (2021). Human resource management and the COVID-19 crisis: Implications, challenges, opportunities, and future organizational directions. *Journal of Management and Organization*, 1–16. <https://doi.org/10.1017/jmo.2021.15>
- Hassan, A., Abu-Jbara, A., Jha, R. & Radev, D. (2011). Identifying the semantic orientation of foreign words. *49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, 2, 592–597.
- Hassan, N. R. & Loebbecke, C. (2017). Engaging scientometrics in information systems. *Journal of Information Technology*, 32(1), 85–109. <https://doi.org/10.1057/jit.2015.29>
- Hennig-Thurau, T., Gwinner, K. P., Walsh, G. & Gremler, D. D. (2004). Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet? *Journal of Interactive Marketing*, 18(1), 38–52. <https://doi.org/10.1002/dir.10073>
- Hess, T., Benlian, A., Matt, C. & Wiesböck, F. (2016). Options for Formulating a Digital Transformation Strategy. *MIS Quarterly Executive*, 15(2), 103–119.
- Hogenboom, A., Bal, D., Frasinca, F., Bal, M., De Jong, F. & Kaymak, U. (2013). Exploiting emoticons in sentiment analysis. *28th Annual ACM Symposium on Applied Computing*, 703–710. <https://doi.org/10.1145/2480362.2480498>
- Hogenboom, A., Boon, F. & Frasinca, F. (2012). A statistical approach to star rating classification of sentiment. In *Advances in Intelligent Systems and Computing* (pp. 251–260). https://doi.org/10.1007/978-3-642-30864-2_24
- Hussein, D. M. E.-D. M. (2016). Analyzing Scientific Papers Based on Sentiment Analysis [Cairo University]. In *Cairo University*. https://www.researchgate.net/profile/Doaa-Mohey-El-Din/publication/313113567_Master_Thesis_of_sentiment_Analysis_Last_Edition/links/58909e56aca272f9a556ba03/Master-Thesis-of-sentiment-Analysis-Last-Edition.pdf
- Hutto, C.J. & Gilbert, E. (2014). VADER: A Parsimonious Rule-based Model for. *Eighth International AAAI Conference on Weblogs and Social Media*, 18. <https://doi.org/10.1609/icwsm.v8i1.14550>
- Ibekwe, F., Bochi, F. & Martínez-Ávila, D. (2021). Mapping the evolution of topics published by Education for Information: Interdisciplinary Journal of Information Studies. *Education for Information*, 37(4), 545–563. <https://doi.org/10.3233/EFI-211559>
- Jordan, M. I. & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255–260. <https://doi.org/10.1126/science.aaa8415>
- Kaity, M. & Balakrishnan, V. (2020). Sentiment lexicons and non-English languages: a survey. *Knowledge and Information Systems*, 62, 4445–4480. <https://doi.org/10.1007/s10115-020-01497-6>
- Kaur, F. & Bhatia, R. (2016). Sentiment Analyzing by Dictionary based Approach. *International Journal of Computer Applications*, 152(5), 32–34. <https://doi.org/10.5120/ijca2016911814>
- Kaur, H. , Mangat, V. & N. (2017). A survey of sentiment analysis techniques. *International Conference on IoT in Social, Mobile, Analytics and Cloud (I-SMAC 2017)*, 921–925. <https://doi.org/10.1109/I-SMAC.2017.8058315>
- Kim, J., Li, J.-J. & Lee, J.-H. (2010). Evaluating multilanguage-comparability of subjectivity analysis systems. *48th Annual Meeting of the Association for Computational Linguistics*, 595–603.
- Kim, S.-M. & Hovy, E. (2006). Identifying and analyzing judgment opinions. *HLT-NAACL 2006*

- *Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings of the Main Conference, 200–207.*
<https://doi.org/10.3115/1220835.1220861>
- King, R. A., Racherla, P. & Bush, V. D. (2014). What we know and don't know about online word-of-mouth: A review and synthesis of the literature. *Journal of Interactive Marketing, 28*(3), 167–183. <https://doi.org/10.1016/j.intmar.2014.02.001>
- Koukaras, P., Tjortjis, C. & Rousidis, D. (2020). Social Media Types: introducing a data driven taxonomy. *Computing, 102*, 295–340. <https://doi.org/10.1007/s00607-019-00739-y>
- Kumar, S., Kar, A. K. & Ilavarasan, P. V. (2021). Applications of text mining in services management: A systematic literature review. *International Journal of Information Management Data Insights, 1*(1), 1–14. <https://doi.org/10.1016/j.ijime.2021.100008>
- Lafferty, J. D., McCallum, A. & Pereira, F. C. N. (2001). Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. *18th International Conference on Machine Learning 2001 (ICML 2001)*, 282–289.
<https://dl.acm.org/doi/10.5555/645530.655813>
- Li, L., Zhang, Q., Tian, J. & Wang, H. (2018). Characterizing information propagation patterns in emergencies: A case study with Yiliang Earthquake. *International Journal of Information Management, 38*(1), 34–41.
<https://doi.org/http://dx.doi.org/10.1016/j.ijinfomgt.2017.08.008>
- Lisboa, J., Coelho, A., Coelho, F. & Almeida, F. (2011). *Introdução à Gestão de Organizações* (3ª Edição). Grupo Editorial Vida Económica.
- Liu, Y., Jiang, C. & Zhao, H. (2019). Assessing product competitive advantages from the perspective of customers by mining user-generated content on social media. *Decision Support Systems, 123*, 1–13. <https://doi.org/10.1016/j.dss.2019.113079>
- Liu, B. (2012). Sentiment Analysis and Opinion Mining. In *Synthesis Lectures on Human Language Technologies* (1st Edition). Morgan & Claypool Publishers.
- Liu, B. (2015). Sentiment Analysis: Mining Opinions, Sentiments, and Emotions. In *Studies in Natural Language Processing* (1st Editio). Cambridge University Press.
- Lo, S. L., Cambria, E., Chiong, R. & Cornforth, D. (2017). Multilingual sentiment analysis: from formal to informal and scarce resource languages. *Artificial Intelligence Review, 48*(4), 499–527. <https://doi.org/10.1007/s10462-016-9508-4>
- Lowry, P., Romans, D. & Curtis, A. (2004). Global Journal Prestige and Supporting Disciplines: A Scientometric Study of Information Systems Journals. *Journal of the Association for Information Systems, 5*(2), 29–77. <https://doi.org/10.17705/1jais.00045>
- Lowry, P. B., Moody, G. D., Gaskin, J., Galletta, D. F., Humpherys, S. L., Barlow, J. B. & Wilson, D. W. (2013). Evaluating journal quality and the association for information systems senior scholars' journal basket via bibliometric measures: Do expert journal assessments add value? *MIS Quarterly, 37*(4), 993–1012.
<https://doi.org/10.25300/MISQ/2013/37.4.01>
- Luo, J. (2020). Bitcoin price prediction in the time of COVID-19. *2020 Management Science Informatization and Economic Innovation Development Conference (MSIEID)*, 243–247.
<https://doi.org/10.1109/MSIEID52046.2020.00050>
- Mahesh, B. (2020). Machine Learning Algorithms - A Review. *International Journal of Science and Research (IJSR)*, 9(1), 381–386. <https://doi.org/10.21275/ART20203995>
- Mantyla, M. V., Graziotin, D. & Kuutilla, M. (2018). The evolution of sentiment analysis—A

- review of research topics, venues, and top cited papers. *Computer Science Review*, 27, 16–32. <https://doi.org/10.1016/j.cosrev.2017.10.002>
- Mehta, P. & Pandya, S. (2020). A Review on Sentiment Analysis Methodologies, Practices and Applications. *International Journal of Scientific and Technology Research*, 9(2), 601–609. <http://www.ijstr.org/final-print/feb2020/A-Review-On-Sentiment-Analysis-Methodologies-Practices-And-Applications.pdf>
- Micu, A., Micu, A. E., Geru, M. & Lixandroi, R. C. (2017). Analyzing user sentiment in social media: Implications for online marketing strategy. *Psychology and Marketing*, 34(12), 1094–1100. <https://doi.org/10.1002/mar.21049>
- Mihalcea, R., Banea, C. & Wiebe, J. (2007). Learning multilingual subjective language via cross-lingual projections. *45th Annual Meeting of the Association for Computational Linguistics*, 976–983. <https://aclanthology.org/P07-1123>
- Mohammad, S. M. & Turney, P. (2010). Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon. *NAACL-HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, 26–34. <https://aclanthology.org/W10-0204>
- Mountrakis, G., Im, J. & Ogole, C. (2011). Support Vector Machines in Remote Sensing: A Review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3), 247–259. <https://doi.org/10.1016/j.isprsjprs.2010.11.001>
- Mukhopadhyay, S. (2018). Opinion mining in management research: the state of the art and the way forward. *Opsearch*, 55(2), 221–250. <https://doi.org/10.1007/s12597-017-0328-3>
- Murphy, K. P. (2012). Machine Learning: A probabilistic Approach. In *Adaptive Computation and Machine Learning series* (Illustrate). The MIT Press. http://noiselab.ucsd.edu/ECE228/Murphy_Machine_Learning.pdf
- Nasukawa, T. & Yi, J. (2003). Sentiment analysis: Capturing favorability using natural language processing. *2nd International Conference on Knowledge Capture, K-CAP 2003*, 70–77. <https://doi.org/https://doi.org/10.1145/945645.945658>
- Nickerson, R. C., Varshney, U. & Muntermann, J. (2013). A method for taxonomy development and its application in information systems. *European Journal of Information Systems*, 22(3), 336–359. <https://doi.org/https://doi.org/10.1057/ejis.2012.26>
- Organização Internacional do Trabalho. (2020). *Advancing Social Justice, Promoting Decent Work – Working Conditions*. <http://www.ilo.ch/global/topics/working-conditions/lang-en/index.htm>
- Pang, B., Lee, L. & Vaithyanathan, S. (2002). Thumbs up? Sentiment Classification using Machine Learning Techniques. *2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)*, 10, 79–86. <https://doi.org/10.3115/1118693.1118704>
- Pang, B. & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, 2(1–2), 1–135. <http://dx.doi.org/10.1561/1500000011>
- Peng, Y., Wang, G., Kou, G. & Shi, Y. (2011). An empirical study of classification algorithm evaluation for financial risk prediction. *Applied Soft Computing Journal*, 11(2), 2906–2915. <https://doi.org/10.1016/j.asoc.2010.11.028>
- Pereira, D. A. (2021). A survey of sentiment analysis in the Portuguese language. *Artificial Intelligence Review*, 54, 1087–1115. <https://doi.org/10.1007/s10462-020-09870-1>
- Perez-Rosas, V., Banea, C. & Mihalcea, R. (2012). Learning sentiment lexicons in Spanish.

- Eighth International Conference on Language Resources and Evaluation (LREC'12)*, 3077–3081.
- Pompian, M. M. (2017). Risk tolerance and behavioral finance. In *Investment Management Consultants Association Inc.*
<https://investmentsandwealth.org/getattachment/0a603493-fce1-4ab9-8d6a-3e78e1178f93/IWM17MayJun-RiskToleranceBehavioralFinance.pdf>
- Pritchard, A. (1969). Statistical Bibliography or Bibliometrics? *Journal of Documentation*, 25, 348–349.
- Rajput, Q., Haider, S. & Ghani, S. (2016). Lexicon-Based Sentiment Analysis of Teachers' Evaluation. *Applied Computational Intelligence and Soft Computing*, 1–12.
<https://doi.org/10.1155/2016/2385429>
- Rathore, A. K. & Ilavarasan, P. V. (2020). Pre- and post-launch emotions in new product development: Insights from twitter analytics of three products. *International Journal of Information Management*, 50, 111–127. <https://doi.org/10.1016/j.ijinfomgt.2019.05.015>
- Sadia, A., Khan, F. & Bashir, F. (2018). An overview of lexicon-based approach for sentiment analysis. *International Electrical Engineering Conference*, 1–6.
- Serras, F. A. (2015). *Métodos de Aprendizagem Automática Um estudo baseado na avaliação e previsão de clientes bancários* [Universidade Nova de Lisboa].
<https://run.unl.pt/bitstream/10362/17371/1/TGIO051.pdf>
- Silva, F., Silvano, P., Leal, A., Oliveira, F., Brazdil, P., Cordeiro, J. & Oliveira, D. (2018). Análise de Sentimento em Artigos de Opinião. *Revista de Estudos Linguísticos Da Universidade Do Porto*, 13, 80–114. <https://hdl.handle.net/10216/120635>
- Smith, S. F. (1992). Knowledge-based production management approaches, results and prospects. *Production Planning & Control*, 3(4), 350–380.
<https://doi.org/10.1080/09537289208919407>
- Stimpson, A. J. & Cummings, M. L. (2014). Assessing intervention timing in computer-based education using machine learning algorithms. *IEEE Access*, 2, 78–87.
<https://doi.org/10.1109/ACCESS.2014.2303071>
- Stone, P.J., Dunphy, D. C., Smith, M. S. & Ogilvie, D. M. (1966). *The General Inquirer: A Computer Approach to Content Analysis*. The MIT Press.
- Subbaraju, P., Chandra Sekhar, K., Raju, P. R.S.S.V., Satyanarayana Raju, K. & Varma, M. K. S. (2020). An Empirical Review of Challenges of and Approaches to Sentiment analysis. *International Journal of Industrial Engineering and Production Research*, 31(2), 317–322. <https://doi.org/10.22068/ijiepr.31.2.317>
- Sun, Y., Liu, X., Chen, G., Hao, Y. & Zhang, Z. (2020). How mood affects the stock market: Empirical evidence from microblogs. *Information & Management*, 57(5), 103181.
<https://doi.org/10.1016/j.im.2019.103181>
- Suryachandra, P. & Reddy, P. V. S. (2019). Methodologies in Sentiment Analysis. *International Journal of Scientific and Technology Research*, 8(9), 531–535.
<https://www.ijstr.org/paper-references.php?ref=IJSTR-0919-22030>
- T.K., B., Annavarapu, C. S. R. & Bablani, A. (2021). Machine learning algorithms for social media analysis: A survey. *Computer Science Review*, 40, 100395.
<https://doi.org/10.1016/j.cosrev.2021.100395>
- Taboada, M., Anthony, C. & Voll, K. (2006). Methods for creating semantic orientation dictionaries. *5th International Conference on Language Resources and Evaluation, LREC*

2006, 427–432.

- Taboada, M., Brooke, J., Tofiloski, M., Voll, K. & Stede, M. (2011). Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics*, 37(2), 267–307. https://doi.org/10.1162/COLI_a_00049
- Tanulia, Y. & Girsang, A. S. (2019). Sentiment analysis on twitter for predicting stock exchange movement. *Advances in Science, Technology and Engineering Systems*, 4(3), 244–250. <http://dx.doi.org/10.25046/aj040332>
- Tranfield, D., Denyer, D. & Smart, P. (2003). Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review. *British Journal of Management*, 14(3), 207–222. <https://doi.org/10.1111/1467-8551.00375>
- Trigueiro, F. M. C. & Marques, N. A. (2014). Teorias da Administração I [Universidade Federal de Santa Catarina]. In *Universidade Federal de Santa Catarina - UFSC* (3ª Edição). https://educapes.capes.gov.br/bitstream/capes/401253/1/Livro_Teorias_da_Administracao_I_WEB_atualizado.pdf
- Turney, P. D. (2002). Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. *40th Annual Meeting of the Association for Computational Linguistics (ACL)*, 417–424.
- Van Eck, N. J. & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84, 523–538. <https://doi.org/10.1007/s11192-009-0146-3>
- Van Eck, N. J. & Waltman, L. (2022). VOSviewer Manual. In *Leiden: Univeristeit Leiden* (Issue 1.6.18). https://www.vosviewer.com/documentation/Manual_VOSviewer_1.6.18.pdf
- Wankhade, M., Rao, A. C. S. & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, 55, 5731–5780. <https://doi.org/10.1007/s10462-022-10144-1>
- Wartick, S. L. & Cochran, P. L. (1985). The Evolution of the Corporate Social Performance Model. *Academy of Management Review*, 10(4), 758–769. <https://doi.org/10.5465/amr.1985.4279099>
- Westbrook, R. A. (1987). Product/Consumption-Based Affective Responses and Postpurchase Processes. *Journal of Marketing Research*, 24(3), 258–270. <https://doi.org/10.2307/3151636>
- Wood, D. J. (1991). Corporate Social Performance Revisited. *The Academy of Management Review*, 16(4), 691–718. <https://doi.org/10.2307/258977>
- Yadav, A., Jha, C. K., Sharan, A. & Vaish, V. (2020). Sentiment analysis of financial news using unsupervised approach. *Procedia Computer Science*, 167, 589–598. <https://doi.org/10.1016/j.procs.2020.03.325>
- Yao, Q., Li, R. Y. M., Song, L. & Crabbe, M. J. C. (2021). Safety knowledge sharing on Twitter: A social network analysis. *Safety Science*, 143, 1–7. <https://doi.org/10.1016/j.ssci.2021.105411>
- Yoo, B. & Rayz, J. (2021). Understanding Emojis for Sentiment Analysis. *The International FLAIRS Conference Proceedings*, 34. <https://doi.org/10.32473/flairs.v34i1.128562>
- Zerfass, A. & Viertmann, C. (2017). Creating business value through corporate communication. *Journal of Communication Management*, 21(1), 68–81. <https://doi.org/10.1108/JCOM-07-2016-0059>

Zhang, D., Li, Y. & Wu, C. (2020). An extended TODIM method to rank products with online reviews under intuitionistic fuzzy environment. *Journal of the Operational Research Society*, 71(2), 322–334. <https://doi.org/10.1080/01605682.2018.1545519>