

Análise de Sentimento: Um estudo de performance entre comentários e vendas de produtos em um e-Commerce pautado em Text Mining

Guilherme Iglesias Alonso, Letícia Panegossi Toniolo, Fábio Silva Lopes

¹Faculdade de Informática e Computação (FCI) -
Universidade Presbiteriana Mackenzie (UPM) São Paulo – SP – Brasil

(32086326, 32086342)@mackenzista.com.br, fabio.lopes@mackenzie.br

Abstract. *The popularization of e-Commerce in the midst of the digital age has brought conveniences to consumers, since it allows purchases to be made with greater agility and practicality. In this context, which seems insecure to those who have never made an online purchase, buyers tend to evaluate products received, making their impressions public on the networks. However, because it is not clear how the evaluations impact the sale of a product, this work aimed to carry out a sentiment analysis of an e-Commerce in order to understand the relationship between evaluation and sales performance, contributing to the improvement business performance as well as scientific, by providing tools that allow analysis in this and other contexts. Seeking to understand the relationships between the variables mentioned, the approach presented as a result sales trends based on the polarities of the reviews made in a period of time.*

Keywords: *e-commerce; sentimental analysis; text mining; product reviews*

Resumo. *A popularização do comércio eletrônico em meio a era digital, trouxe consigo comodidades aos consumidores, uma vez que possibilita que compras sejam feitas com maior agilidade e praticidade. Neste contexto, que parece inseguro àqueles que nunca realizaram uma compra online, compradores tendem a avaliar produtos recebidos, deixando pública suas impressões nas redes. Porém, por não estar claro ao certo de que forma as avaliações impactam na venda de um produto, este trabalho visou realizar uma análise de sentimento de um e-Commerce a fim de entender a relação entre avaliação e performance de vendas, contribuindo para a melhoria de desempenho empresarial como também científico, ao disponibilizar ferramentas que permitam análises neste e em outros contextos. Buscando entender as relações entre as variáveis citadas, a abordagem apresentou como resultado tendências de vendas baseadas nas polaridades das avaliações feitas em um período de tempo.*

Palavras-Chave: *e-commerce; análise de sentimentos; text mining; avaliação de produtos*

1. Introdução

Com o surgimento da internet, tornou-se possível a conexão e interação entre as pessoas, permitindo também, que muitas empresas participassem e se inserissem neste meio [Müller 2013]. Por isso, ao passar dos anos como consequência de sua evolução, estas mesmas empresas passaram a comercializar seus produtos e serviços de forma on-line,

utilizando-se do comércio eletrônico para tal e possibilitando de maneira benéfica, que a comunidade pudesse realizar compras com maior rapidez e praticidade [Cruz e Silva 2012].

Dessa forma, o e-commerce - como é amplamente conhecido - tem se expandido a cada dia, e a expectativa, de acordo com um estudo realizado pela Grand View Research (2019), é de que seu mercado global atinja US\$ 27,15 trilhões até 2027, o que implica dizer que é um modelo de negócio em ascensão.

Contudo, apesar de suas vantagens, este mesmo formato de comercialização também traz consigo alguns problemas, principalmente no que diz respeito à experiência do cliente, uma vez que ao realizar compras on-line, existe o risco de por exemplo, obter produtos que divirjam daqueles anunciados ou não estejam de acordo com o que se espera, sendo por isso, necessárias garantias a fim de permitir que o consumidor possa retornar o que foi comprado e ser por fim reembolsado no caso de insatisfação [Devderea e Toader 2019].

Estes casos, afetam negativamente a percepção dos clientes acerca do e-commerce, e repelem novas compras uma vez que, uma das razões pelas quais consumidores deixam de realizar suas compras on-line, é a falta de confiança [Petrovic et al. 2003]. Uma vez que os consumidores não têm certeza do comportamento de um vendedor, e se este cumprirá com o combinado da venda de forma ética, a confiança se torna um pilar primordial neste ambiente [Gefen e Straub 2003].

A construção da confiança envolve conhecer o vendedor antes de comprar um produto para evitar situações indesejáveis [Gefen e Straub 2003]. Impressões de outros compradores deixados sobre uma compra anteriormente feita também se mostram relevantes, uma vez que de acordo com uma pesquisa feita pela Bright Local (2016), 84% dos entrevistados confiam e se baseiam em resenhas on-line.

Por isso, muitos dos clientes se auxiliam na construção da confiança se manifestando através de comentários e reviews, ao tornarem públicas suas opiniões sobre o produto ou serviço adquirido aos demais interessados, o que lhes causa conseqüentemente impressões, sejam elas positivas ou negativas, e que tendem a impactar diretamente no crescimento da venda de determinados bens [Hendrawan, Suryani e Oktavia 2017].

Visto a dificuldade de se identificar produtos avaliados e entender a relação existente entre avaliação e desempenho, além do obstáculo encontrado pelo setor de e-commerce para análise de seus produtos devido a falta de ferramentas que possibilitem-na, podemos identificar a pergunta da pesquisa sendo: é possível criar estratégias ou ferramentas que permitam avaliar a relação entre performance de vendas e avaliação de produtos feita pelos clientes?

Dito isso, a fim de responder a questão, a atual pesquisa tem como objetivo realizar a análise de sentimentos de determinados produtos, ao observar seus comentários, utilizando técnicas de data analytics como a mineração de texto, conhecida também como Text Mining de forma a relacionar e compreender as duas variáveis, além de visar colaborar para a melhoria das atividades de e-commerce de modo a performar melhor seus produtos à venda.

Portanto, a contribuição deste trabalho está direcionada tanto a uma melhoria de

desempenho 1) empresarial, uma vez que possibilita adoção de novas técnicas que auxiliarão no aumento da performance de vendas quanto 2) científico, ao permitir o uso da aplicação de algoritmo em contextos diversos, colaborando para a generalização das ferramentas em outros cenários que se apliquem.

Este trabalho foi estruturado da seguinte forma: na seção 2 do presente documento, estão apresentados os referenciais teóricos pertinentes à mineração de dados e texto, como também à análise de sentimentos; a seção 3 contém as etapas da metodologia escolhida; e as seções 4 e 5 possuem respectivamente, os resultados e conclusões obtidas a partir da presente análise.

2. Referencial Teórico

2.1. Data Analytics e Big Data

Segundo o McKinsey Global Institute (2011, p.01), “Big Data” refere-se a conjuntos de dados cujo tamanho está além da capacidade das ferramentas típicas de software de banco de dados para capturar, armazenar, gerenciar e analisar”. Esse conceito é subjetivo, uma vez que não existe uma delimitação formal do tamanho dos dados que essas ferramentas poderiam ou não processar, dado o rápido avanço da tecnologia.

Os dados englobados no conceito de “Big Data” não possuem uma forma única, pois trata-se de tipos variados de informações, que requerem o uso de diferentes tecnologias e técnicas analíticas para serem trabalhados.

Entre as técnicas disponíveis para manipulação desses dados [Maltby 2014], destaca-se:

1. *Data Mining*: Uma combinação de técnicas de estatística e machine learning utilizada pra identificar padrões em grandes conjuntos de dados [Osman 2019]
2. Regras de Associação: Comumente utilizadas em *data mining*, é uma forma de encontrar ligações entre variáveis,
3. *Análise de Cluster*: Um tipo de data mining que divide os dados em categorias menores, de acordo com características em comum entre eles,
4. *Crowdsourcing*:, Técnica para coletar dados de um número de pessoas, geralmente via chamadas on-line para tarefas (pagas ou voluntárias). Exemplos de sites que coletam esse tipo de informação incluem o Galaxy Zoo, para coletar informações sobre a identificação de galáxias pelos usuários e o eBird, que recebe cantos de pássaros gravado por usuários para identificação da distribuição geográficas de aves [Shepherd 2012];
5. *Machine Learning*: Capacidade de aprender e sofrer mudanças de acordo com estímulos externos, adaptativamente, é o conceito central do “machine learning”. A máquina deve utilizar as informações obtidas para tomar as melhores decisões de acordo com o objetivo estabelecido, mesmo que essas informações não estejam completas. [Bonaccorso 2017].

Além das técnicas de *data analytics* existentes apresentadas, existe o *Text Mining*, sendo este uma extensão da mineração de dados, ou *data mining* como também é conhecido. *Text Mining* é uma técnica de manipulação de dados capaz de processar e extrair informações significativas provenientes de bases semi estruturadas ou não estruturadas [Machado et al. 2010].

Tais informações são provenientes de documentos no formato de texto, tais como e-mails, conteúdo de páginas WEB, documentos corporativos etc. A técnica “envolve outras áreas correlatas da informática como *machine learning*, linguística computacional, estatística e principalmente, a mineração de dados” [Hotho, Nurnberger e Paass 2005], que auxiliam na formatação textual, de forma que o computador consiga interpretar o conteúdo.

2.2. Text Mining

De acordo com Schmitz (2015), *Text Mining* é uma técnica capaz de analisar grandes quantidades de dados textuais, a fim de extrair informações relevantes que permitem obter conhecimento sobre assuntos específicos. A partir de técnicas de mineração de textos, - como a análise de sentimentos - é possível classificar documentos de texto com base nas opiniões e sentimentos ali presentes, de forma a mensurar a apreciação de uma marca, produto ou serviço por exemplo [Ferrari e Castro 2016, p. 40].

A metodologia básica aplicada e principal objetivo da mineração de texto é estabelecer padrões de relacionamento textuais, a partir de um documento ou texto, baseando-se na frequência em que certos termos aparecem, além de identificar aqueles considerados relevantes dentro do contexto analisado [Serapião et al. 2010].

O processo inicial de mineração de texto, consiste na aplicação de técnicas de pré-processamento como a tokenização, stemming e remoção de stop words [Mohan 2015], como no modelo da Figura 1:

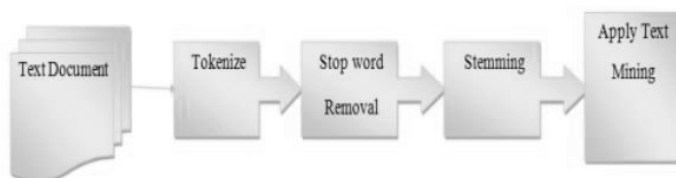


Figura 1. Flow Chart de Pré-Processamento de Text Mining [Soni, Kumar e Mathur 2015]

Estas etapas são necessárias para que os documentos sejam estruturados e padronizados, mantendo suas características originais, para que posteriormente seja possível a utilização de algoritmos de forma que estes consigam manipular os dados textuais igualmente [Pezzini 2016].

Após o pré-processado o texto, se torna possível por fim a extração do conhecimento, na qual é feita a interpretação dos resultados obtidos com auxílio de um usuário especializado e ferramentas que permitam filtrá-los [Schmitz 2015].

Segundo Soni, Kumar e Mathur (2015), a fim de extrair informações de fontes não estruturadas, o Text mining envolve etapas como: 1) sumarização do texto 2) classificação, 3) clusterização, 4) extração de informação e 5) visualização.

Na classificação de documentos, modelos como o Bag of Words (BoW) possibilitam selecionar e classificar o número de ocorrências distintas de palavras que são encontradas em um texto, e que por fim são apresentadas visualmente como um histograma [Qader, Ameen e Ahmed 2019].

O modelo, que é aplicado na fase de pré-processamento, é uma técnica amplamente utilizada para análise de sentimentos [El-Din 2016]. A importância do BoW portanto, se dá devido permitir classificar textos de acordo com o peso atribuído à cada uma das palavras encontradas [Qader, Ameen e Ahmed 2019]. Na figura abaixo, é possível identificar um exemplo de funcionamento do Bag of Words, na qual um texto é classificado a partir da quantidade de ocorrências identificadas de cada palavra:

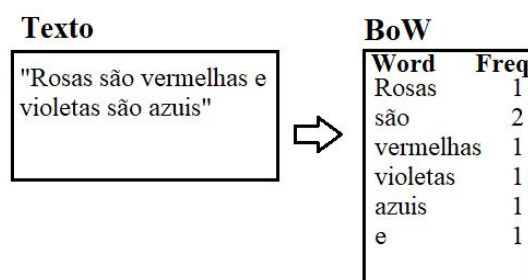


Figura 2. Bag of Words. (Elaborado pelos Autores)

Uma vez aplicadas técnicas de pré-processamento, a etapa de clusterização, ou agrupamento de texto, consiste em realizar a extração textual proveniente da grande quantidade de dados presentes no banco, de forma precisa e rápida [Soni, Kumar e Mathur 2015]. Desta forma, termos semelhantes em um documento podem finalmente ser agrupados [Jayabharathy e Kanmani 2014], como representado na Figura 3.

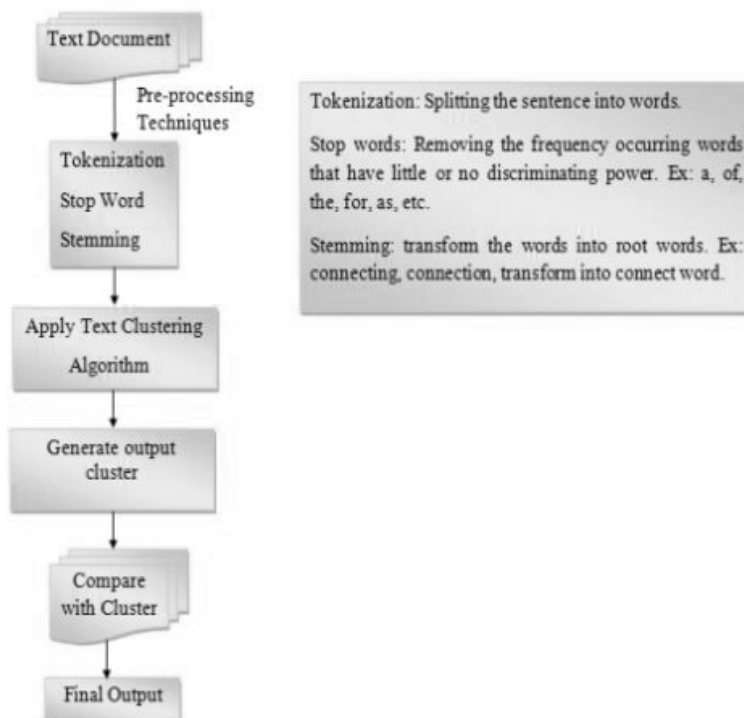


Figura 3. Flow Chart da Clusterização de Texto [Soni, Kumar e Mathur 2015]

A partir da clusterização, os documentos de texto são divididos automaticamente relacionando-se os níveis de similaridade entre si, em grupos denominados clusters. Tal

organização é realizada de maneira que exista 1) alta similaridade entre os documentos de um mesmo cluster e 2) baixa similaridade entre os elementos de clusters distintos [Gonçalves 2012].

As principais técnicas que envolvem a mineração de texto atualmente, compreendem 1) a *Extração de Informação (EI)*, responsável por realiza a extração das partes relevantes de um texto, atribuindo a ele propriedades específicas, e que por sua vez é considerada uma técnica mais limitada, no que diz respeito à compreensão da linguagem natural completa [Machado et al. 2010] 2) a *Recuperação de Informação (RI)*, a qual utiliza métodos estatísticos para processar dados textuais de maneira automática, buscando recuperar documentos a partir de palavras-chave, assim como ferramentas de busca, podendo ser considerado um método que se tornou mais popular com a ascensão da internet [Machado et al. 2010] e 3) o *Processamento de Linguagem Natural (PLN)*: que, segundo Abney (1991) citado por Machado et al. (2010), faz o uso de computadores e técnicas a fim de aumentar a rapidez no processamento de textos, além de melhorar a assimilação acerca da linguagem natural. Compreende a linguagem natural de inquéritos e manipulação de strings.

2.2.1. Tokenização

Segundo Ferreira (2019), token é uma sequência de caracteres encontrados em textos, e que são agrupados em unidade semântica útil, podendo ser uma única palavra ou em alguns casos, frases e parágrafos.

O processo de tokenização é uma etapa essencial no pré-processamento da linguagem natural (NLP) onde é muito utilizada [Michelbacher e Brenz 2013]. Ela é capaz de padronizar termos de um documento textual, removendo acentos, caracteres especiais e transformando letras maiúsculas em minúsculas [Soares, Prati e Monard, s.d.].

Desta maneira, tokenizar vocábulos na fase de pré-processamento segregando-os em unidades textuais significativas, auxilia na eficiência do uso de algoritmos, uma vez efetuada a análise dos dados. Todavia, a metodologia requer uso de técnicas em conjunto para aperfeiçoamento dos tokens, dado que em alguns casos, os espaços que funcionam como delimitadores podem dividir palavras compostas de forma imprecisa, em mais de um token, como nos casos de “São Paulo” e “Rio de Janeiro” [Ferreira 2019].

Para estes casos, o algoritmo é capaz de auxiliar na detecção dos termos compostos, uma vez identificada uma grande frequência de ocorrências e assim empregar, quando possível, dicionários de expressões [Schmitz 2015]. A figura abaixo mostra um exemplo do processo de tokenização:

Texto de entrada	Texto após a Tokenização
A ligeira <i>raposa marrom</i> salta sobre o <i>cão</i> preguiçoso.	"A", "ligeira", "raposa", "marrom", "salta", "sobre", "o", "cão", "preguiçoso"

Figura 4. Tokenização de palavras [Schmitz 2015]

Com o avanço da Linguagem de processamento natural (LPN), muitas ferramen-

tas e métodos surgiram para auxiliar no processo de tokenização de palavras, como é o caso do NLTK, spaCy, Flair e coreNLP [Sarker 2021]. Apesar de grande parte dos métodos suportarem apenas a língua inglesa, o NLTK (Natural Language ToolKit) é um dos exemplos que permite a tokenização de palavras presentes na língua portuguesa.

2.2.2. Stop Words

São consideradas como stop words, todas as palavras tidas como irrelevantes para a compreensão de um texto, como por exemplo artigos (a, as, o, os), preposições (de, por, para) e outras classes de palavras auxiliares, e que tendem a ser removidas uma vez realizado o pré-processamento da análise textual [Barbosa et al. 2017].

Texto de entrada	Texto após a Remoção de Stopwords
"A", "ligeira", "raposa", "marrom", "salta", "sobre", "o", "cão", "preguiçoso"	"ligeira", "raposa", "marrom", "salta", "sobre", "cão", "preguiçoso"

Figura 5. Exemplo de Stopwords [Schmitz 2015]

Para Moraes e Ambrósio (2007) estas palavras são suprimidas devido a não captarem a essência do documento, além de que anexá-las apenas aumentaria o tamanho do índice de maneira irrelevante. Atualmente existem diversos estudos que providenciam listas de stopwords (as chamadas stoplists) para uso público [Schmitz 2015].

As stoplists, vistas como um dicionário de stopwords, mudam conforme o idioma do documento que está sendo analisado (para o inglês por exemplo, seriam comuns palavras como "this", "the", "on", "at", o que não ocorreria para o português, que possui outras palavras como "a", "o", "para", "de") e no que diz respeito ao cenário em que se tenta analisar o problema [Gonçalves 2012]. Em um documento em que o conteúdo seja sobre futebol, "jogador" e "bola" podem por sua vez serem consideradas stopwords.

2.2.3. Stemming

Na gramática da língua portuguesa, o processo de formação de palavras ocorre por meio da derivação, que compreende a formação de outras novas, a partir de um radical comum, onde são adicionados prefixos, sufixos ou ambos [Correa 2009, p. 63].

No pré-processamento de Text Mining, a técnica de Stemming compreende reduzir palavras de diferentes classificações gramaticais - como verbos, pronomes etc - à sua forma base [Jivani 2011] visando eliminar possíveis derivações nas buscas por dados textuais, ao manter apenas seu radical comum.

Desta forma, as palavras uma vez reduzidas, trazem resultados mais acurados e com maior relevância ao varrerem documentos muito extensos [Haroon 2018], além de poupar tempo de processamento e espaço de memória [Mohan 2015]. Um exemplo, são as palavras Close, Closed, Closely e Closing, que através do Stemming serão suprimidas ao radical Clos [Haroon 2018].

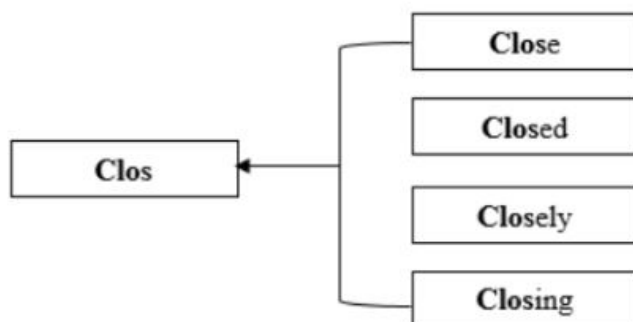


Figura 6. Exemplo de Stemming [Haroon 2018]

Acima, é possível visualizar este exemplo (Figura 6) na qual palavras da língua inglesa, derivadas de um mesmo radical, foram suprimidas à sua forma base pelo processo de stemming.

Neste caso específico, as palavras tiveram apenas seu sufixo removido, todavia é comum que além disso também haja remoção da prefixação dos termos em questão.

Segundo Haroon (2018), existem diferentes tipos de algoritmos de Stemming e esta variação está relacionada à capacidade, potência e estilo pela qual palavras sofrerão o processo de stemming. Ebecken et. al (2003) cita alguns clássicos, amplamente utilizados para processar palavras de língua inglesa, sendo:

- *Stemmer S*: Realiza a supressão das letras finais de palavras, como por exemplo de sufixos que formam plurais.
- *Stemmer de Porter*: Conhecido como Algoritmo de Porter, é capaz de extrair sufixos mais comuns de palavras e que se baseia em regras relacionadas às vogais-consoantes das mesmas para atuação do stemmer [Álvares 2014].
- *Stemmer de Lovins*: Remove apenas um sufixo por palavra, sendo este o mais longo encontrado. Para palavras da língua inglesa, o algoritmo é capaz de identificar e remover mais de 250 sufixos distintos.

No caso da Língua Portuguesa, existe o algoritmo Removedor de Sufixos da Língua Portuguesa (RSLP) proposto por Orenge e Huyck (2001) que se assemelha ao algoritmo de Porter, por ser capaz de realizar a remoção da sufixação, mas que diferentemente do primeiro, evita que sufixos que se assemelham a terminações de palavras sejam removidos, como por exemplo, “o sufixo de diminutivo “inho” pode ser removido de “passarinho”, mas não da palavra “linho”, apesar desta terminar com uma sequência de caracteres idêntica ao sufixo em questão” [Coelho 2007].

3. Trabalhos Correlatos

Existem diversos projetos que abordam a análise de sentimentos e que fazem uso da técnica de Text Mining para fins semelhantes ao proposto no presente artigo.

Hornikx e Hendriks (2015) apresentam uma análise de tweets postados por usuários da plataforma Twitter, a fim de entender como a rede social é utilizada para

avaliar marcas e comunicar avaliações, identificando as impressões das postagens deixadas pelos mesmos e a tendência de serem positivos ou negativos.

Lak e Turetken (2014) nos mostram outra abordagem para a análise de texto, onde comentários fornecidos por clientes de um e-commerce são analisados com técnicas que detectam automaticamente a polaridade das postagens, e que por fim são comparadas às classificações por estrelas de produtos, estipuladas pelos clientes para mensurar o teor de uma avaliação.

Por sua vez, Asif et al. (2020) realiza a análise de sentimento de postagens consideradas extremistas feitas por usuários do Facebook, utilizando a ferramenta de Data Miner Scrapper. A técnica utilizada em comentários e postagens, é capaz de rotulá-los entre quatro classes, incluindo: moderado, neutro, baixo extremo e alto extremo. Um classificador linear é utilizado como suporte com finalidade de classificar os comentários analisados.

Ao correlacionar a abordagem adotada para o desenvolvimento do projeto, pode-se afirmar que este se assemelha ao proposto por Lak e Turetken, uma vez que o intuito também compreende analisar comentários de um e-commerce, estipulando polaridades para análise das avaliações. Por outro lado, também há coincidência em relação a análise feita por Asif et al. no que diz respeito à técnica optada para rotulação de postagens em classes distintas, que busca classificar comentários em grupos.

4. Metodologia

4.1. Definição de Métodos

O método de pesquisa utilizado para realização da análise proposta foi de cunho experimental. De acordo com Wazlawick (2020, p. 42), a pesquisa experimental é uma das caracterizações clássicas de pesquisa que consiste em promover mudanças no cenário pesquisado de maneira sistêmica, de forma a criar situações propositais a fim de analisar se estas mesmas intervenções geram os resultados almejados e qual a reação provocada no ambiente.

Para realização do experimento portanto, foi adotado como metodologia o CRISP-DM sendo este um processo de mineração de dados amplamente utilizado em pesquisas experimentais.

O CRISP-DM visa solucionar problemas de negócio, sendo descrito como um modelo hierárquico, na qual um conjunto de tarefas é executado e dividido em seis fases [Sastry et al. s.d.] que englobam: compreensão do negócio, entendimento dos dados, preparação dos dados, modelagem, avaliação e implantação.

O modelo, utilizado por pesquisadores, teve seu desenvolvimento em decorrência da indústria, onde havia a necessidade de um modelo de processo que fosse padronizado, na mineração de dados [Schröer, Kruse e Gómez 2021].

Além disso, a metodologia CRISP-DM possibilita que um projeto de DM seja criado por usuários mesmo que iniciantes e não possui etapas rígidas, sendo possível revisita-las enquanto o projeto avança [Ramos et al. 2020].

Abaixo é possível visualizar as etapas do modelo, e a sequência em que cada tarefa ocorre, por meio da imagem a seguir:

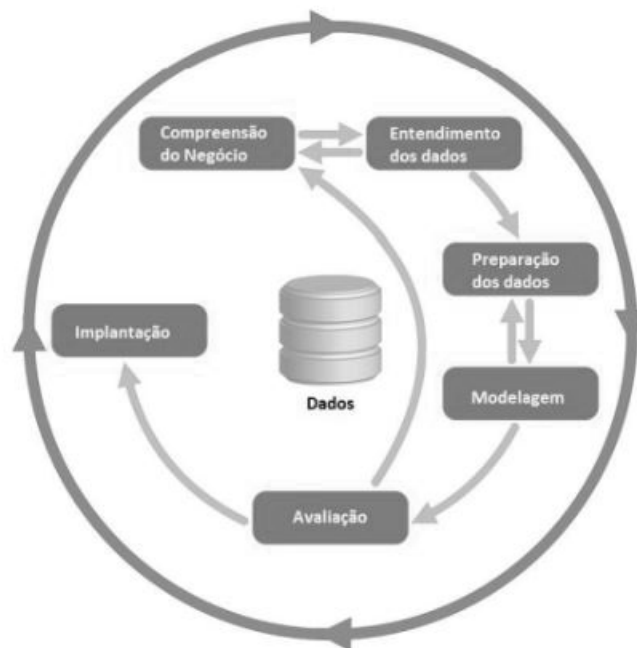


Figura 7. Ciclo do CRISP-DM [Shearer 2000]

Segundo Sastry et al. (s.d.), as etapas da metodologia se dividem da seguinte maneira:

1. *Compreensão do Negócio*: Consiste em entender o projeto e seus objetivos, em uma perspectiva de negócios a fim de definir o problema a ser tratado utilizando-se a mineração de dados;
2. *Entendimento dos Dados*: Inicia-se a coleta inicial de dados, a fim de familiarizar-se com os mesmos, para então detectar possíveis problemas de qualidade ali presentes além de explorar ideias;
3. *Preparação dos Dados*: Nesta etapa, atividades para construir o conjunto de dados final são realizadas a partir dos dados iniciais. Além disso, os dados são preparados para serem então utilizados em ferramentas de modelagem posteriormente. Algumas das tarefas desta fase incluem limpeza de dados e a seleção de atributos e tabelas;
4. *Modelagem*: Várias ferramentas e técnicas de modelagem são selecionadas e aplicadas. Parâmetros também são calibrados. A escolha de uma modelagem depende do problema a ser resolvido e cada uma possui requisitos específicos.
5. *Avaliação*: O modelo construído é avaliado antes de ser implantado, garantindo que os resultados e conhecimentos possam ser utilizados pelo negócio. Nesta fase também são avaliados os problemas do negócio para certificar que foram considerados.
6. *Implantação*: A etapa de implantação os novos *insights* são utilizados pelo negócio de forma que possa trazer melhorias e tornar processos de decisão mais eficientes [Ramos et al. 2020]. Isso inclui a geração de um relatório ou algo mais complexo, como implementar um processo de mineração de dados.

4.2. Etapas CRISP-DM

No presente projeto, a fase de *compreensão do negócio* é constituída pela problemática de se identificar produtos avaliados em um e-Commerce, e associá-los a sua performance de vendas, visando observar se as variáveis “performance” e “avaliações positivas” são associações causais e se relacionam de alguma forma.

Por isso, a fim de se realizar o experimento, foi escolhida para estudo, uma empresa de móveis e decoração que possui um sistema de e-Commerce já estabelecido no Brasil e que se encontra atualmente em expansão.

Na etapa de *entendimento dos dados*, a princípio foram coletados os dados de treinamento a fim de possibilitar a construção da análise e que, portanto, compuseram os dados iniciais. Visto que a ideia era analisar avaliações de produtos, inicialmente, foram utilizadas como dados de treino, 1.185 comentários, provenientes de avaliações de produtos variados de e-Commerces, os quais foram divididos entre três polaridades, sendo estas: Positivo, Neutro e Negativo.

Os comentários foram classificados manualmente, verificando o teor do conteúdo dos mesmos, ou seja, de acordo com a experiência descrita pelo consumidor. Comentários positivos possuem experiências exclusivamente positivas, apresentando palavras de satisfação e felicidade; experiências negativas possuem conteúdo totalmente negativo, contendo palavras de insatisfação, nojo, desprezo e etc e comentários neutros possuem percepções indiferentes dos consumidores.

Assim, uma vez finalizada a base de treinamento dos dados, deu-se início a fase de *preparação dos dados*, onde ocorreu a criação do script utilizando a plataforma Databricks.

Para preparar os dados, primeiramente houve a execução do treinamento da base anteriormente selecionada de 1.185 comentários, onde foram realizadas as técnicas de pré-processamento de text mining como remoção das stopwords (colocar exemplos de stopwords), stemming de vocábulos - na qual o objetivo foi realizar a remoção dos radicais das palavras - para ao fim do processo, ser feita a tokenização.

Uma vez realizada a tokenização, as frases dos comentários foram divididas em palavras, e classificadas com base no Bag of Words, de forma que cada palavra foi atrelada a um peso, o qual por fim foi atribuído ao sentimento que a frase de teste possuía. Este processo possibilitou o código entender que caso uma frase futura (do conjunto de dados final) possuísse aquele peso anteriormente estipulado, poderia ser classificado como sendo pertencente a determinado sentimento da análise.

Finalizado o treino, ocorreu a extração dos comentários da empresa de e-Commerce. Para este fim, utilizou-se o Google BigQuery para realização da extração, sendo o banco de dados que a empresa a qual concedeu os dados de estudo, possui acesso.

Após a extração destes dados, na *modelagem*, com a obtenção dos resultados dos testes anteriormente realizados com os dados de treino, ajustou-se o script, utilizando-se de uma função que ao ler os comentários dos produtos da empresa, foi capaz de realizar a comparação entre os pesos e palavras contidas neles e o resultado anterior dos testes, conseguindo assim classificar qual o percentual de polaridade que um determinado comentário teve em relação aos três sentimentos (positivo, neutro e negativo) contidos na

base de treino.

Ainda nesta etapa de modelagem, a escolha da plataforma Databricks para criação do script utilizado, se deu por ser uma plataforma online, com integração direta com os bancos de dados que foram utilizados no projeto e também devido permitir o uso de linguagens como SQL e Python na construção de um mesmo script.

Além disso, feita a escolha da plataforma, utilizamos a linguagem Python para a criação do código, com as bibliotecas Pandas para criação e manipulação de dados.

A escolha da linguagem ocorreu com intuito de desempenhar tarefas de NLP uma vez que esta linguagem permite acesso à biblioteca de NLTK (Natural Language Toolkit) sendo também uma das principais linguagens de pré-processamento de texto existentes, e que contém algoritmos e conjunto de dados relevantes para esta área de análise escolhida.

O script então passou posteriormente, pela fase de *avaliação* a qual foi dividida em duas etapas. A primeira, foi realizar testes com base nos dados de treinamento disponibilizados, permitindo que novas palavras fossem treinadas e a segunda, consistiu em realizar uma validação simples, por uma matriz de confusão, que serviu como base para validação da acurácia do código desenvolvido.

Na matriz de confusão, dos 389 comentários considerados como negativos, 362 foram realmente classificados como corretos, 13 foram classificados como neutro e 14 como positivos. Então, após verificar que o treino possuía uma acurácia aceitável, os dados do e-commerce puderam finalmente ser utilizados no script.

	Negativo	Neutro	Positivo
Negativo	362	13	14
Neutro	10	299	32
Positivo	2	11	439

Figura 8. Matriz de Confusão. (Elaborado pelos Autores)

A *implantação* consistiu em disponibilizar a análise à empresa estudada, de forma que a mesma conseguisse gerar e captar novos insights por meio dos resultados obtidos.

Esta implantação tem como objetivo melhorar o processo de decisão empresarial, permitindo que 1) a análise gere um relatório recorrente sobre os produtos disponíveis na plataforma, contendo suas respectivas relação entre as avaliações deixadas pelos clientes e performance de venda, de modo a estudar maneiras de aumentar o desempenho e visibilidade de cada produto e 2) seja gerada uma base contendo apenas produtos que tenham sua venda impactada por comentários, com intuito de como na opção anterior, estabelecer formas de impulsionar o comportamento de cada produto no site.

Outra forma de implantação da análise, é a disponibilização da mesma para fins científicos, ao possibilitar que o script criado possa ser utilizado em outros contextos, podendo ser disseminado.

5. Resultados e Discussão

Para a base de treino aplicada na pesquisa proposta, como supracitado, foram coletados 1.185 comentários de produtos aleatórios, provenientes de um e-Commerce já também estabelecido no Brasil, assim como a empresa estudada, na qual, 452 resultaram como positivos, 389 negativos e 341 neutros de acordo com o script criado para análise de sentimento.

Atualmente existem bibliotecas prontas para realização de análises de sentimento, porém mesmo com o conhecimento sobre estas bibliotecas que poderiam ser aplicadas no experimento, foi considerado o período de curva de aprendizado sobre as mesmas e, portanto não foram utilizadas para a primeira versão do projeto.

Então, após realizados os testes e definidos os parâmetros, foram selecionados 780 comentários de produtos diversos, oriundos da empresa escolhida para realização do experimento.

De acordo com a distribuição de notas do website (em uma escala de 0 a 5 estrelas, sendo 5 a classificação mais alta), 77.4% dos comentários ficaram entre os melhores avaliados, o que corresponde ao valor de 604, 5.1% obtiveram nota 4 (totalizando 40), 5.9% foram classificados com 3 estrelas (46), 3% resultaram em nota igual a 2 (23), e 8.6% tiveram o pior resultado de uma única estrela na avaliação final (67).

A partir dos gráficos abaixo, é possível visualizar a relação entre o número total de comentários e sua respectiva porcentagem.

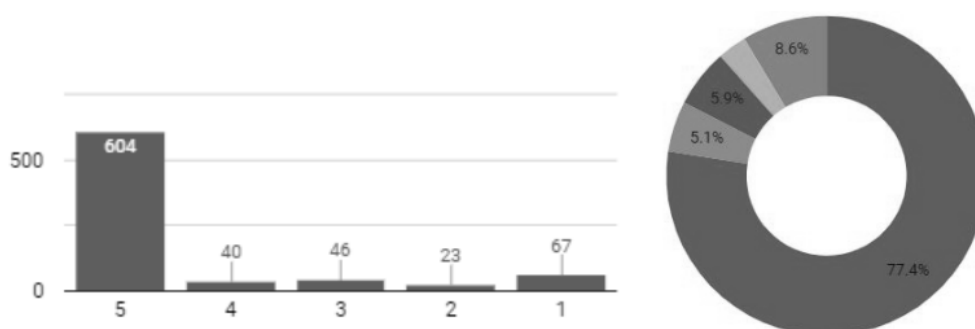


Figura 9. Distribuição de Notas de comentários de um e-Commerce. (Elaborado pelos Autores)

No código, após seleção e treinamento dos comentários de treino como visto acima, realizou-se então, a separação dos 780 comentários da empresa, em 3 sentimentos distintos (positivo, neutro e negativo) que foram classificados a partir da tokenização de palavras e do processo de mineração de texto, provenientes da base de treinamento.

Estes sentimentos por sua vez, obtiveram uma distribuição pouco distribuída, uma vez que mais de 71% foram classificados pelo script como sendo um sentimento positivo:

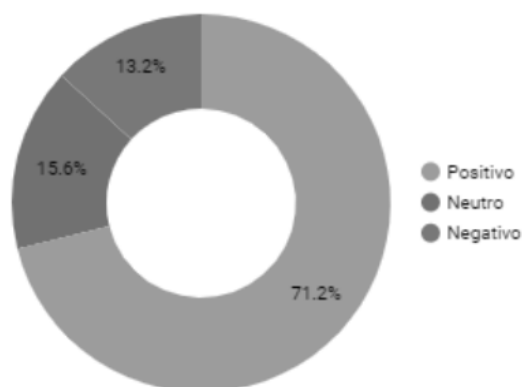


Figura 10. Resultado da Distribuição de Sentimentos. (Elaborado pelos Autores)

Já a classificação de comentários por item, se deu da seguinte forma, quando consideradas as polaridades:

Comentários de SKUs	Quantidade	% Relativa
Somente Positivos	489	69.36%
Somente Negativos	85	12.06%
Somente Neutros	103	14.61%
Somente Positivos e Negativos	11	1.56%
Somente Negativos e Neutros	2	0.28%
Somente Neutros e Positivos	13	1.84%
Somente Positivos, Negativos e Neutros	2	0.28%

Figura 11. Resultado da Distribuição de Comentários. (Elaborado pelos Autores)

Ao relacionar a performance dos produtos levando em consideração os comentários e os sentimentos obtidos, foram realizadas as seguintes etapas:

Primeiramente, foi preciso definir regras para realização de cálculos de performance de vendas de um produto (comparando o antes vs. depois de uma avaliação), uma vez que um único produto, pode ter inúmeros comentários em diversas datas distintas e os mesmos poderiam ter sido alocados dentro das 3 polaridades de sentimento. Sendo assim, foi escolhida para análise de performance de vendas, a primeira data de comentário do produto, independente de sua polaridade.

Em sequência, sabendo que a base do cálculo de performance seria a primeira data do comentário feito para um determinado produto, foi escolhido o período que compreendia 15 dias anteriores ao comentário feito, e 15 dias após o primeiro comentário para realização da análise. O período citado foi escolhido pois por se tratar de um e-commerce, é necessário levar em consideração o tempo de entrega do produto para que o consumidor pudesse avaliar sua compra.

Com isso, é possível observar que a grande maioria dos itens contém somente comentários positivos, seguidos de comentários considerados neutros e por fim, negativos. Produtos com dois ou mais sentimentos, estão em menor escala chegando a pouco menos

de 4 da amostra total.

Por outro lado, ao relacionar os comentários à performance do produto, temos os seguintes cenários:

Comentários de SKUs	Vendas Antes dos Comentários	Venda Depois dos Comentários	% Relativa
Todos produtos sem sentimentos atribuidos	4960	3244	-34,6%
Somente Positivo	3131	2046	-34,65%
Somente Negativos	565	309	-45,31%
Somente Neutros	586	366	-37,54%
Somente Positivos e Negativos	227	167	-26,43%
Somente Positivo e Neutro	306	195	-36,27%
Somente Negativo e Neutro	8	4	-50,00%
Somente Positivos, Negativos e Neutros	137	157	14,6%

Figura 12. Comentários vs. Vendas de Produtos para 15 dias. (Elaborado pelos Autores)

Considerando a primeira linha métrica da tabela, onde é possível identificar as vendas dos produtos da amostra, sem levar em consideração qual o sentimento atribuído a elas, nota-se que há um decréscimo de 34,6%, o que pode levar a concluir que se um produto tem qualquer comentário, seja qual for sua polaridade, há uma queda nas vendas. Porém, ao olhar esses mesmos produtos, dividindo suas respectivas polaridades, observa-se que os comentários podem afetar a venda dos produtos ainda mais. Produtos com somente comentários Negativos por exemplo, caem sua venda em 45%, ao passo que nos produtos com somente comentários positivos, o valor sobe e fica próximo dos -35%.

Os únicos produtos que tiveram alta em suas vendas, foram os que obtiveram as 3 polaridades comentadas, mesmo sendo somente 2 itens. No caso acima, eles obtiveram um aumento de vendas em 14,6 o que leva a crer ser o resultado de um compartilhamento de opiniões que tende fazer com que o consumidor compre ou não um produto.

Aumentando o histórico das vendas para um acompanhamento de como a performance e os comentários se relacionam, foram estipulados os períodos de 30 e 45 dias respectivamente.

Comentários de SKUs	Vendas Antes dos Comentários	Venda Depois dos Comentários	% Relativa
Todos produtos sem sentimentos atribuidos	7836	4986	-36,37%
Somente Positivo	4962	3001	-39,52%
Somente Negativos	797	546	-31,49%
Somente Neutros	939	565	-39,83%
Somente Positivos e Negativos	376	264	-29,79%
Somente Positivo e Neutro	447	323	-27,74%
Somente Negativo e Neutro	12	9	-25,00%
Somente Positivos, Negativos e Neutros	303	278	-8,25%

Figura 13. Comentários vs. Vendas de Produtos para 30 dias. (Elaborado pelos Autores)

Ao analisar os dados, percebe-se que quanto mais tempo se passa, mais o produto performa, independente do seu comentário.

Em 30 dias, houve um impacto bastante expressivo em produtos os qual obtiveram comentários não importando sua polaridade com uma redução de 36% nas vendas. Já produtos que tiveram somente uma avaliação, o impacto foi semelhante ou em alguns casos, como nos de polaridade positiva, até pior. O interessante é reparar que produtos

que obtiveram as 3 polaridades nos comentários, também tiveram venda relativa muito maior que os demais produtos classificados.

Comentários de SKUs	Vendas Antes dos Comentários	Venda Depois dos Comentários	% Relativa
Todos produtos sem sentimentos atribuídos	9874	7920	-19,79%
Somente Positivo	6156	4567	-25,81%
Somente Negativos	1034	849	-13,54%
Somente Neutros	1212	820	-32,34%
Somente Positivos e Negativos	457	399	-12,69%
Somente Positivo e Neutro	560	629	12,32%
Somente Negativo e Neutro	15	16	6,67%
Somente Positivos, Negativos e Neutros	440	595	35,23%

Figura 14. Comentários vs. Vendas de Produtos para 45 dias. (Elaborado pelos Autores)

Não obstante, quando observado o período de 45 dias, visualiza-se um cenário diferente, em que o impacto do comentário dentro da venda dos produtos foi o menor se comparado aos demais períodos (19%), ao passo que produtos com mais de um sentimento identificado nas avaliações, obtiveram uma melhor performance que os classificados somente em uma única polaridade, com destaque nos produtos que tiveram as 3 polaridades. Estes obtiveram um aumento de 35% nas vendas neste período.

6. Conclusões

Visto o crescimento dos sistemas de comércio online e consequentemente, de clientes que, ao realizarem este tipo de consumo, tornam públicas suas avaliações sobre suas experiências de compra nas redes, surge como discutido anteriormente, a dificuldade de se identificar produtos avaliados em um e-commerce e qual a relação desta variável com a performance de vendas, tanto pelo fato de faltarem ferramentas ao setor que possibilitem quaisquer análises, quanto por ser um setor ainda em ascensão e que necessita adequação de estratégias.

Por conseguinte, o projeto teve como principal objetivo auxiliar na identificação e análise de comentários de produtos de um e-Commerce estabelecido no Brasil e atualmente em expansão, submetendo-os à uma análise de sentimento por meio de técnicas de text mining, a fim de compreender as duas variáveis e como elas se comportaram ao decorrer do período estipulado no experimento.

A partir da análise de mineração de texto realizada para uma gama de produtos da empresa nos períodos de 15, 30 e 45 dias, na qual foram classificados os comentários deixados em 3 polaridades (positivo, neutro e negativo) e verificado o histórico de vendas no mesmo período, observou-se como resultado que produtos tendem a aumentar sua performance de vendas ao decorrer do tempo, sem estarem diretamente relacionados ao comentário deixado.

De forma geral, produtos que possuíram as 3 polaridades de avaliação, foram os que tiveram maior desempenho em suas vendas, chegando a atingir 35% de melhoria quando constatado por exemplo, 45 dias.

Com base nos resultados portanto, é possível afirmar que os comentários podem sim ter um impacto na decisão de compra de um cliente. Os produtos com mais polaridades, ajudam consumidores na hora de finalizar uma compra, pois geram diversas opiniões e como consequência, o ajudam a fazer melhores escolhas como compradores.

Acredita-se que o processo de estudo para futuros trabalhos não esteja finalizado e sim precise ser aperfeiçoado melhorando a base de treino com a inserção de mais amostras de comentários e um olhar mais profundo na questão da interação do usuário com o comentário, a fim de aumentar a acurácia do algoritmo.

Isso poderia ser feito, analisando se um cliente de fato observou os comentários na página do produto ao realizar sua compra, ou se este fora na verdade, adicionado diretamente no carrinho.

Além disso, outras melhorias dizem respeito à 1) classificação dos comentários que, em vez de serem associados como sendo neutro, positivo ou negativo de acordo com o termo utilizado em cada avaliação, poderia ser feita a partir da atribuição de um score, na qual valores positivos seriam atribuídos como sentimentos positivos, valores negativos seriam vistos como negativos e valores nulos (ou próximos de zero de acordo com algum critério) seriam verdadeiramente neutros e 2) sazonalidades de vendas, em que se verificaria se as oscilações nas observadas nas vendas e períodos escolhidos para análise, ocorreram devido a sazonalidades conhecidas do mercado ou não.

Por outro lado, em suma, o resultado foi satisfatório e possibilita que a empresa em questão possa com a análise, estar ciente de como seus produtos estão performando e conseqüentemente, como estão sendo observados por seus compradores, podendo desta forma criar novas estratégias de vendas para melhorar a percepção dos mesmos, e além disso, permitindo também que a análise possa ser utilizada por outros departamentos como mídias sociais.

E, em um contexto científico, a pesquisa agora permite que o algoritmo e ferramentas criadas para análise possam ser disseminados e possam contribuir para análise de comentários neste e em outros contextos, caso julgado pertinente.

Referências

- Abney, S. (1991). *Principle-Based Parsing: Computation and Psycholinguistics*. Norwell: Kluwer Academic Publishers.
- Alvares, R. V. (2014) *Algoritmos de Stemming e o Estudo de Proteomas*. Universidade Federal do Rio de Janeiro.
- Asif, M. et al. (2020) "Sentiment analysis of extremism in social media from textual information," *Telematics and Informatics*, 48.
- Barbosa, J. L. N. et al. (2017) "Introdução ao Processamento de Linguagem Natural usando Python," in *Livro Anais - Artigos e Minicursos. III Escola Regional de Informática do Piauí*, pp. 336–360.
- Bonaccorso, G. (2017). *Machine Learning Algorithms*. Birmingham: Packt Publishing.
- BrightLocal (2016) "Local Consumer Review Survey". Disponível em: <https://www.brightlocal.com/learn/local-consumer-review-survey/>.
- de Castro, L. N. e Ferrari, D. G. (2016) *Introdução à Mineração de Dados: Conceitos Básicos, Algoritmos e Aplicações*. Editora Saraiva.
- Chapman, P. et al. (1999) "The CRISP-DM User Guide"
- Coelho, A. R. (2007) *Stemming para a língua portuguesa: estudo, análise e melhoria do*

algoritmo RSLP. Universidade Federal do Rio Grande do Sul.

Correa, V. L. (2009) *Língua Portuguesa: Da Oralidade À Escrita*. Curitiba, PR: IESDE Brasil.

Cruz, C. A. B. da e Silva, L. L. da (2012) “A sociedade e a Influência do e-Commerce,” *Interfaces Científicas*, 01, pp. 81–90.

Devderea, C. e Toader, F. (2018) “Consumer Behavior Towards Apparel E-Commerce in Romania”, *Management Dynamics in the Knowledge Economy*, 6(3), pp. 471–487.

Ebecken, N., Lopes, M. e Costa, M. (2003) *Mineração de Textos*, chapter 13, pp. 337–370.

El-Din, D. M. (2016) “Enhancement Bag-of-Words Model for Solving the Challenges of Sentiment Analysis,” *International Journal of Advanced Computer Science and Applications*, 7, pp. 244–252.

Ferreira, H. H. (2019) *Processamento de Linguagem Natural e Classificação de textos em Sistemas Modulares*. Universidade de Brasília.

Gefen, D. e Straub, D. (eds.) (2003) *Managing User Trust in B2C e-Services*. *e-Service Journal*. doi: 10.2979/ESJ.2003.2.2.7.

Gonçalves, E. C. (2012) “Mineração de Texto,” *SQL Magazine*, 105, pp. 31–44.

Grand View Research (2019) “E-commerce Market Size, Share Trends Analysis Report By Model Type (B2B, B2C), By Region (North America, Europe, APAC, Latin America, Middle East Africa), And Segment Forecasts 2020 - 2027.” Available at: <https://www.grandviewresearch.com/industry-analysis/e-commerce-market>.

Haroon, M. (2018) “Comparative Analysis of Stemming Algorithms for Web Text Mining,” *I.J. Modern Education and Computer Science*, 9, pp. 20–25.

Hendrawan, R. A., Suryani, E. e Oktavia, R. (2017) “Evaluation of E-Commerce Product Reviews Based on Structural, Metadata, and Readability Characteristics,” *Procedia Computer Science*, pp. 280–286.

Hornikx, J. e Hendriks, B. (2015). Tweets de consumidores sobre marcas: uma análise de conteúdo de tweets de sentimento sobre bens e serviços. *Journal of Creative Communications*, 10 (2), pp. 176–185

Hotho, A., Nurnberger, A. e Paass, G. (2005). A brief survey of text mining. *LDV Forum - GLDV Journal for Computational Linguistics and Language Technology*, v. 20, n. 1, pp. 19-62

Jayabharathy, J. e Kanmani, S. (2014) Correlated concept based dynamic document clustering algorithms for newsgroups and scientific literature.

Jivani, A. G. (2011) “A Comparative Study of Stemming Algorithms,” *International Journal of Circuit Theory and Applications* 2, pp. 1930–1938.

Lak, P. e Turetken, O. (2014) “Star Ratings versus Sentiment Analysis – A Comparison of Explicit and Implicit Measures of Opinions” 47th Hawaii International Conference on System Sciences, pp. 796-805, doi: 10.1109/HICSS.2014.106.

- Machado, A. P. et al. (2010) “Mineração de Texto em Redes Sociais Aplicada à Educação a Distância,” *Colabor@ - Revista Digital da CVA*, 6.
- Maltby, D. (2011). *Big Data Analytics*. In: ASIST 2011, New Orleans, USA, pp. 1-6. McKinsey Global Institute. (2011). *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Company.
- Michelbacher, L. (2014). Multi-word tokenization for natural language processing.
- Mohan, V. (2015). *Preprocessing Techniques for Text Mining - An Overview*.
- Morais, E. A. M. e Ambrósio, A. P. L. (2007) *Mineração de Textos*. Universidade Federal de Goiás.
- Müller, V. N. (2013) *e-Commerce: Vendas pela Internet*. Fundação Educacional do Município de Assis.
- Osman, A. S. (2019). *Data Mining techniques: Review*. IJDSR, v. 2, issue 1, pp. 1-4
- Petrovic, O. et al. (2003) *Trust in the Network Economy*. Santa Clara, CA: Springer-Verlag TELOS.
- Pezzini, A. (2016) “Mineração de Textos: Conceito, Processo e Aplicações,” *R. Eletr. do Alto Vale do Itajaí – REAVI*, 5, pp. 01–13.
- Qader, W. A., Ameen, M. M. e Ahmed, B. I. (2019) “An Overview of Bag of Words: Importance, Implementation, Applications, and Challenges.”
- Ramos, J. L. C. et al. (2020) *CRISP-EDM: uma proposta de adaptação do Modelo CRISPDM para mineração de dados educacionais*. Universidade Federal do Vale do São Francisco.
- Sarker, S. (2021) *BNLP: Natural language processing toolkit for Bengali*. Begum Rokeya University.
- Sastry, H. (2013) “Implementation of CRISP Methodology for ERP Systems,” *International Journal of Computer Science Engineering* 2, pp. 203–217.
- Schmitz, F. E. B. (2015) *Aplicação da Técnica de Text Mining para Comentários Relacionados ao Contexto do Turismo*. Centro Universitário Luterano de Palmas.
- Schröer, C., Kruse, F. e Gómez, J. M. (2021) “A Systematic Literature Review on Applying CRISP-DM Process Model,” *Procedia Computer Science*, 181, pp. 526–534.
- Serapião, P. R. B., Suzuk, K. M. F. e Marques, P. M. A. (2010) “Uso de mineração de texto como ferramenta de avaliação da qualidade informacional em laudos eletrônicos de mamografia,” *Radiol Bras*, pp. 103–107.
- Shepherd, H. (2012). *Crowdsourcing*. *Contexts*, v. 11, n. 2, pp. 10-11.
- Soares, M. V. B., Prati, R. C. e Monard, M. C. (no date) *Melhorando o Desempenho Computacional e a Geração de Atributos na Ferramenta de Pré-Processamento de Textos Pretext*. Laboratório de Inteligência Computacional – LABIC .
- Soni, A. K., Kumar, A. e Mathur, R. P. (2015) “Enhancing the Stemming Algorithm in Text Mining,” *International Journal of Applied Engineering Research*, 10.
- Wazlawick, R. S. (2020) *Metodologia de Pesquisa para Ciência da Computação*. Elsevier.