



INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO PIAUÍ
CAMPUS PEDRO II
TECNOLOGIA EM ANÁLISE E DESENVOLVIMENTO DE SISTEMAS

ABDENAIDE LEITE RIBEIRO

**WEBCRAWLER INTEGRADO AO CHATGPT PARA ANÁLISE DE SENTIMENTOS
DE CLIENTES EM MARKETPLACES**

PEDRO II, PIAUÍ
2023

ABDENAIDE LEITE RIBEIRO

WEBCRAWLER INTEGRADO AO CHATGPT PARA ANÁLISE DE SENTIMENTOS DE
CLIENTES EM MARKETPLACES

Trabalho de Conclusão de Curso (artigo científico) apresentado como exigência parcial para obtenção do diploma do Curso de Análise e Desenvolvimento de Sistemas do Instituto Federal de Educação, Ciência e Tecnologia do Piauí, Pedro II.

Orientador: Prof.Es.Cleber da Silva Araujo

PEDRO II, PIAUÍ
2023

WEBCRAWLER INTEGRADO AO CHATGPT PARA ANÁLISE DE SENTIMENTOS DE CLIENTES EM MARKETPLACES

Abdenaide Leite Ribeiro
Prof.Es.Cleber da Silva Araujo

RESUMO

Este artigo apresenta a integração eficaz do WebCrawler com a API do ChatGPT para realizar análises de sentimentos em avaliações de clientes em plataformas de marketplace. Embora a extração de dados por meio de Web Crawlers e a análise de sentimentos em ambientes online sejam tópicos crescentemente populares, este estudo se destaca ao focar na aplicação específica a marketplaces. Além de descrever detalhadamente a metodologia utilizada, incluindo a coleta e organização de dados, o artigo enfatiza a importância da compreensão dos sentimentos e preferências dos clientes para o sucesso nos negócios. Ao explorar a valência emocional dos textos, classificando-os como positivos, negativos ou neutros, a pesquisa busca contribuir não apenas para o entendimento do sentimento expresso em avaliações de marketplace, mas também para a avaliação do desempenho e compreensão de uma inteligência artificial quando aplicada a comentários de usuários. Espera-se que este estudo forneça insights valiosos para áreas como marketing, publicidade, análise de opinião pública e previsão de tendências de mercado.

Palavras-chaves: ChatGPT, WebCrawler, Análise de Sentimentos, Marketplace.

ABSTRACT

This article presents the effective integration of WebCrawler with the ChatGPT API to perform sentiment analysis on customer reviews on marketplace platforms. While data extraction through Web Crawlers and sentiment analysis in online environments are increasingly popular topics, this study stands out by focusing on the specific application to marketplaces. In addition to detailing the methodology used, including data collection and organization, the article emphasizes the importance of understanding customer sentiments and preferences for business success. By exploring the emotional valence of texts, classifying them as positive, negative, or neutral, the research aims to contribute not only to understanding the sentiment expressed in marketplace reviews but also to evaluating the performance and understanding of artificial intelligence when applied to user comments. This study is expected to provide valuable insights for areas such as marketing, advertising, public opinion analysis, and market trend forecasting.

Keywords: ChatGPT, WebCrawler, Sentiment Analysis, Marketplace.

Data de aprovação: 10/01/2024

1 INTRODUÇÃO

Nos últimos anos, a extração de dados com *Web Crawlers* e a análise de sentimentos têm se tornado cada vez mais populares com o grande volume de dados não estruturados em ambientes online, como redes sociais e fóruns (BOUDIN; LIGOZAT, 2018). *Web Crawlers* são ferramentas que permitem coletar informações de maneira automatizada em páginas web, permitindo que grandes volumes de dados sejam processados em um curto período (BAE; KIM, 2020). A análise de sentimentos, por sua vez, é uma técnica que permite identificar a valência emocional de um texto, classificando-o como positivo, negativo ou neutro (LIU, 2012).

Essas técnicas têm sido aplicadas em diversas áreas, desde *marketing* e publicidade até análise de opinião pública e previsão de tendências de mercado (CAO *et al.*, 2020). A importância da opinião para as empresas é evidente nesse contexto, uma vez que a compreensão das emoções e preferências dos clientes é crucial para o sucesso nos negócios. Um dos campos promissores para a aplicação da extração de dados com *Web Crawlers* e a análise de sentimentos é a possibilidade da integração com a API¹ do *ChatGPT*², podendo ser aplicada em diversas áreas, tais como redes sociais, fóruns, sites de notícias e também no *marketplace* (BISWAS, 2023). No contexto do *marketplace*, a análise de sentimentos pode ser usada para identificar as valências dos clientes durante as interações e, a partir dessas informações, personalizar o atendimento ao cliente, detectar problemas e identificar áreas de melhoria (EL-ANSARI; BENI-HSSANE, 2023).

Nesse contexto, a aplicação da técnica de extração de dados com *Web Crawlers* integrado à API do *ChatGPT* para análise de sentimentos em avaliações de *marketplace* pode ser uma abordagem promissora para identificar tendências de mercado, preferências de clientes e áreas de melhoria em produtos e serviços (KASNECI *et al.*, 2023). A combinação dessas técnicas permite uma análise mais precisa e abrangente dos dados coletados em diversas plataformas online, como redes sociais, fóruns e sites de avaliações (BISWAS, 2023). O *ChatGPT*, por sua vez, pode ser utilizado para compreender e interpretar de forma mais profunda a linguagem natural utilizada pelos usuários nas avaliações, permitindo uma análise de sentimentos mais sofisticada e precisa (BROWN *et al.*, 2020). Além disso, essa abordagem pode ser usada para personalizar a experiência do cliente, oferecendo recomendações e respostas mais relevantes e personalizadas.

¹ <<https://aws.amazon.com/pt/what-is/api/>>

² <<https://chat.openai.com/>>

1.1 OBJETIVOS GERAIS

Prover um panorama das percepções de usuários de *marketplaces* baseado em análise de sentimentos por meio de técnicas de *web crawling* em conjunto com a API do *ChatGPT*.

1.2 OBJETIVOS ESPECÍFICOS

- Desenvolver um *Crawler* para extração de dados;
- Estruturar e organizar os dados coletados por *Web Crawlers*;
- Desenvolver um algoritmo com interação que sejam capazes de identificar com precisão o sentimento expresso em avaliações de *marketplaces*;
- Avaliar o desempenho e a compreensão de uma IA quando colocada para analisar comentários de usuários.

2 FUNDAMENTAÇÃO TEÓRICA

Esta seção apresenta conceitos fundamentais relevantes para compreensão deste artigo.

A extração de dados usando um *web crawler* e a análise de sentimento estão se tornando cada vez mais comuns na era da informação digital, principalmente devido à grande quantidade de dados não estruturados disponíveis em ambientes online como redes sociais e fóruns. Essas técnicas são amplamente utilizadas em vários campos, desde marketing e publicidade até análise de opinião pública e previsão de tendências de mercado (CAO *et al.*, 2020; GUO *et al.*, 2019)

2.1 WEB CRAWLER E ANÁLISE DE SENTIMENTOS

Web crawler é uma técnica de extração de dados que permite coletar informações de maneira automatizada em páginas web, permitindo que grandes volumes de dados sejam processados em um curto período (COELHO, 2022). Isso é possível por meio de softwares que rastreiam a web em busca de informações específicas, como *posts* em redes sociais ou comentários em fóruns.

A partir desses dados coletados, é possível realizar análises mais profundas, como a análise de sentimentos, essa técnica permite identificar a polaridade emocional de um texto, classificando-o como positivo, negativo ou neutro (LIU, 2012). Esse tipo de análise é possível graças à utilização de algoritmos que conseguem identificar padrões de linguagem e sentimentos expressos nas palavras.

Com o aumento da quantidade de dados não estruturados disponíveis em ambientes online, as técnicas de extração de dados usando um *web crawler* e a análise de sentimentos estão se tornando cada vez mais comuns na era da informação digital, essas técnicas têm diversas aplicações, desde marketing até análise de opinião pública e previsão de tendências de mercado (CAO *et al.*, 2020; GUO *et al.*, 2019).

2.2 PROCESSAMENTO DE LINGUAGEM NATURAL

O processamento de Linguagem Natural (PLN) é um método de análise de texto que permite que máquinas interpretem e avaliem a linguagem natural e é muito útil para automatizar tarefas que exigem muito trabalho manual sendo considerado um domínio interdisciplinar de pesquisa para complementar a análise de dados textuais extraídos de sites (CHEEMA; TARIQ; PIRES, 2023).

O objetivo do PLN é fazer com que o computador possa se comunicar em linguagem humana, em diferentes níveis de entendimento e/ou geração de sons, palavras, sentenças e discursos, a representação do significado de uma sentença,

independentemente do contexto, é obtida através de sua forma lógica, que codifica os possíveis sentidos de cada palavra e identifica os relacionamentos semânticos entre palavras e frases (GONZALEZ; LIMA, 2003).

A pesquisa sobre opiniões e sentimentos das pessoas, em especial com o uso de PLN, tem se tornado uma área ativa de pesquisa (RAMOS; FREITAS, 2019). Uma das razões para isso é a ampla gama de aplicações em diversos domínios, o que motivou o desenvolvimento de aplicações comerciais de análise de sentimento (ALBUQUERQUE, 2022).

2.3 COMÉRCIO ELETRÔNICO E *MARKETPLACE*

O comércio eletrônico teve origem com a popularização do acesso à internet nos anos 2000. No entanto, seu embrião remonta à década de 70 nos Estados Unidos, quando ocorreu o primeiro indício dessa evolução com a utilização de arquivos para a troca de solicitações de pedidos. Nesse contexto, as lojas passaram a receber informações sobre o interesse do cliente em produtos ou serviços por meio do compartilhamento de documentos através do Intercâmbio Eletrônico de Dados (IED) (GUNASEKARAN *et al.*, 2002). Além disso, o *marketplace* tem se mostrado uma ferramenta poderosa para as Pequenas e Médias Empresas (PMEs), oferecendo oportunidades únicas de iniciar negócios com pouco investimento e possibilitando um alcance global, redução de custos e melhorias na cadeia de fornecimento (GIL, 2022).

Desde então, o comércio eletrônico evoluiu significativamente e se tornou uma indústria globalizada e altamente desenvolvida. Com o avanço da tecnologia, a expansão da conectividade e o surgimento de plataformas especializadas, o comércio eletrônico alcançou novos patamares de sofisticação e conveniência (RAMOS, 2015). Os *marketplaces* surgiram como um marco significativo na transformação da forma como os consumidores realizam transações pela internet, reunindo múltiplos vendedores e introduzindo modelos inovadores, como "gerador de leads, serviços agregados" e "gerador de pedido." Essa diversidade melhorou a experiência de compra dos consumidores e reforçou a importância central dessas plataformas no comércio eletrônico. (SILVA; RESCH; PEREIRA, 2022).

Atualmente, os consumidores têm acesso a uma ampla variedade de produtos e serviços online, podendo realizar compras a qualquer hora e de qualquer lugar, utilizando dispositivos móveis, tablets e computadores. Além disso, as empresas adotaram estratégias de personalização, oferecendo recomendações personalizadas com base no histórico de compras e preferências dos clientes (PARAISO, 2011).

O comércio eletrônico moderno também se beneficia de avanços na logística, com opções de entrega rápida e eficiente, além de processos simplificados de pagamento, garantias de segurança aprimoradas e uma grande gama de comentários diversificados (NASCIMENTO; SILVA; SANTOS, 2009). Esses avanços têm impulsionado o

crescimento exponencial do comércio eletrônico em todo o mundo, transformando como as pessoas compram e impulsionando o surgimento de novos modelos de negócios online (TOMÉ, 2021).

2.4 CHATGPT

O *ChatGPT* é um recurso de inteligência artificial (IA) desenvolvido pela OpenAI, um renomado laboratório de pesquisa em IA nos Estados Unidos. O termo "GPT" é uma sigla para "*Generative Pre-Trained Transformer*", que pode ser traduzido como "Transformador Pré-Treinado Generativo". Essa tecnologia simula a linguagem humana de forma impressionante, permitindo interações avançadas e naturais (BISWAS, 2023).

Quando falamos em "pré-treinamento" do *ChatGPT*, estamos nos referindo ao processo pelo qual o modelo de linguagem utilizado pelo programa é treinado antecipadamente com uma vasta quantidade de dados de textos e conteúdos (KASNECI *et al.*, 2023). Essa etapa é fundamental, pois permite que o *ChatGPT* compreenda de maneira mais aprofundada a linguagem natural e gere respostas precisas e contextualizadas para as perguntas dos usuários.

3 METODOLOGIA

Nesta seção, apresentaremos a metodologia empregada neste artigo, a qual demonstra como o estudo foi planejado e executado, incluindo as estratégias utilizadas. Este estudo tem uma abordagem predominantemente quantitativa, envolvendo análise e interpretação de dados primários e secundários, com foco em compreender a eficiência do *ChatGPT* na análise de sentimentos

3.1 ESPECIFICAÇÕES DO AMBIENTE

Para esse trabalho o ambiente de execução constitui em uma máquina com as seguintes configurações: *Windows 10* home 64 bits, processador *intel core i5 8th Gen* com placa de vídeo Integrada Intel UHD Graphics 620, memória RAM de *12GB* e armazenamento *SSD* de *450GB*. Para a implementação do modelo foi utilizada a linguagem de programação *Python* na versão 3.10.12 em conjunto com a ferramenta *Colab*, que consiste em uma plataforma de desenvolvimento online desenvolvida pela *Google*.

Nesse trabalho foram utilizadas as seguintes bibliotecas:

1. **Urlopen** (*urllib.request*): A função *urlopen* faz parte do módulo *urllib.request* em *Python* e é usada para abrir URLs. Ela permite fazer solicitações HTTP a servidores web para recuperar dados, como páginas da web, arquivos JSON, imagens, etc. Essa função é frequentemente usada para acessar conteúdo da web.
2. **BeautifulSoup** (*bs4*): é uma biblioteca *Python* muito útil para fazer a análise de documentos HTML e XML. Ela fornece ferramentas para extrair informações de páginas da web de forma estruturada e facilita a navegação e manipulação de elementos HTML, como links, parágrafos, tabelas, etc. É uma escolha popular para a extração de dados de páginas da web.
3. **TextBlob**: é uma biblioteca de processamento de linguagem natural (PLN) em *Python*. Ela oferece uma API simples para tarefas comuns de PLN, como análise de sentimentos, tradução de texto, detecção de idioma, extração de frases-chave e muito mais. É uma ferramenta útil para lidar com texto em várias aplicações.
4. **Google Translate API** (*googletrans*): A biblioteca *googletrans* é uma interface não oficial para a API de Tradução do *Google*. Ela permite a tradução de texto de um idioma para outro usando o serviço de tradução do *Google*. Sendo útil para internacionalizar aplicativos ou para traduzir conteúdo entre idiomas.

5. **Requests:** A biblioteca *requests* é uma ferramenta popular para fazer solicitações HTTP em *Python*. Ela simplifica o envio de solicitações GET, POST e outras solicitações HTTP para servidores web. É amplamente utilizada para interagir com APIs da web e recuperar dados de serviços online.
6. **Openai:** A biblioteca *openai* é uma referência a uma API ou SDK fornecido pela OpenAI, uma empresa de inteligência artificial. A OpenAI oferece APIs que permitem realizar tarefas de processamento de linguagem natural, como geração de texto, tradução, sumarização e muito mais. A biblioteca *openai* pode ser usada para integrar as funcionalidades da OpenAI em aplicativos *Python*.
7. **Pandas:** é uma biblioteca muito popular para manipulação e análise de dados em *Python*. Ela fornece estruturas de dados flexíveis, como *DataFrames*, que facilitam a leitura, limpeza, análise e visualização de dados tabulares. O *Pandas* é frequentemente usado em tarefas de ciência de dados e análise de dados.

Essa configuração serviu perfeitamente para o propósito, sem dificuldades de processamento ou qualquer outro problema. Foi optado por utilizar a plataforma *Coolab* por facilitar o processo de desenvolvimento, já que possui algumas ferramentas pré-instaladas, além de facilitar o processo de instalação de novas ferramentas e fornecer um sistema de organização de códigos eficaz.

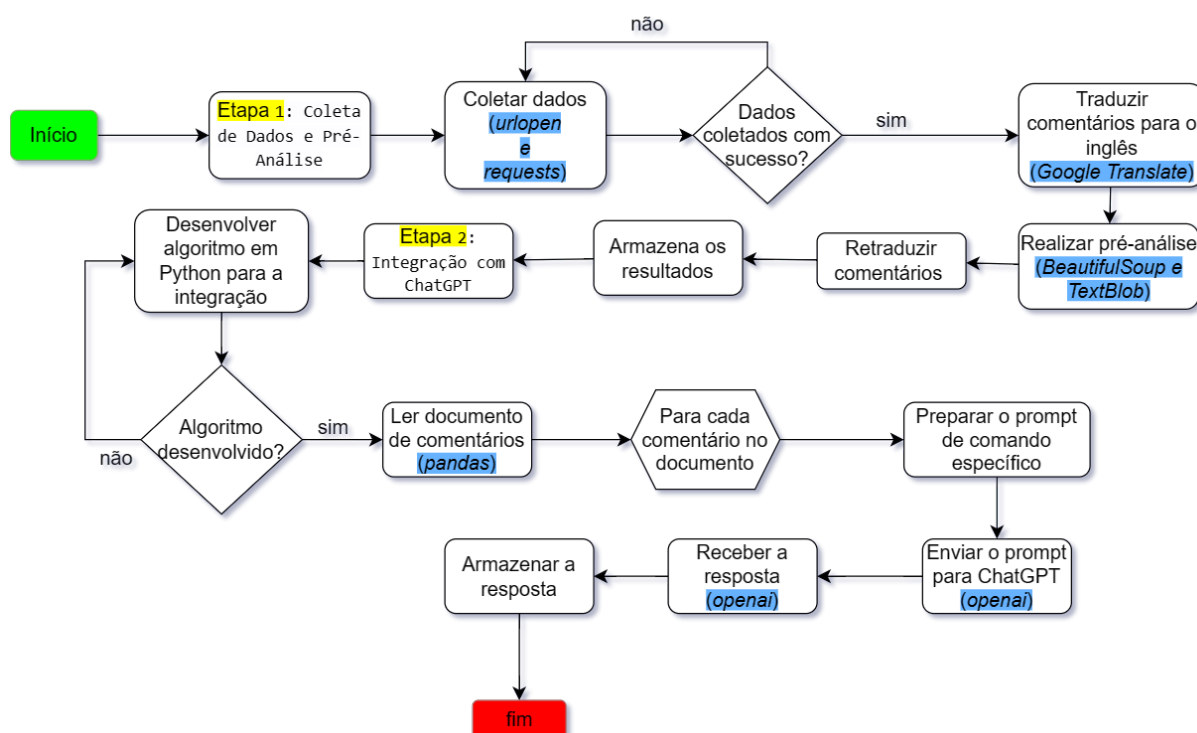
3.2 DESENVOLVIMENTO DO MODELO

O desenvolvimento foi dividido em duas etapas, ambas desenvolvidas na linguagem de programação *Python*. A primeira etapa focou na coleta dos dados, e para isso foram utilizadas as bibliotecas *requests*, *BeautifulSoup* e *urlopen* para a obtenção das informações. Uma pré-análise dos dados foi realizada utilizando as bibliotecas *Pandas*, *TextBlob* e *googletrans*, esta última utilizada para traduzir todos os comentários para o inglês, facilitando o processo de pré-análise, uma vez que a biblioteca *TextBlob* é nativa da língua inglesa e interpreta melhor se os comentários estiverem na mesma. Posteriormente, os comentários foram revertidos para a língua original para facilitar a leitura.

Na segunda etapa, foi realizada a integração com o *ChatGPT*, utilizando a API disponibilizada pela Plataforma da *OpenAI*. Esta etapa consiste em desenvolver um algoritmo capaz de ler o documento com os comentários, selecionar cada um deles e repassá-los para a API com um *prompt* de comando especificando o que deve ser feito.

A figura 1 apresenta como funciona o fluxo das etapas do desenvolvimento.

Figura 1 – Fluxograma do Desenvolvimento



Fonte: Produzido pelo autor

3.3 COLETA DE DADOS

A estratégia utilizada no projeto consiste em coletar informações de diversas fontes de dados por meio do rastreador web *crawler*, empregando na linguagem *python*, e posteriormente organizá-las em um formato estruturado para serem analisadas conforme as necessidades do usuário. Além disso, o algoritmo pode auxiliar na obtenção dos resultados desejados ao permitir a execução de iterações específicas para descartar informações irrelevantes provenientes das URLs de origem (THOMAS; MATHUR, 2019).

É importante que a coleta de dados seja cuidadosamente planejada e executada para garantir a qualidade e a diversidade das informações, possibilitando uma análise mais precisa e completa das emoções e opiniões dos clientes em relação aos produtos ou serviços oferecidos (EL-ANSARI; BENI-HSSANE, 2023). As questões de privacidade e ética foram abordadas de forma a garantir a proteção de dados dos usuários. Nesse processo, os nomes ou qualquer outro atributo que possa revelar a identidade dos usuários não foram coletados, a fim de manter suas identidades de forma anônima. Essa abordagem é consistente com a Lei Geral de Proteção de Dados (LGPD)¹, que

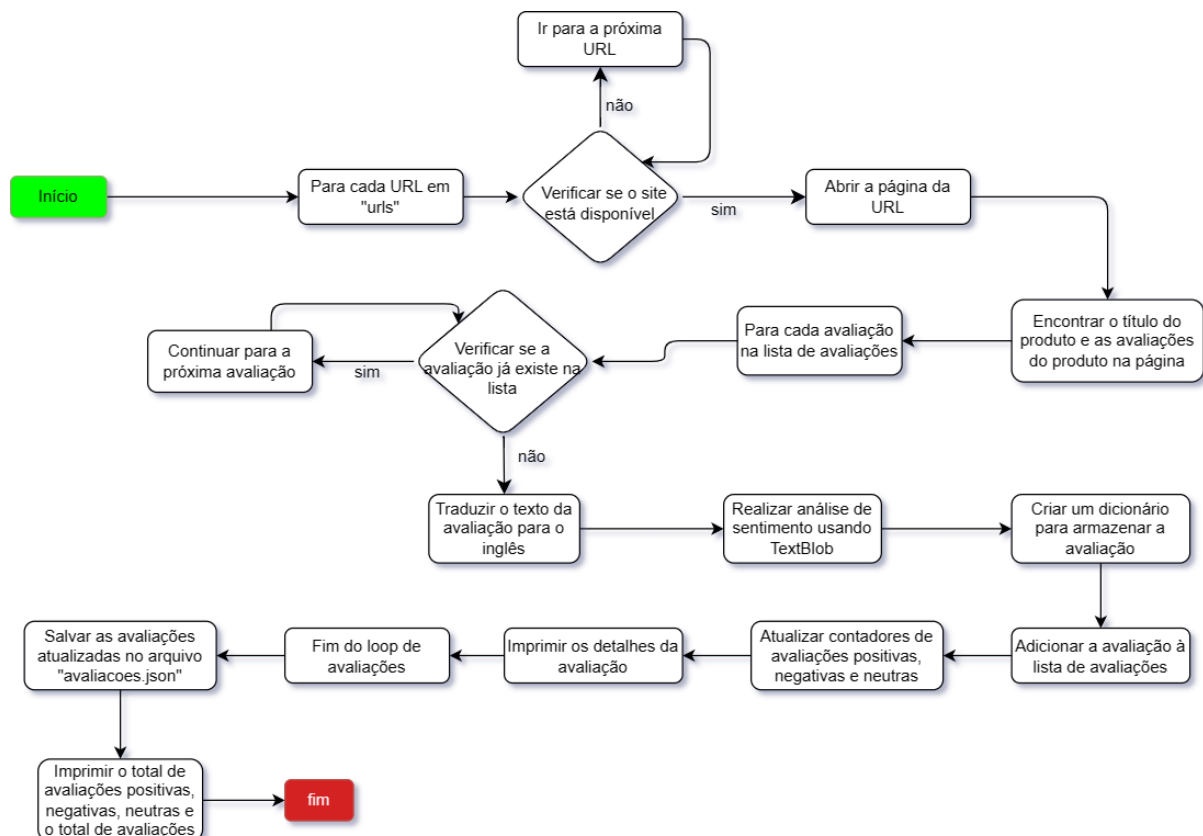
¹ <https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm>

estabelece que os dados pessoais só podem ser coletados com o consentimento do titular e devem ser tratados de forma segura e transparente.

Os sites escolhidos para realizar a coleta de dados foram: **Amazon, Mercado Livre, Magazine Luiza e Americanas**. A escolha desses sites específicos ocorreu após observar e analisar diversos critérios, como a disponibilidade dos dados, a quantidade de informações disponíveis e a ausência de protocolos de defesa específicos contra o *crawler* utilizado. Alguns desses sites apresentam apenas algumas limitações em relação à quantidade de comentários que podem ser obtidos em uma única solicitação, resultando na realização de várias solicitações em dias diferentes para ampliar a diversidade dos comentários. Por outro lado, alguns sites testados demonstraram maior resistência à coleta de dados, chegando até mesmo a bloquear completamente as tentativas. Estes últimos foram descartados após testes que confirmaram a incapacidade de obter os dados desejados.

A figura 2 apresenta a fluxo de funcionamento do algoritmo de coleta de dados

Figura 2 – Fluxograma da coleta de dados



Fonte: Produzido pelo autor

A figura 3 apresenta um exemplo fictício de como os comentários são apresentados nos sites. Na figura são demonstrados como os comentários ficam expostos nos sites, destacando as ferramentas de apoio, como avaliações com estrelas e selos de

verificação, que contribuem para comprovar a eficácia do produto. Neste trabalho serão levados em consideração apenas os comentários.

Figura 3 – Exemplo de Avaliação



Fonte: Produzido pelo autor

3.3.1 Código da Coleta dos Dados

O trecho de código a seguir é responsável por criar um loop para verificar a disponibilidade dos sites. Caso estejam disponíveis, a coleta dos dados é realizada. As linhas iniciadas com "#" são comentários dentro do próprio código, fornecendo explicações mais detalhadas sobre o funcionamento de cada parte.

```
# Cria um loop para coletar as avaliações de cada URL
for url in urls:
    try:
        # Verifica se o site está disponível
        requests.get(url)

    except requests.exceptions.RequestException as e:
        print(f"Erro: e")
        continue

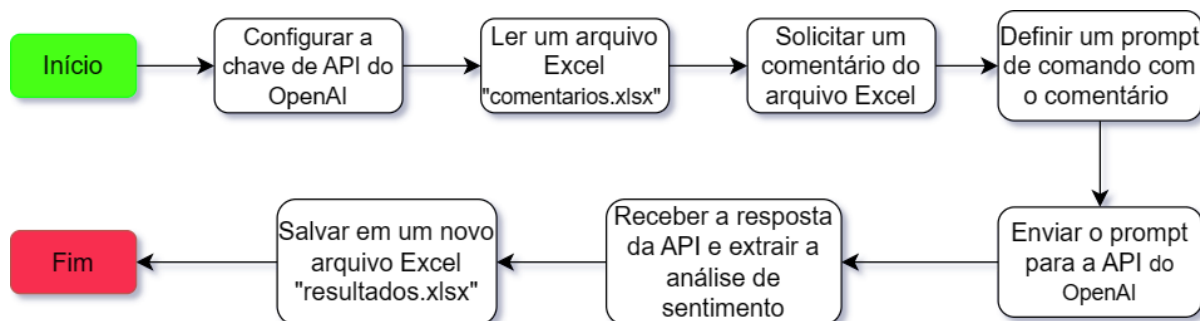
    # Abre a página da URL e coleta o HTML
    html = urlopen(url)
```

3.4 INTEGRAÇÃO COM CHATGPT

A conexão do código com o *ChatGPT* é feita por uma chave de API disponibilizada na plataforma da *OpenAI*. A integração com o *ChatGPT* permite que o sistema reconheça o contexto e o tom da conversa, permitindo uma análise mais precisa e personalizada dos comentários dos usuários (BROWN *et al.*, 2020).

A figura 4 apresenta o fluxo de interação entre o código e a API.

Figura 4 – Interação com a API



Fonte: Produzido pelo autor

O código desenvolvido permite fazer a leitura de um arquivo *excel* contendo os comentários, processo feito utilizando a ferramenta *pandas*. Com o *prompt* certo, especificando como a avaliação deve ser feita é possível ter uma resposta consistente para ser armazenada.

Com essa abordagem, o processo de análise de sentimentos automatizada tem um potencial de se tornar mais escalável, permitindo que as empresas respondam aos comentários dos usuários de forma mais rápida e efetiva (BISWAS, 2023). Além disso, a integração com o *ChatGPT* possibilita que as empresas obtenham *insights* valiosos sobre a experiência do usuário, permitindo que elas ajustem suas estratégias de marketing e aprimorem seus produtos e serviços de acordo com as necessidades e expectativas dos usuários (JAIN *et al.*, 2023).

3.4.1 Prompt

Este tópico apresenta o *prompt* que foi utilizado durante a pesquisa. Ele define como o modelo deve realizar a análise, especificando a escala utilizada e como o resultado deve ser apresentado.

"Por favor, realize uma análise profunda e detalhada dos sentimentos expressos na frase delimitada por {}, classificando-a como positiva, negativa ou neutra e avalie a intensidade de cada sentimento de -1 a 1 sendo que -1 é negativo, 0 é neutro e 1 é positivo. Indique quais palavras-chave na sentença contribuíram para os sentimentos, com base no contexto da frase. Por fim, explique como você chegou a essa conclusão."

3.5 TRABALHOS RELACIONADOS

Na seção atual, abordaremos os trabalhos correlatos e as semelhanças e diferenças com este trabalho.

O autor (ALBUQUERQUE, 2022) tem como principal objetivo desenvolver um modelo de processamento de linguagem natural para análise de sentimento sobre comentários de plataformas de *e-commerce* em português/BR, de maneira automática. O intuito é ajudar as empresas a avaliarem seus processos e tomarem decisões de negócios com base nas opiniões e sentimentos dos consumidores expressos em comentários de *e-commerce*. O autor utilizou uma abordagem baseada em redes neurais BiLSTM, operações de agrupamento e mecanismo de atenção para criar o modelo de análise de sentimento em comentários e *reviews* coletados de diferentes plataformas de *e-commerce*.

(SUDIRJO *et al.*, 2023) visa utilizar o *ChatGPT* em empresas, com o intuito de desenvolver esse sistema e melhorar a análise de sentimento do cliente para empresas, visando entender e atender aos requisitos, tendências e níveis de satisfação do cliente. A abordagem utilizada foi principalmente qualitativa, com uma observação rigorosa e documentação de dados, seguida de técnicas de análise de dados, como redução de dados, visualização e inferência, para extrair resultados significativos.

Em seu trabalho, (BELAL; SHE; WONG, 2023) investigaram a eficácia do *ChatGPT* como uma ferramenta de anotação de texto para tarefas de análise de sentimento. Para isso, realizaram experimentos e avaliações usando conjuntos de dados de referência estabelecidos para tarefas de análise de sentimento. O desempenho do *ChatGPT* foi comparado com outras técnicas de análise de sentimento, incluindo abordagens baseadas em léxico e rotulagem humana. O objetivo final foi demonstrar a superioridade do *ChatGPT* em relação a outras técnicas de análise de sentimento baseadas em léxico e destacar seu potencial como uma ferramenta valiosa para a análise de sentimento em diferentes setores e domínios.

Em seu trabalho a autora (SAMPAIO, 2021) teve como principal objetivo responder a seguinte questão de pesquisa: "As redes sociais contribuíram para a análise de sentimento de que maneira?". Para alcançar esse objetivo, a abordagem utilizada foi uma revisão bibliográfica sobre o tema e um estudo de caso para analisar a contribuição das redes sociais para a análise de sentimentos. O intuito ao desenvolver este trabalho foi realizar estudos sobre as redes sociais e o uso das suas fontes textuais para identificar opiniões e sentimentos, além de realizar estudos sobre métodos para fazer análise de sentimentos nas redes sociais.

Este estudo se assemelha aos trabalhos de (SUDIRJO *et al.*, 2023) e (ALBUQUERQUE, 2022) no uso do *ChatGPT* como ferramenta de análise e na concentração na análise do comércio eletrônico. No entanto, difere desses trabalhos no método de busca e extração de comentários, que envolve o uso de um *Web crawler*. Este método desempenha um papel crucial no desenvolvimento deste estudo.

3.5.1 Tabela de Relacionados

A Tabela 1 a seguir destacará os principais artigos relacionados, juntamente com suas ênfases e distinções em relação ao estudo proposto neste trabalho.

Tabela 1 – Trabalhos relacionados

Referência	Título	Ano	Objetivo
Sampaio. (SAMPAIO, 2021)	ANÁLISE DE SENTIMENTOS	2021	Apresentação de uma visão geral sobre a área de análise de sentimentos e como ela pode ser aplicada na sociedade, destacando também a importância das mídias sociais como fonte de informações.
Da Silva Albuquerque. (ALBUQUERQUE, 2022)	Análise de sentimento sobre comentários em sites de e-commerce no idioma Português/BR	2022	Utilizou abordagem baseada em redes neurais BiLSTM para análise de sentimento em comentários de e-commerce.
Sudirjo <i>et al.</i> (SUDIRJO <i>et al.</i> , 2023)	Application of ChatGPT in Improving Customer Sentiment Analysis for Businesses	2023	Fez o uso do ChatGPT, para melhorar a análise de sentimento do cliente de empresas com o objetivo de entender e atender aos requisitos, tendências e níveis de satisfação do cliente.
Belal <i>et al.</i> (BELAL; SHE; WONG, 2023)	Leveraging ChatGPT As Text Annotation Tool For Sentiment Analysis	2023	Realizou uma comparação do <i>ChatGPT</i> com outras ferramentas de análise de sentimentos, com o objetivo de avaliar a eficácia do mesmo para essa tarefa.
Este Trabalho	WebCrawler integrado ao ChatGPT para Análise de sentimentos de clientes em e-commerces	2023	Realiza uma integração de um WebCrawler com a API do ChatGPT personalizando a análise dos comentários dos usuários de E-commerces.

4 RESULTADOS

Esta seção apresenta os resultados obtidos neste trabalho.

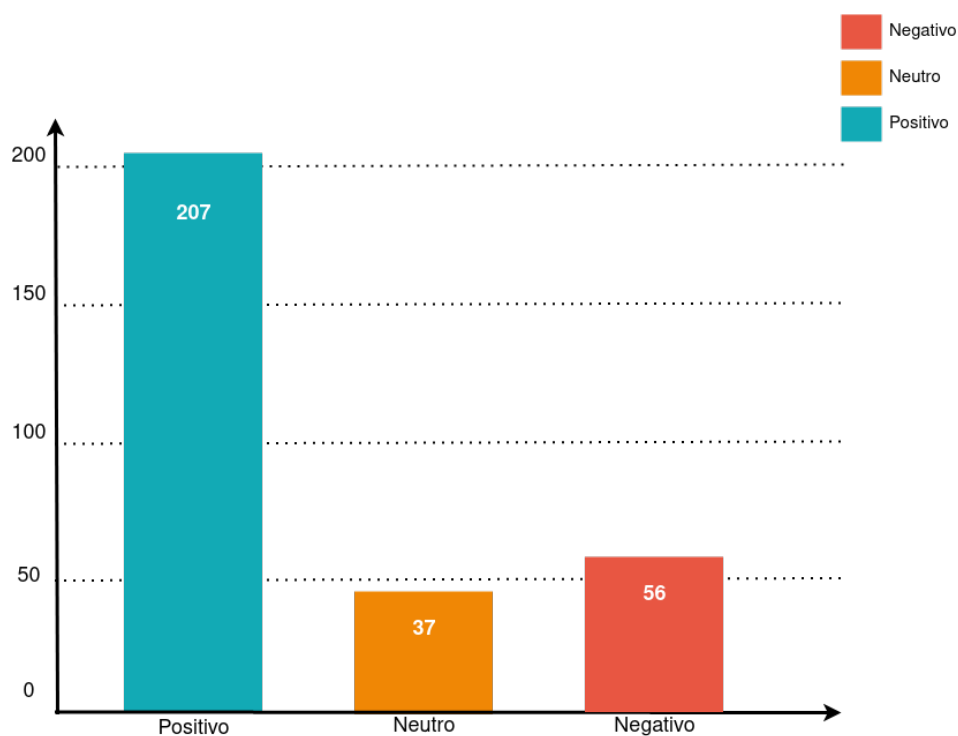
4.1 RESULTADOS DOS COMENTÁRIOS

Foi adotada uma escala semelhante a descrita em (BENEVENUTO; RIBEIRO; ARAÚJO, 2015), que varia de 1 a -1 para a classificação dos comentários. Comentários com valores no intervalo de 0.1 a 1 foram considerados positivos, aqueles entre -0.1 a -1 foram classificados como negativos, enquanto os que obtiveram uma classificação de 0 foram categorizados como neutros.

Foram coletados 300 comentários dos quatro sites selecionados para análise durante um mês. Ao final da coleta a divisão se deu da seguinte forma: 207 positivos, 56 negativos e 37 neutros. Essa escolha baseou-se na relevância para a pesquisa, uma vez que essa quantidade de dados nos proporciona uma ampla variedade de comentários diversificados.

A figura 5 apresenta a quantidade de cada comentário positivo, negativo e neutro.

Figura 5 – Percentual dos comentários



Fonte: Produzido pelo autor

4.2 EXEMPLOS DE COMENTÁRIOS

Nesta seção, exemplificamos diferentes tipos de comentários e suas análises correspondentes geradas pelo ChatGpt, proporcionando uma compreensão prática do processo de análise de sentimentos.

4.2.1 Positivo

Comentário:

"Este é o melhor livro de Daniel Goleman. Muitos outros são bons, mas este é sem dúvida, para mim, o mais bem elaborado e que cujo tema é muito útil."

Análise:

"Sentimento: Positivo;
Intensidade: 0.9;
Palavras-chave que contribuíram para o sentimento positivo: melhor, bons, bem elaborado, útil.

Explicação: A análise da frase indicada mostra que a sentença expressa um sentimento positivo. Isso pode ser observado pela presença de palavras-chave como "melhor", "bons", "bem elaborado", "útil". Essas palavras têm conotação positiva e indicam uma apreciação e satisfação em relação ao livro de Daniel Goleman. Além disso, o uso de advérbios intensificadores como "mais" e "muito" também contribui para a intensidade positiva. Portanto, com base nas palavras-chave e no contexto da frase, a análise de sentimentos classificou o sentimento como positivo, com uma intensidade de 0.9."

4.2.2 Neutro

Comentário:

"Tema mais científico e menos entretenimento."

Análise:

"Sentimento: Neutro; Intensidade: 0;
Palavras-chave contribuintes: científico, entretenimento.

Explicação: A frase "Tema mais científico e menos entretenimento" expressa uma atitude neutra. Os adjetivos "científico" e "entretenimento" são relevantes para a análise dos sentimentos expressos na frase. Eles sugerem uma preferência por um tema mais orientado para a ciência em vez de entretenimento. No entanto, a frase em si não indica claramente se essa preferência é positiva ou negativa, resultando em um sentimento neutro. Portanto, a intensidade é avaliada como 0, indicando neutralidade. "

4.2.3 Negativo

Comentário:

"Na descrição de funcionalidade e autonomia da pulseira consta 14 dias de duração da carga porém, a mesma está durando aproximadamente 03hs. Ressalto que as funções descritas com alto consumo de bateria, informado no app mi fit não foram acionadas."

Análise:

"A análise dos sentimentos expressos na frase é a seguinte:

- Sentimento: negativo; - Intensidade: -0.8;

Palavras-chave que contribuíram para o sentimento negativo: funcionamento, autonomia, carga, durando, consumo de bateria, não foram acionadas.

Explicação: A análise foi feita identificando as palavras-chave relacionadas a funcionalidade, autonomia e carga. Nesse caso, a expectativa era de que a pulseira tivesse uma duração de carga de 14 dias, porém está durando apenas cerca de 03hs. Isso representa uma diferença significativa e negativa, resultando em um sentimento negativo.

Além disso, ao mencionar que as funções que têm alto consumo de bateria não foram acionadas, implica que a queda na duração da carga não pode ser atribuída ao uso dessas funções, o que pode aumentar a frustração e contribuir para um sentimento negativo ainda maior.

Sendo assim, com base nas palavras-chave relacionadas ao funcionamento e autonomia, e considerando o contexto da frase, foi possível determinar que o sentimento expresso é negativo com uma intensidade de -0.8. "

4.2.4 Matriz de Confusão

A Figura 6 representa a Matriz de Confusão¹, que ilustra a relação entre os resultados de referência (comentários) e os classificados pela análise do modelo. A matriz destaca a divisão entre os comentários corretamente classificados (em azul) e aqueles incorretamente classificados (em laranja). A Matriz de Confusão é uma ferramenta crucial em problemas de classificação, proporcionando a atribuição de rótulos ou categorias a instâncias de dados (HEYDARIAN; DOYLE; SAMAVI, 2022).

¹ <<https://cienciaenegocios.com/o-que-e-a-matriz-de-confusao/>>

Figura 6 – Matriz de Confusão



Fonte: Produzido pelo autor

4.2.5 Métricas de Avaliação

A Tabela 2 resume os resultados das análises realizadas pelo ChatGPT, destacando a quantidade de comentários corretos e incorretos para cada tipo (positivo, neutro, negativo). Os valores de acurácia, precisão média, *recall* e *F1-Score* são apresentados para avaliar o desempenho geral do modelo na classificação de sentimentos. Essas métricas fornecem uma visão abrangente da eficácia do modelo em lidar com diferentes tipos de comentários.

Tabela 2 – Métricas

Métricas	Resultados
Total	300
Corretos	280
Incorretos	20
Acurácia	93.3%
Precisão	86%
Recall	88.18%
F1-Score	97.1%

Fonte: Produzido pelo autor

Segundo (STRAUSS; JÚNIOR; FERREIRA, 2022):

- **Acurácia** é uma métrica de avaliação de modelos de classificação que mede a proporção de instâncias que foram classificadas corretamente. No contexto

de classificação binária, a acurácia é calculada como a proporção de instâncias positivas e negativas que foram classificadas corretamente.

- **Precisão** é outra métrica de avaliação de modelos de classificação que mede a proporção de instâncias positivas que foram classificadas corretamente. No contexto de classificação binária, a precisão é calculada como a proporção de instâncias positivas que foram classificadas como positivas.
- **Recall** é outra métrica de avaliação de modelos de classificação que mede a proporção de instâncias positivas que foram detectadas. No contexto de classificação binária, o *recall* é calculado como a proporção de instâncias positivas que foram classificadas como positivas ou negativas.
- **F1-score** é uma métrica de avaliação de modelos de classificação que combina precisão e *recall* em uma única medida. O *F1-score* é calculado como a média harmônica de precisão e *recall*.

Os cálculos referentes à acurácia, precisão, *recall* e *F1-score* seguem a seguinte ordem: os valores para as categorias positivo, neutro e negativo são apresentados, respectivamente, abaixo:

$$\textbf{Acurácia} = ((204 + 26 + 50) / (300)) * 100 = \textbf{93.3\%}$$

$$\textbf{Precisão} = (204/207) + (26/37) + (50/56) = (98.5\% + 70.3\% + 89.3\%)/3 = \textbf{86\%}$$

$$\textbf{Recall} = (204/210) + (26/32) + (50/58) = (97.1\% + 81.25\% + 86.2\%)/3 = \textbf{88,18\%}$$

$$\textbf{F1-Score} = (2 * \text{Precisão} * \text{Recall}) / (\text{Precisão} + \text{Recall}) = \textbf{97.1\%}$$

4.2.6 Apresentação dos resultados

A ferramenta desenvolvida para análise de sentimentos, integrada à API do ChatGPT, possui diversas aplicações práticas que podem beneficiar empresas e usuários no ambiente de *e-commerce*. Abaixo estão alguns exemplos de como os resultados obtidos podem ser aplicados:

A tabela 3 apresenta os resultados obtidos pela pesquisa.

Tabela 3 – Resultados

Resultado	Descrição
Aprimoramento de Produtos e Serviços	A análise de sentimentos pode fornecer <i>insights</i> valiosos sobre as preferências e opiniões dos clientes em relação a produtos específicos. Empresas podem usar esses dados para realizar melhorias contínuas, ajustando recursos, design ou funcionalidades com base no <i>feedback</i> dos usuários.
Compatibilidade com a Literatura Existente	A ferramenta segue as práticas e avanços da análise de sentimentos descritos na literatura.
Personalização da Experiência do Cliente	Compreender os sentimentos expressos pelos clientes permite personalizar interações e ofertas. Empresas podem adaptar estratégias de marketing, recomendações de produtos e promoções com base nas preferências individuais, aumentando a satisfação do cliente.
Gestão de Reputação Online	A ferramenta pode ser empregada na monitorização da reputação online da empresa. Identificando e classificando comentários positivos e negativos, as organizações podem responder de maneira proativa a <i>feedbacks</i> negativos, mitigando possíveis impactos negativos na reputação da marca.
Tomada de Decisões Estratégicas	Ao analisar os sentimentos expressos em larga escala, empresas podem embasar decisões estratégicas. Isso inclui a identificação de tendências de mercado, avaliação de campanhas de marketing e ajuste de estratégias de comunicação para alinhar-se melhor às expectativas do público.
Monitoramento Contínuo de Satisfação do Cliente	Implementar a análise de sentimentos de forma contínua permite às empresas monitorar a satisfação do cliente em tempo real. Com essa abordagem, é possível identificar mudanças nas percepções dos clientes imediatamente e responder rapidamente a problemas emergentes.
Identificação de Problemas e Oportunidades	A detecção de padrões nos sentimentos dos clientes pode revelar problemas específicos que precisam de atenção imediata, bem como oportunidades de inovação. Identificar áreas de insatisfação ou demandas não atendidas pode orientar ações para melhorar a oferta de produtos e serviços.

5 CONCLUSÃO

Os resultados obtidos indicam que o *ChatGPT* pode ser considerado como uma alternativa viável como uma ferramenta para análise de sentimentos, especialmente ao fornecer serviços por meio de uma API. Durante a avaliação da ferramenta, o serviço ofereceu dados precisos para a compreensão da experiência dos usuários em sistemas de comércio eletrônico. Com base neste estudo, conclui-se que a técnica de interação com a API do ChatGPT pode ser eficiente na tarefa de análise dos sentimentos de comentários de *marketplace*. Os resultados deste trabalho mostram que, dos 300 comentários selecionados, obteve-se 93,3% de acurácia, resultados semelhantes aos de (BELAL; SHE; WONG, 2023), confirmando que com essa técnica as empresas podem identificar problemas e tendências de mercado, permitindo que se adaptem às necessidades dos cliente e melhorem a satisfação e fidelidade dos mesmos. No entanto, ainda existem algumas limitações, como o uso restrito da API (foi utilizado apenas o pacote gratuito disponibilizado pela plataforma) e a falta de mais bases de dados (sites) para análise.

5.1 TRABALHOS FUTUROS

Pesquisas futuras podem explorar o aprimoramento da metodologia usada neste trabalho, bem como expandir a análise para outras plataformas de *marketplace*. Além disso, é possível investigar melhorias no algoritmo de análise de sentimentos e considerar o uso de outras Inteligências Artificiais para comparação de desempenho entre elas. Sugiro incluir também a avaliação do *Iramuteq* e *Atlas.ti* como ferramentas potenciais para a análise de dados qualitativos, enriquecendo a abordagem metodológica.

REFERÊNCIAS

- ALBUQUERQUE, C. L. D. S. Análise de sentimento sobre comentários em sites de e-commerce no idioma português/br fortaleza 2022. 2022. Citado 3 vezes nas páginas 7, 15 e 16.
- BAE, M.; KIM, H. J. Web crawler-based approaches for data-driven tourism research. **Tourism Management Perspectives**, Elsevier, v. 35, p. 100686, 2020. Citado na página 4.
- BELAL, M.; SHE, J.; WONG, S. Leveraging chatgpt as text annotation tool for sentiment analysis. **arXiv preprint arXiv:2306.17177**, 2023. Citado 3 vezes nas páginas 15, 16 e 23.
- BENEVENUTO, F.; RIBEIRO, F.; ARAÚJO, M. Métodos para análise de sentimentos em mídias sociais. **Sociedade Brasileira de Computação**, 2015. Citado na página 17.
- BISWAS, S. The function of chat gpt in social media: According to chat gpt. **Available at SSRN 4405389**, 2023. Citado 3 vezes nas páginas 4, 8 e 14.
- BOUDIN, F.; LIGOZAT, A.-L. Deep learning for sentiment analysis: A survey. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, Wiley Online Library, v. 8, n. 4, p. e1253, 2018. Citado na página 4.
- BROWN, T. B. *et al.* Language models are few-shot learners. **arXiv preprint arXiv:2005.14165**, 2020. Citado 2 vezes nas páginas 4 e 13.
- CAO, H. *et al.* An empirical study on the applications of sentiment analysis in education. **Computers & Education**, Elsevier, v. 149, p. 103814, 2020. Citado 2 vezes nas páginas 4 e 6.
- CHEEMA, S. M.; TARIQ, S.; PIRES, I. M. Natural language interface for automatic generation of data flow diagram. **J. King Saud Univ. - Comput. Inf. Sci.**, v. 35, n. 2, p. 626–640, 2023. ISSN 1319-1578. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S131915782300006X>>. Citado na página 6.
- COELHO, M. R. S. **Web crawlers para netnografia: uma proposta para análise e mapeamento de redes sociais**. 2022. Disponível em: <<https://bdm.ufpa.br:8443/jspui/handle/prefix/4219>>. Trabalho de Conclusão de Curso (Tecnólogo em Geoprocessamento) – Campus Universitário de Ananindeua, Universidade Federal do Pará, Ananindeua, 2022. Citado na página 6.
- EL-ANSARI, A.; BENI-HSSANE, A. Sentiment analysis for personalized chatbots in e-commerce applications. **Wireless Personal Communications**, Springer, p. 1–22, 2023. Citado 2 vezes nas páginas 4 e 11.
- GIL, S. R. P. Fatores críticos de sucesso na internacionalização de pmes através de marketplaces de referência mundial. 2022. Citado na página 7.

GONZALEZ, M.; LIMA, V. L. S. Recuperação de informação e processamento da linguagem natural. In: **XXIII Congresso da Sociedade Brasileira de Computação**. [S.l.: s.n.], 2003. v. 3, p. 347–395. Citado na página 7.

GUNASEKARAN, A. *et al.* E-commerce and its impact on operations management. **International journal of production economics**, Elsevier, v. 75, n. 1-2, p. 185–197, 2002. Citado na página 7.

GUO, K. *et al.* Sentiment analysis on chinese social media platforms: A survey. **Journal of Information Science**, SAGE Publications Sage UK: London, England, v. 45, n. 1, p. 26–41, 2019. Citado na página 6.

HEYDARIAN, M.; DOYLE, T. E.; SAMAVI, R. Mlcm: Multi-label confusion matrix. **IEEE Access**, IEEE, v. 10, p. 19083–19095, 2022. Citado na página 19.

JAIN, V. *et al.* The prospects and challenges of chatgpt on marketing research and practices. **Emmanuel, The Prospects and Challenges of ChatGPT on Marketing Research and Practices (March 23, 2023)**, 2023. Citado na página 14.

KASNECI, E. *et al.* Chatgpt for good? on opportunities and challenges of large language models for education. **Learning and Individual Differences**, Elsevier, v. 103, p. 102274, 2023. Citado 2 vezes nas páginas 4 e 8.

LIU, B. **Sentiment analysis and opinion mining**. [S.l.]: Morgan & Claypool Publishers, 2012. Citado 2 vezes nas páginas 4 e 6.

NASCIMENTO, A. R. d.; SILVA, B. F. d.; SANTOS, G. G. d. E-commerce: o melhor caminho no mercado atual. 2009. Citado na página 7.

PARAISO, G. J. B. O e-commerce nas redes sociais: estudo sobre os desdobramentos do comércio eletrônico na atualidade. **Undergraduate thesis (Bachelor of Business Administration)–Faculdade de Ciências da Administração, Universidade de Pernambuco, Recife, Pernambuco**, 2011. Citado na página 7.

RAMOS, B.; FREITAS, C. Sentimento de quê?” uma lista de sentimentos para a análise de sentimentos. **STIL**, p. 15–18, 2019. Citado na página 7.

RAMOS, E. **E-commerce**. [S.l.]: Editora FGV, 2015. Citado na página 7.

SAMPAIO, A. G. Análise de sentimentos. Pontifícia Universidade Católica de Goiás, 2021. Citado 2 vezes nas páginas 15 e 16.

SILVA, J. do N.; RESCH, S.; PEREIRA, J. A. Marketplaces. **Encontro Internacional de Gestão, Desenvolvimento e Inovação (EIGEDIN)**, v. 6, n. 1, 2022. Citado na página 7.

STRAUSS, E.; JÚNIOR, M. V. B.; FERREIRA, W. L. L. A importância de utilizar métricas adequadas de avaliação de performance em modelos preditivos de machine learning. **Projectus**, v. 7, n. 2, p. 52–62, 2022. Citado na página 20.

SUDIRJO, F. *et al.* Application of chatgpt in improving customer sentiment analysis for businesses. **Jurnal Teknologi Dan Sistem Informasi Bisnis**, v. 5, n. 3, p. 283–288, 2023. Citado 2 vezes nas páginas 15 e 16.

THOMAS, D. M.; MATHUR, S. Análise de dados por web scraping usando python. In: **2019 3ª Conferência Internacional de Eletrônica, Comunicação e Tecnologia Aeroespacial (ICECA)**. [S.l.: s.n.], 2019. p. 450–454. Citado na página 11.

TOMÉ, L. M. Comércio eletrônico. Banco do Nordeste do Brasil, Fortaleza, v. 6, n. 205, dezembro 2021. (Caderno Setorial Etene). Citado na página 8.