

INSTITUTO FEDERAL DO ESPÍRITO SANTO
SISTEMAS DE INFORMAÇÃO

WILLIAM DESTEFFANI SOARES

**AVALIAÇÃO DE PRODUTOS BASEADA EM ANÁLISE DE SENTIMENTOS
APLICADA A POSTAGENS TEXTUAIS EM REDES SOCIAIS**

Cachoeiro de Itapemirim

2023

WILLIAM DESTEFFANI SOARES

**AVALIAÇÃO DE PRODUTOS BASEADA EM ANÁLISE DE SENTIMENTOS
APLICADA A POSTAGENS TEXTUAIS EM REDES SOCIAIS**

Trabalho de Conclusão de Curso apresentado à Coordenadoria do Curso de Sistemas de Informação do Instituto Federal do Espírito Santo, Campus Cachoeiro de Itapemirim, como requisito parcial para a obtenção do título de Bacharel em Sistemas de Informação.

Orientador: Prof. Dr. Ricardo Maroquio Bernardo

Cachoeiro de Itapemirim

2023

(Biblioteca do Campus Cachoeiro de Itapemirim)

S676a Soares, William Desteffani.

Avaliação de produtos baseada em análise de sentimentos aplicada a postagens textuais em redes sociais / William Desteffani Soares. - 2023.
114 f. : il. ; 30 cm.

Orientador: Ricardo Maroquio Bernardo

TCC (Graduação) Instituto Federal do Espírito Santo, Campus Cachoeiro de Itapemirim, Sistemas de Informação, 2023.

1. Inteligência artificial. 2. Linguagem de programação - computadores. 3. Aprendizado por computador. 4. Comércio eletrônico. 5. Redes sociais on line. I. Bernardo, Ricardo Maroquio. II. Título III. Instituto Federal do Espírito Santo.

CDD: 006.3

Bibliotecário/a: Jacqueline Machado Silva CRB-ES nº 640



**MINISTÉRIO DA EDUCAÇÃO
INSTITUTO FEDERAL DO ESPÍRITO SANTO
CAI - COORDENADORIA DO CURSO DE BACHARELADO
EM SISTEMAS DE INFORMACAO**



FOLHA DE APROVAÇÃO-TCC N° 19 / 2023 - CAI-CCSI (11.02.18.01.08.02.13)

N° do Protocolo: 23151.004355/2023-44

Cachoeiro De Itapemirim-ES, 15 de dezembro de 2023.

WILLIAM DESTEFFANI SOARES

**AVALIAÇÃO DE PRODUTOS BASEADA EM ANÁLISE DE SENTIMENTOS
APLICADA A POSTAGENS TEXTUAIS EM REDES SOCIAIS**

Trabalho de Conclusão de Curso apresentado à Coordenadoria do Curso de Sistemas de Informação do Instituto Federal do Espírito Santo, Campus Cachoeiro de Itapemirim, como requisito parcial para a obtenção do título de Bacharel em Sistemas de Informação.

Orientador: Prof. Dr. Ricardo Maroquio Bernardo

Aprovado em 11 de dezembro de 2023

COMISSÃO EXAMINADORA

Prof. Dr. Ricardo Maroquio Bernardo
Instituto Federal Do Espírito Santo
Orientador

Profa. Dra. Susana Brunoro Costa de Oliveira
Instituto Federal Do Espírito Santo

Prof. Dr. Rafael Vargas Mesquita dos Santos
Instituto Federal Do Espírito Santo

(Assinado digitalmente em 15/12/2023 13:02)
RAFAEL VARGAS MESQUITA DOS SANTOS
PROFESSOR DO ENSINO BASICO TECNICO E TECNOLOGICO
CAI-CCSI (11.02.18.01.08.02.13)
Matricula: 1544937

(Assinado digitalmente em 15/12/2023 12:28)
RICARDO MAROQUIO BERNARDO
PROFESSOR DO ENSINO BASICO TECNICO E TECNOLOGICO
CAI-CCSI (11.02.18.01.08.02.13)
Matricula: 2152606

(Assinado digitalmente em 15/12/2023 12:47)
SUSANA BRUNORO COSTA DE OLIVEIRA
PROFESSOR DO ENSINO BASICO TECNICO E TECNOLOGICO
CAI-CCTI (11.02.18.01.08.02.07)
Matricula: 1505999

Visualize o documento original em <https://sipac.ifes.edu.br/public/documentos/index.jsp> informando seu número: **19**, ano: **2023**, tipo: **FOLHA DE APROVAÇÃO-TCC**, data de emissão: **15/12/2023** e o código de verificação: **d2745ec559**

DECLARAÇÃO DO AUTOR

Declaro, para fins de pesquisa acadêmica, didática e técnico-científica, que este Trabalho de Conclusão de Curso pode ser parcialmente utilizado, desde que se faça referência à fonte e ao autor.

Cachoeiro de Itapemirim, 11 de Dezembro de 2023.

William Desteffani Soares

Dedico este trabalho, em profunda gratidão, a Deus, fonte inesgotável de inspiração e guia em minha jornada acadêmica. A todos que, com generosidade e apoio, contribuíram para a concretização deste trabalho, meu sincero agradecimento. Suas contribuições foram essenciais, e este projeto não seria possível sem a presença de cada um de vocês em meu caminho. Que este trabalho seja não apenas uma expressão do meu esforço, mas também um reflexo da força da união e colaboração. Muito obrigado a todos que tornaram este momento possível.

AGRADECIMENTOS

Gostaria de expressar minha sincera gratidão a todas as pessoas que contribuíram para a realização deste Trabalho de Conclusão de Curso. Este projeto representa não apenas o meu esforço, mas também a colaboração e o apoio de muitas pessoas incríveis.

Em primeiro lugar, agradeço ao meu orientador, Ricardo Maroquio Bernardo, pela orientação, paciência e valiosas sugestões que foram essenciais para o desenvolvimento deste trabalho. Sua expertise e dedicação foram fundamentais para o meu crescimento acadêmico.

À minha família e amigos, meu agradecimento especial. Obrigado por serem meu alicerce, por compreenderem as ausências e por oferecerem apoio incondicional durante todo o processo. Vocês são a fonte do meu sucesso e da minha perseverança.

Aos professores e colegas de curso, agradeço por compartilharem conhecimentos, desafios e conquistas. Cada interação moldou minha jornada acadêmica de maneira única.

Por último, mas não menos importante, agradeço a todos que, direta ou indiretamente, contribuíram para este trabalho. Cada palavra de incentivo e apoio foi significativa.

Este trabalho é dedicado a todos vocês. Muito obrigado por fazerem parte desta conquista.

"A persistência é o caminho do êxito."
Charles Chaplin

RESUMO

A análise de sentimentos tem sido utilizada em uma variedade de aplicações, incluindo a análise do impacto de eventos em redes sociais online (OSNs) e o resumo da opinião pública sobre produtos e marcas em discussões nesses sistemas. Esse recurso ensina os computadores a reconhecer a emoção humana presente em diferentes tipos de texto, porém, o maior desafio é equilibrar simplicidade e precisão. Os métodos variam desde o uso de listas de palavras relacionadas à emoção até o aprendizado profundo e técnicas como incorporação de palavras, redes neurais e mecanismos de atenção. Ao longo dos últimos anos, a análise de sentimentos tem sido usada em um número crescente de aplicações, portanto, o presente trabalho tem por objetivo identificar e classificar emoções expressadas em textos, mensagens ou discursos de feedback público sobre marcas e produtos registrados em redes sociais, a fim de auxiliar empresas, especialmente aquelas do setor de e-commerce, a proporcionar melhores estratégias de vendas, bem como facilitar a tomada de decisões pelos consumidores antes de efetuarem compras, e, também oferecer às fabricantes a oportunidade de identificar tendências de mercado e fabricar produtos personalizados de acordo com as demandas identificadas.

Palavras-chave: Aprendizado de Máquina, Inteligência Artificial, Processamento de Linguagem Natural, Análise de Sentimentos, Feedbacks de Produtos, Avaliação de Produtos.

ABSTRACT

Sentiment analysis has been employed in a variety of applications, including assessing the impact of events on online social networks (OSNs) and summarizing public opinion on products and brands within these systems. This feature teaches computers to recognize human emotion present in different types of text, but the major challenge is striking a balance between simplicity and accuracy. Methods range from using lists of emotion-related words to deep learning and techniques such as word embedding, neural networks, and attention mechanisms. Over the past few years, sentiment analysis has been utilized in a growing number of applications. Thus, this paper aims to identify and classify emotions expressed in texts, messages, or public feedback speeches about brands and products recorded on social networks. Its purpose is to assist businesses, especially those in the e-commerce sector, in providing better sales strategies and facilitating consumer decision-making before making purchases. Additionally, it aims to offer manufacturers the opportunity to identify market trends and manufacture customized products according to identified demands.

Keywords: Machine Learning, Artificial Intelligence, Natural Language Processing, Sentiment Analysis, Product Feedback, Product Evaluation.

LISTA DE FIGURAS

Figura 1 – Categorização de Elementos de uma Frase Utilizando o Método Simbólico da NLP - Estado 1	52
Figura 2 – Categorização de Elementos de uma Frase Utilizando o Método Simbólico da NLP - Estado 2	53
Figura 3 – Categorização de Elementos de uma Frase Utilizando o Método Simbólico da NLP - Estado 3	53
Figura 4 – Categorização de Elementos de uma Frase Utilizando o Método Simbólico da NLP - Estado 4	53
Figura 5 – Categorização de Elementos de uma Frase Utilizando o Método Simbólico da NLP - Estado 5	54
Figura 6 – Classificação de Elementos Frasais por Meio do Método Estatístico da NLP - Estado 1	55
Figura 7 – Classificação de Elementos Frasais por Meio do Método Estatístico da NLP - Estado 4	59
Figura 8 – <i>N-grams</i>	65
Figura 9 – <i>Named Entity Recognition (NER)</i>	70
Figura 10 – Modelo IMDB-PT	80
Figura 11 – Gráfico BERT	82
Figura 12 – Curva ROC	95
Figura 13 – Acurácia	95
Figura 14 – Formulário de Pesquisa	108
Figura 15 – Resultados Gerais de Pesquisa para o Termo "Notebook Positivo" .	109
Figura 16 – Resultados Específicos de um Vídeo de Pesquisa para o Termo "Notebook Positivo"	109
Figura 17 – Resultados Gerais de Pesquisa para o Termo "Notebook Asus" . . .	110
Figura 18 – Resultados Específicos de um Vídeo de Pesquisa para o Termo "Notebook Asus"	110

LISTA DE TABELAS

Tabela 1 – Classificação de Elementos Frasais por Meio do Método Estatístico da NLP - Estado 2	56
Tabela 2 – Classificação de Elementos Frasais por Meio do Método Estatístico da NLP - Estado 3	58

SUMÁRIO

1	INTRODUÇÃO	19
2	TRABALHOS RELACIONADOS	23
3	FUNDAMENTAÇÃO TEÓRICA	27
3.1	Redes Sociais e sua Importância	27
3.2	Análise de Sentimentos	31
3.2.1	Importância da análise de sentimentos	32
3.2.2	Casos de uso de análise de sentimentos	34
3.2.3	Funcionamento da análise de sentimentos	36
3.2.4	Abordagens para a análise de sentimentos	39
3.2.4.1	Com base em regras (palavras-chave e afinidade léxica)	39
3.2.4.2	Aprendizado de máquina/ML	40
3.2.4.3	Orientação semântica	40
3.2.4.4	Baseado em conceitos	41
3.2.5	Diferentes tipos de análise de sentimentos	41
3.2.5.1	Pontuação refinada	41
3.2.5.2	Com base em aspectos	41
3.2.5.3	Com base em intenções	42
3.2.5.4	Detecção emocional	42
3.2.6	Desafios na análise de sentimentos	42
3.2.6.1	Sentimentos implícitos e sarcasmo	42
3.2.6.2	Dependência de domínio	43
3.2.6.3	Expectativas frustradas	43
3.2.6.4	Pragmática	43
3.2.6.5	Conhecimento de mundo	44
3.2.6.6	Detecção de subjetividade	44
3.2.6.7	Identificação de entidade	45
3.2.6.8	Negação	45
3.2.7	Análise semântica	45
3.2.7.1	Análise de sentimento versus análise semântica	46
3.3	Processamento de Linguagem Natural	47

3.3.1	Como funciona o NLP	48
3.3.1.1	Pré-processamentos	48
3.3.1.2	Processamentos	49
3.3.2	Pirâmide do Processamento de Linguagem Natural	50
3.3.2.1	Pragmática	50
3.3.2.2	Semântica	51
3.3.2.3	Sintaxe	51
3.3.2.4	Morfologia	51
3.3.3	Os 4 tipos de abordagem do processamento NLP	52
3.3.3.1	Simbólica	52
3.3.3.2	Estatística	54
3.3.3.3	Conexionista	59
3.3.3.4	Híbrida	60
3.3.4	Vantagens e desafios do NLP	61
3.3.5	Principais termos do NLP	62
3.3.5.1	Corpus	62
3.3.5.2	Tokenization ou Tokenização	63
3.3.5.3	<i>Normalization</i> ou Normalização	64
3.3.5.4	<i>N-grams</i>	64
3.3.5.5	<i>Lexicons</i> ou Léxicos	65
3.3.5.6	Pré processamento e limpeza dos dados	66
3.3.6	Principais técnicas usadas no NLP	66
3.3.6.1	<i>Stop Words</i>	66
3.3.6.2	<i>Stemming</i> e <i>Lemmatization</i>	67
3.3.6.3	<i>Bag of words</i>	67
3.3.6.4	TF-IDF	68
3.3.6.5	<i>Word Embeddings</i>	68
3.3.6.6	<i>Named Entity Recognition</i> (NER)	69
3.3.6.7	<i>Parts-of-speech</i> (POS) <i>Tagging</i>	70
3.3.7	As 5 principais aplicações do NLP	71
3.3.8	3 exemplos de NLPs na prática	75
4	METODOLOGIA	78
4.1	Ferramentas e Recursos Utilizados	79

4.1.1	Google API e YouTube Data API	79
4.1.2	Preparação dos Dados	80
4.1.3	Modelo BERT para Análise de Sentimentos	80
4.1.4	Flask	83
4.2	Coleta de Dados	85
4.2.1	Busca por Vídeos de Avaliações de Produtos	85
4.2.2	Captura de Comentários	85
4.3	Processamento de Dados	86
4.3.1	Limpeza e Tratamento de Comentários	86
4.4	Análise de Sentimentos	87
4.4.1	Classificação com BERT	87
4.4.2	Limiar de Classificação	88
4.5	Interface de Usuário	89
4.5.1	Formulário de Pesquisa	89
4.5.2	Apresentação de Resultados	90
4.6	Treinamento do Modelo	91
4.6.1	Instalação de Pacotes	91
4.6.2	Importação de Bibliotecas	91
4.6.3	Carregamento e Exploração do <i>Dataset</i>	92
4.6.4	Tokenização do <i>Dataset</i>	92
4.6.5	Definição do Modelo BERT	93
4.6.6	Treinamento do Modelo	93
4.6.7	Avaliação do Modelo no Conjunto de Desenvolvimento	94
4.6.8	Avaliação do Modelo e Visualização das Métricas de Desempenho	94
4.6.9	Salvando o Modelo Treinado	95
4.6.10	Importação e Uso do Modelo Treinado	96
4.6.11	Classificação de Sentimentos em um Novo Texto	96
4.7	Utilização do Modelo BERT Treinado	96
4.7.1	Importação de Bibliotecas	96
4.7.2	Carregamento do Modelo BERT Pré-treinado e Tokenizador	97
4.7.3	Definição de Parâmetros	97
4.7.4	Classe para Representar Resultados de Vídeos	97
4.7.5	Função para Buscar Vídeos de Avaliações de Produtos no YouTube	98

4.7.6	Função para Processar e Salvar Comentários	98
4.7.7	Configuração da Aplicação Web com Flask	98
4.7.8	Rotas da Aplicação Web	98
4.7.9	Execução da Aplicação Web	99
4.7.10	Integração do Modelo BERT no Processamento de Comentários	99
4.7.11	Aplicação da Lógica de Classificação	99
4.7.12	Atualização dos Resultados e Exibição na Aplicação Web	100
4.8	<i>Scripts</i> de Processamento	100
4.8.1	Evento de Submissão do Formulário	100
4.8.2	Prevenção do Envio Padrão	101
4.8.3	Captura da Consulta de Pesquisa	101
4.8.4	Envio da Consulta para o Servidor	101
4.8.5	Processamento da Resposta do Servidor	101
4.8.6	Exibição dos Resultados na Página	101
4.9	Utilização da Interface Web - Detalhamento	102
4.9.1	Meta Tags e Título	102
4.9.2	Bootstrap e Estilo	102
4.9.3	Cor de Fundo e Layout Responsivo	103
4.9.4	Formulário de Pesquisa	103
4.9.5	Campos do Formulário	103
4.9.6	Botão de Pesquisa	103
4.9.7	Estilização Adicional	104
4.9.8	Estilo CSS Personalizado	104
4.9.9	Bibliotecas Externas	104
4.9.10	Dados Dinâmicos e Bootstrap	104
4.9.11	Informações da Pesquisa	104
4.9.12	Gráfico Geral	105
4.9.13	Recomendação Geral	105
4.9.14	Gráficos por Vídeo	105
4.9.15	Scripts para Gráficos	106
4.9.16	Script para Gráfico Geral	106
5	RESULTADOS	107
5.1	Resumo dos Resultados	107

5.1.1	Análise de Sentimentos Eficiente	107
5.1.2	Classificação de Comentários	107
5.1.3	Interface Intuitiva	107
5.2	Contribuições e Impacto	108
5.2.1	Auxílio na Tomada de Decisão	108
5.2.2	Otimização do Investimento e Tomada de Decisão para Empresá- rios, Consumidores e Fabricantes	111
5.2.3	Adaptabilidade a Diferentes Segmentos	111
5.3	Conclusões e Perspectivas Futuras	111
6	CONCLUSÃO	112
	REFERÊNCIAS	113

1 INTRODUÇÃO

Desde sua eclosão, em março de 2020, a pandemia de Covid-19 permitiu que o comércio virtual, conhecido como *e-commerce*, ganhasse muito espaço no Brasil e no mundo. Em terra tupiniquim, o varejo online deve crescer ainda mais rápido, 56% até 2024, mesmo com a reabertura das lojas físicas (ALBUQUERQUE, 2022b).

O termo *e-commerce*, que em português significa “comércio eletrônico”, refere-se ao método de vendas realizadas de forma virtual. A pandemia do COVID-19 paralisou grande parte do comércio presencial, resultando em um aumento sem precedentes no número de transações no mercado digital (OLIVEIRA, 2020a). Por isso, pesquisar a satisfação do consumidor é fundamental para empresas e fabricantes que buscam se manter competitivas no cenário atual. Porém, há uma grande dificuldade enfrentada por essas empresas no quesito de monitoração da percepção dos usuários em relação aos seus produtos em meio a uma grande quantidade de dados gerados nas redes sociais, por exemplo, e isso pode afetar, de maneira geral, a reputação e as vendas dessas empresas.

Um estudo realizado pela MarketingSherpa (instituto de pesquisa especializado em rastrear o que funciona em todos os aspectos do marketing) constatou que 95% dos adultos entre 18 e 34 anos seguem marcas nas redes sociais, e embora esse estudo tenha sido realizado no mercado americano, essa realidade também é vista no Brasil, onde milhares de brasileiros consomem produtos e serviços de maneira online (ZENDESK, 2022).

Além disso, à medida que a tecnologia se fortalece e acompanha o imediatismo do consumidor 4.0, as empresas têm visto as redes sociais como uma forma de se aproximar dos clientes, criando mais canais de experiência de conversação com eles por meio de uma comunicação mais horizontal (ZENDESK, 2022).

Nos dias de hoje, as redes sociais têm um papel fundamental na maneira como os consumidores interagem com as empresas e seus produtos. Ao mesmo tempo, elas geram uma quantidade cada vez maior de dados, que podem ser utilizados

para monitorar a percepção dos usuários em relação a determinados produtos e serviços. No entanto, a grande quantidade de dados gerados pode tornar o processo de monitoramento bastante desafiador para as empresas, fabricantes e, também, para os consumidores. Isso envolve o uso de ferramentas especializadas para coletar informações sobre a opinião dos usuários em relação aos produtos, serviços e marca. Em seguida, é necessário analisar os dados coletados e entender o que eles significam em termos de reputação da marca e vendas.

Outro desafio é a identificação de tendências e *insights* importantes. As empresas precisam estar atentas às mudanças na percepção dos usuários em relação aos seus produtos e serviços, bem como às tendências emergentes no mercado. Isso pode ajudá-las a tomar decisões informadas sobre a forma como seus produtos são desenvolvidos e comercializados.

A capacidade das empresas de monitorar a percepção dos usuários em relação aos seus produtos é essencial para o sucesso no mercado. O monitoramento da satisfação do cliente permite a detecção precoce de problemas relacionados à qualidade, desempenho e funcionalidade de um produto ou serviço, bem como expectativas não atendidas do cliente. É claro que o sucesso de uma organização depende de seus clientes, portanto, entender as necessidades atuais e futuras dos clientes permitirá que a organização atenda a seus requisitos e exceda suas expectativas (CHRISTINO, 2021). Nesse contexto, a análise de sentimentos propõe um conjunto de métodos que ajudam a identificar a satisfação do consumidor por meio de opiniões expressas textualmente (avaliações, mensagens, etc.), principalmente, quando a mesma é aplicada nas redes sociais. Para extrair o máximo de informações dessa enorme quantidade de dados não estruturados, a análise de sentimento geralmente requer a aplicação de técnicas de processamento de linguagem natural (NLP).

O Processamento de Linguagem Natural (PLN), mais conhecido como *Natural Language Processing* (NLP), é uma das áreas do aprendizado de máquina (*Machine Learning*) que cresce a cada dia. Essa é uma área que está inserida no cotidiano de grande parte da sociedade, como as pesquisas que são realizadas na plataforma do Google, por exemplo, e preenche a lacuna entre a comunicação humana e a da

máquina.(PREMEBIDA, 2021)

Além disso, visando a utilização de NLP para a captura de emoções registradas no meio digital, cada vez mais as redes sociais online, tais como Twitter e Facebook, são utilizadas como um local para as pessoas desabafarem alguma insatisfação com empresas ou com produtos adquiridos ou desejados, ou até mesmo prestar elogios a referidos negócios e instituições comerciais. Os clientes notaram que obtêm respostas mais rápidas das empresas devido à postagens realizados nesses meios de comunicação (CASTRO, 2013). Nesse viés, essas redes acabam sendo canais muito eficazes de comunicação com o cliente. Por outro lado, as empresas estão cientes de que a repercussão das críticas "ruins" nas redes sociais podem afetar rapidamente um grande número de clientes e, assim, afetar a reputação da empresa no mercado, que é maior do que a propaganda tradicional na TV. Outro fator importante das redes sociais é a capacidade de atingir um público específico, o que é muito mais eficaz do que a publicidade em massa.

A motivação primordial deste trabalho é impulsionar de maneira abrangente o setor de *e-commerce*, especializado na venda e distribuição de produtos. Porém, busca-se não apenas facilitar a identificação de tendências e *insights* valiosos de mercado para otimizar as vendas e a gestão de produtos personalizados e estoque, mas também estender esses benefícios aos consumidores finais. Para concretizar esse objetivo principal, o projeto explorará as áreas de tecnologia e desenvolvimento, concentrando-se em inteligência artificial, aprendizado de máquina e processamento de linguagem natural.

A abordagem incluirá a realização de buscas, coleta e extração de grandes volumes de dados e conteúdos provenientes de vídeos da plataforma YouTube. Esses dados serão meticulosamente filtrados e selecionados com base em comentários que apresentem conteúdo de maior relevância em relação aos produtos do setor de venda. Seguindo essa premissa, será desenvolvido um sistema com o objetivo final de processar todo esse conteúdo, abrangendo aspectos positivos e negativos avaliados sobre os produtos. O resultado será a geração de relatórios, estatísticas e indicações das corretas tendências de mercado identificadas em um ambiente social virtual de fácil acesso.

Dentro desse contexto, os passos essenciais para a concepção e desenvolvimento do projeto serão detalhados, apresentando os resultados alcançados ao longo desse processo. Por fim, é válido ressaltar, que esta abordagem cumpre com o objetivo de não apenas beneficiar as empresas no aprimoramento de suas estratégias, mas também proporcionar aos clientes uma ferramenta eficaz para avaliar a adequação e recomendação de produtos com base nas análises de sentimentos fornecidas pelo projeto.

2 TRABALHOS RELACIONADOS

Atualmente há um grande número de trabalhos e pesquisas disponíveis em livros, revistas e principalmente na Internet relacionados área de análise de sentimentos, tendências de mercado e tecnologias como o *machine learning*, inteligência artificial e o processamento de linguagem natural. Dentre esses, destacam-se diversos estudos voltados para a análise de conteúdo gerado pelos usuários da Internet e seu processamento, visando gerar informações relevantes principalmente para o setor de *e-commerce*, mas também para os consumidores finais e fabricantes.

No trabalho de Grassi et al. (2011) foi criado um *website* no qual é possível assistir e navegar entre 100 vídeos retirados do Youtube. Entretanto, o *website* vai além da navegação que estamos acostumados, pois, usando técnicas da inteligência artificial e da web semântica para classificar os vídeos, ele torna possível fazer buscas de acordo com os sentimentos associados aos vídeos. Por exemplo, o usuário pode listar todos vídeos que causam medo ou vídeos engraçados.

A ferramenta SPOONS do Netflix, conforme Augustine et al. (2012), é um exemplo prático da utilidade da análise de sentimentos em redes sociais. O SPOONS usa mensagens compartilhadas pelos usuários do Twitter¹ (*tweets*), para detectar falhas de disponibilidade do Netflix. Várias técnicas são utilizadas para tentar detectar falhas no serviço através dos *tweets*, entre elas o SPOONS classifica as mensagens em categorias, por exemplo, feliz, neutro e reclamação, para saber se muitos usuários estão reportando problemas na rede social. Isto é possível através do uso de classificadores que são treinados com *tweets* que foram rotulados em categorias manualmente por especialistas.

O trabalho de Davidov, Tsur e Rappoport (2010) usou uma ideia simples e inteligente para treinar um classificador de sentimentos em *tweets*. Em vez de usar especialistas para rotular manualmente as mensagens do conjunto de treinamento, foram utilizadas 50 *hashtags* e 15 *emoticons*, estruturas que aparecem naturalmente nos *tweets*, como

¹ Rede social e serviço de microblog, que permite aos usuários enviar e receber atualizações pessoais de outros contatos em textos de até 280 caracteres.

os possíveis rótulos. Dessa forma, o trabalho de classificar as mensagens do conjunto de treinamento passou a ser automático, possibilitando gerar um conjunto muito maior em um tempo muito menor.

Uma abordagem um pouco diferente foi proposta por Hassan, Qazvinian e Radev (2010), neste trabalho não foram classificados quais sentimentos estão relacionados a uma mensagem textual, o foco dele foi avaliar a interação entre usuários através de mensagens de texto, para saber se um usuário tem uma atitude positiva ou negativa em relação aos outros usuários.

Com relação as tecnologias de processamento de linguagem natural e análise de sentimentos, junto ao envolvimento de ambientes de comércio eletrônico ou *e-commerce*, o trabalho de Albuquerque (2022a) tem como projeto final um modelo de processamento de linguagem natural para análise de sentimento sobre comentários de plataformas de *e-commerce*, dentre as quais estão plataformas como B2W², Buscapé³, Olist⁴ e UTLCorpus⁵, de maneira automática e no idioma português/BR. Posto isto, os resultados obtidos através de uma análise experimental mostram que o modelo proposto fornece uma acurácia acima de 85% para todos os conjuntos de dados testados. Desta maneira, os resultados obtidos mostraram que o modelo proposto apresenta ótimos resultados para análise de sentimento positivo e negativo sobre dados de comentários de plataformas de *e-commerce*, desse modo, auxiliando empresas em compreender melhor seus usuários e a criar novas estratégias de vendas.

Seguindo esse mesmo contexto de processamento de linguagem natural, no trabalho de Oliveira (2020b) é abordado um estudo aprofundado a respeito do banco de dados da Olist Store⁶, contendo 100 mil registros, utilizando processamento de linguagem natural (PNL) para criação de modelos lineares e não lineares de classificação da avaliação de compradores de lojas virtuais através dos seus comentários. Diante disso, como resultados, foi possível observar que diferentes técnicas de análise indicam que o atraso é o principal fator de insatisfação dos clientes. Foram testados 8 modelos

² Empresa de varejo online na América Latina.

³ Plataforma do comércio eletrônico brasileiro.

⁴ Startup brasileira que atua no segmento de tecnologia para varejo.

⁵ Corpus de opiniões sobre filmes e aplicativos para smartphones.

⁶ É um canal de vendas digital que está presente nos maiores marketplaces do país.

de classificação e escolhidos os 4 melhores (2 lineares e 2 não lineares) para a experimentação aprofundada. Dentre os modelos experimentados, os não lineares alcançam os melhores desempenhos. Em todas as análises realizadas observa-se a dificuldade em prever a classe menos frequente do banco de dados.

Conforme trabalho focado em ponderação de desempenho de um *e-commerce* realizado por Alonso, Toniolo e Lopes (2020), um projeto de análise de sentimento de um *e-commerce* foi elaborado e realizado a fim de entender a relação entre avaliação e desempenho de vendas, contribuindo para a melhoria de desempenho empresarial como também científico, ao disponibilizar ferramentas que permitam análises neste e em outros contextos. Dessa maneira, o modelo de análise de sentimentos foi construído utilizando técnicas de *data analytics*⁶ como a mineração de texto, conhecida também como *Text Mining*⁷, de forma a relacionar e compreender as duas variáveis. Buscando entender as relações entre as variáveis citadas, com base no estudo, a abordagem apresentou como resultado tendências de vendas baseadas nas polaridades (positivo, neutro e negativo) das avaliações feitas em um período de tempo.

Um caso mais tendencioso foi desenvolvido no trabalho de Rodrigues, Flávio e Neto (2021), em que foi identificada uma oportunidade de analisar o sentimento do consumidor através destas opiniões para que seja possível expressar graficamente o seu sentimento em relação a determinado produto. Para isso, foi utilizada a biblioteca *syuzhet*, cujo através dela, é possível obter pontuações para cada palavra que representa alguma emoção, essas emoções foram definidas de acordo com o modelo de Plutchik, criado pelo psicólogo com o mesmo nome, Robert Plutchik. Ele define que existem 8 emoções básicas: alegria, confiança, medo, surpresa, tristeza, expectativa, raiva e desgosto. Observar os resultados obtidos do processo de análise de sentimentos dos produtos mostrados anteriormente, pode auxiliar os consumidores na decisão de compra, por exemplo analisando o sentimento de quem comprou o *smartphone* Iphone da Apple a avaliação negativa é menor que a metade da avaliação positiva, claramente os consumidores que adquiriram este produto ficaram bastante satisfeitos com a aquisição. Contudo, acredita-se que esse tipo de análise utilizando

⁶ Processo de inspeção, limpeza, transformação e modelagem de dados com o objetivo de descobrir informações úteis, informar conclusões e apoiar a tomada de decisões.

⁷ Processo de obtenção de informações importantes de um texto.

uma importante estratégia de mercado pode auxiliar consumidores a visualizarem o sentimento de outros consumidores que adquiriram o produto que se deseja antes de efetuar a compra e assim atingir o nível de satisfação que esperamos obter após adquirir o produto desejado.

Por fim, um grande fator responsável pelo sucesso de um *e-commerce* é analisar e desenvolver boas estratégias de Marketing Digital, fato que consta na obra acadêmica de Mena et al. (2019). Para tal trabalho, tornou-se por objetivo analisar os critérios de consumo em relação à influência das estratégias de marketing digital de empresas de comércio eletrônico, para o qual foi utilizada uma metodologia qualitativo-quantitativa, explorando os resultados obtidos com 192 usuários de produtos ou serviços prestados por *e-commerce*. A hipótese de pesquisa foi verificada, ou seja, as estratégias de marketing digital em empresas de comércio eletrônico influenciam a perspectiva do consumidor. Concluindo, sob a ótica do consumidor, 41% das empresas já se aventuraram a estabelecer seus negócios na Internet e, desta forma, aplicar estratégias de marketing digital.

3 FUNDAMENTAÇÃO TEÓRICA

Os capítulos apresentados anteriormente serviram para introduzir o projeto, de modo a auxiliar no entendimento do trabalho. Este capítulo apresenta conceitos-chave referentes às áreas relacionadas.

3.1 REDES SOCIAIS E SUA IMPORTÂNCIA

As redes sociais revolucionaram a forma como as pessoas se comunicam e interagem na Internet. Elas oferecem uma nova maneira de se conectar com outras pessoas, independentemente da distância geográfica ou das barreiras culturais. Desde o surgimento do Orkut em 2004 até plataformas como o TikTok mais recentemente, o impacto global das redes sociais tem sido enorme.

Esses ambientes sociais virtuais são estruturas compostas por pessoas e empresas, dentro ou fora da internet. Dessa forma, seus usuários se conectam por terem interesses, objetivos e valores em comum. A entrada da internet na sociedade mundial fez as mídias sociais ganharem força a ponto de criar redes sociais referentes, como o Instagram, o Facebook, o Twitter e também, como mencionado anteriormente, o TikTok (SULZ, 2020).

Um dos principais pontos fortes das redes sociais é a capacidade de reunir pessoas que compartilham interesses, ideias e objetivos comuns. Por meio delas, um indivíduo pode se conectar com amigos, familiares, colegas de trabalho, além de conhecer novas pessoas e se envolver em causas sociais e políticas.

O domínio do ambiente online e seu impacto na sociedade é enorme, principalmente quando o assunto é marketing digital e/ou transformação digital. Atualmente, são mais de 3,8 bilhões de pessoas em todo o mundo conectadas a mundos virtuais (SULZ, 2020). Além disso, as redes sociais também são ferramentas poderosas para fins profissionais. Elas permitem que empresas e empreendedores se conectem com seu público-alvo, aumentem a visibilidade da marca, anunciem produtos e serviços e recebam *feedback* em tempo real.

Conforme E-commerce (2023), para reforçar a importância dessas plataformas no comércio eletrônico, existem 8 principais benefícios a serem destacados, são eles:

a) Criar uma relação extra comercial com os consumidores

Manter interações exclusivamente comerciais é algo que dificilmente atrai e fideliza clientes. Com as redes sociais, eles esperam muito mais de uma empresa do que apenas comprar e vender. É preciso ter diálogo, conteúdo, confiança e até mesmo descontração. São detalhes que humanizam mais a relação entre marcas e consumidores, e criam diferenciais para que o público não seja convencido apenas por preço ou conveniência.

b) Extrair *insights* sobre o comportamento dos consumidores

Ao estar nas redes sociais, há a oportunidade de ver o que as pessoas comentam sobre a própria loja, seus conteúdos, produtos, concorrentes e diversos outros assuntos relevantes. Isso permite entender melhor como potenciais e atuais clientes pensam e agem. Dessa forma, será possível adequar as estratégias de marketing e vendas às preferências desses públicos.

c) Trazer mais visibilidade para a loja

De acordo com o relatório Social Media Trends de 2017, a maior visibilidade é destacada por 83,2% das empresas como uma das vantagens de usar as redes sociais. E esse benefício é bem notório. Ao criar perfis e executar uma estratégia de marketing nas redes sociais, um *e-commerce* está expandindo suas fronteiras. Diante disso, a empresa se torna mais localizável para pessoas que dificilmente a encontrariam de outras maneiras pela internet.

d) Aumentar a reputação da marca

Uma página que promove discussões interessantes, que compartilha conteúdos úteis para os seguidores e que não se envolve em polêmicas ajuda a melhorar (e muito) a credibilidade da marca ou empresa por ela representada. Além disso,

a própria participação dos consumidores pode elevar a confiança na marca. Por exemplo, se um cliente satisfeito com uma compra fizer uma avaliação positiva na *fan page* da loja, aquilo será uma prova social valiosíssima para o *e-commerce*.

e) Gerar mais engajamento com o público

Esses tipos de comportamento que foram citados no tópico anterior aumentam não só a reputação, como também o engajamento da marca com o público. As pessoas deixam de ser indiferentes a cada *post* referente à loja. Em vez disso, elas passam a reagir, opinar, criticar, elogiar e compartilhar publicações da empresa. A propósito, a interação com o público e o engajamento com a audiência são citados por pouco mais de 60% das empresas no Brasil como vantagens de estar presente nas redes sociais, segundo a mesma pesquisa Social Media Trends.

f) Melhorar os resultados de SEO (*Search Engine Optimization*)

A visibilidade, reputação ou engajamento nas redes sociais não são necessariamente fatores que colocarão sua loja na primeira página do Google. No entanto, elas ajudam indiretamente nessa missão. As redes sociais podem contribuir de duas formas. Uma, e a mais simples, é com a indexação mais rápida das suas páginas nos resultados de pesquisa. Quando um *post* é compartilhado em alguma plataforma social, o Google detecta mais rapidamente a existência dele. A segunda, e tão simples quanto a primeira, tem a ver com o ganho de tráfego. A quantidade de visitas que um site recebe é um dos fatores de ranqueamento. Portanto, se a loja estiver recebendo muitos acessos das mídias sociais, aquilo será premiado pelo Google e outros buscadores.

g) Segmentar a divulgação dos produtos

Outro benefício é o poder de segmentação das redes, especialmente em suas plataformas de anúncios. Ferramentas como o Facebook Ads, por exemplo, permitem praticamente a escolha a dedo as pessoas impactadas em uma campanha. Trata-se de uma segmentação que vai além de dados demográficos. Esse direcionamento pode ser feito de acordo com perfil de compra, interesses pessoais, páginas curtidas,

atividades recentemente publicadas, etc. Todo esse detalhamento ajuda a criar ações mais acertadas e econômicas de marketing e vendas.

h) Impulsionar as vendas

Há uma ideia de que 75% dos consumidores compraram alguma oferta só porque a viram anteriormente nas redes sociais? Pois é, esse dado da Sprout Social mostra como uma publicação na própria página da empresa, um *post* pago para ser promovido ou um conteúdo trazido pelo público podem influenciar outros consumidores a comprarem algo. Além do mais, com uma visibilidade maior, é natural que a loja conquiste mais oportunidades de negócio e, conseqüentemente, mais conversões em clientes.

Por meio dessas plataformas, as lojas online podem se tornar mais conhecidas, confiáveis, inteligentes e financeiramente desenvolvidas. Independentemente disso, os benefícios da mídia social variam amplamente. No entanto, para experimentar todos esses benefícios, é necessário adotar uma estratégia consistente de melhores práticas. Personalizar a presença da empresa, compartilhar conteúdos interessantes e construir bons relacionamentos são apenas algumas das estratégias para o sucesso. As redes sociais, juntamente com e-mail e blogs, são os principais canais para as empresas de comércio eletrônico interagirem com consumidores e clientes, desempenhando um papel vital na construção de uma presença online eficaz (E-COMMERCE, 2023).

A influência das redes sociais na experiência de compra é notável. A dinâmica e a interatividade dessas plataformas transformaram a maneira como as pessoas descobrem, consideram e decidem adquirir produtos online. O compartilhamento de experiências e opiniões, a construção de uma comunidade virtual baseada na confiança e a criação de uma presença autêntica por meio de campanhas e engajamento ativo são elementos-chave para o sucesso nas redes sociais.

Além disso, a presença visualmente estimulante nas redes sociais, com conteúdo atraente como fotos de produtos e vídeos explicativos, contribui significativamente para influenciar as decisões de compra. A utilização de estratégias de marketing de influenciadores também se destaca, aproveitando a credibilidade e autenticidade

de figuras conhecidas para impactar positivamente as decisões dos consumidores (E-COMMERCE, 2023).

No entanto, é crucial reconhecer que o uso responsável das redes sociais é fundamental. Apesar de sua capacidade transformadora, é necessário equilibrar o engajamento com a responsabilidade. Em resumo, as redes sociais são uma ferramenta poderosa e essencial na sociedade atual, conectando pessoas, ampliando possibilidades de interação e criando novas formas de comunicação. Utilizá-las com sabedoria e estratégia é fundamental para maximizar seus benefícios.

3.2 ANÁLISE DE SENTIMENTOS

O objetivo da pesquisa TM (*Text Mining*) é encontrar padrões em textos de linguagem não natural e extrair informações para um determinado propósito. Geralmente, é composto por três etapas, a saber: pré-processamento, análise dos dados (extração) e avaliação dos resultados. A etapa de pré-processamento está relacionada à limpeza dos dados, normalmente, realiza-se a correção ortográfica, a remoção de *stop words* (artigo, rejeição, preposição) e o *stemming* (remoção de variações de plural, gerúndio e sufixos temporais). A fase de análise de dados refere-se ao uso de algoritmos de aprendizagem para extrair conhecimento das informações contidas em documentos de texto. Esses algoritmos podem executar tarefas como associação, resumo, classificação e agrupamento. A avaliação das descobertas valida o conhecimento extraído e relatórios podem ser gerados com base em tendências, padrões ou regularidades observadas nos documentos analisados (EVANGELISTA; PADILHA, 2014).

A TM aplica algoritmos computacionais para processar grandes volumes de textos em linguagem natural não estruturados, a fim de extrair informações relevantes, úteis e inovadoras que não podem ser recuperadas usando métodos padrões de pesquisa e consulta, identificando regularidades, tendências e padrões não triviais (ARANHA; PASSOS, 2006) e (BARION; LAGO, 2008).

A análise de sentimento é um ramo da TM que visa classificar o texto não por tópico, mas pelo sentimento ou opinião contida em determinado documento (SANTOS, 2010). Neste caso, é o estudo computacional das opiniões, sentimentos e emoções expressos

em um texto, utilizando técnicas para identificar e extrair informações subjetivas contidas em determinado material, verificar quais sentimentos eles representam e decidir se o documento exprime uma opinião positiva, negativa, neutra ou, até mesmo, tentar classificar qual emoção está presente como raiva, felicidade, tristeza (LIU, 2010) e (NASCIMENTO; OSIEK; XEXÉO, 2012).

A classificação de sentimentos é semelhante à classificação clássica de texto (por tópicos). A classificação de sentimentos funciona principalmente em palavras que contêm conteúdo subjetivo (que carregam consigo um conteúdo opinativo) que expressam sentimento positivo ou negativo (opinião), como ótimo, excelente, maravilhoso, péssimo e ruim (LIU, 2010).

Em resumo, a análise de sentimento é uma técnica que utiliza inteligência artificial para analisar e entender o sentimento expresso em textos, como postagens em redes sociais, comentários em blogs, avaliações de produtos e serviços, etc. Empresas e organizações usam essa tecnologia extensivamente para entender a percepção do público sobre suas marcas, produtos e serviços e para identificar tendências e padrões de comportamento.

3.2.1 Importância da análise de sentimentos

A análise de sentimentos, também conhecida como mineração de opiniões, emerge como uma poderosa ferramenta de *business intelligence* com amplo potencial. Além de capacitar as empresas na melhoria de produtos e serviços, essa abordagem oferece benefícios substanciais tanto para os clientes quanto para as fabricantes. Para os consumidores, proporciona decisões de compra mais informadas, permitindo a avaliação da adequação dos produtos às suas necessidades por meio da compreensão das opiniões de outros clientes. Para as fabricantes, oferece uma análise detalhada do *feedback* do consumidor, permitindo a identificação de áreas de melhoria, ajustes estratégicos e adaptação de produtos para melhor atender às expectativas do mercado. Assim, a análise de sentimentos promove uma dinâmica equilibrada e vantajosa entre empresas, clientes e fabricantes, gerando benefícios abrangentes para todas as partes envolvidas. Diante disso, há alguns pontos específicos em que a análise de sentimentos pode ser muito útil e eficaz, são eles:

a) Fornecer *insights* objetivos

O uso de ferramentas de análise de sentimento, conforme destacado por Amazon (2023), proporciona benefícios substanciais tanto para os clientes quanto para as fabricantes e *e-commerces*. Para os clientes, isso se traduz em avaliações mais imparciais e confiáveis, facilitando decisões de compra fundamentadas. Ao evitar tendenciosidades humanas, exemplificado pela avaliação do processador, os consumidores podem confiar em informações mais objetivas. Para as fabricantes, a implementação da análise de sentimento oferece a capacidade de acessar *feedbacks* mais imparciais e confiáveis. Ao evitar vieses pessoais, as empresas podem confiar em resultados consistentes para ajustar produtos e estratégias, garantindo uma abordagem mais objetiva ao atender às demandas do mercado.

b) Criar produtos e serviços melhores

A aplicação de sistemas de análise de sentimento não apenas beneficia os clientes, mas também proporciona vantagens significativas para as fabricantes e *e-commerces*. Para os clientes, essa abordagem significa produtos e serviços aprimorados com base em *feedbacks* genuínos. Ao identificar objetos ou situações associadas a emoções negativas, a tecnologia de IA permite que as empresas ajustem seus produtos de acordo. No exemplo do processador, os clientes podem esperar melhorias específicas no gerenciamento térmico, refletindo diretamente em uma experiência aprimorada. Para as fabricantes, essa análise oferece *insights* valiosos e acionáveis para aprimorar produtos com base em *feedbacks* concretos. Ao entender as associações entre emoções negativas e entidades específicas, as empresas podem direcionar seus esforços para melhorias precisas, garantindo uma abordagem centrada no cliente (AMAZON, 2023).

c) Análise em alta escala

As empresas continuam a extrair *insights* de extensos volumes de dados não estruturados, que incluem e-mails, transcrições de *chatbots*, pesquisas, registros de gerenciamento de relacionamento com clientes e análises de produtos. A adoção de ferramentas de análise de sentimento baseadas na nuvem não só simplifica, mas

também torna acessível o processo de revelar as emoções dos clientes em dados de texto. Esta abordagem eficaz beneficia não apenas os clientes, proporcionando uma compreensão mais aprofundada de suas percepções, mas também é vantajosa para os fabricantes, permitindo que dimensionem suas operações de maneira eficiente e econômica (AMAZON, 2023).

d) Resultados em tempo real

Na dinâmica contemporânea de mudanças rápidas, as empresas precisam de respostas ágeis a possíveis crises ou tendências de mercado. Profissionais de marketing confiam nas capacidades do software de análise de sentimento para compreender em tempo real como os clientes percebem a marca, produtos e serviços de uma empresa, agindo imediatamente de acordo com essas percepções. Esta abordagem não só beneficia os clientes, proporcionando uma resposta ágil a suas necessidades e sentimentos, mas também é vantajosa para os fabricantes, permitindo-lhes ajustar estratégias e operações de forma imediata e precisa (AMAZON, 2023).

3.2.2 Casos de uso de análise de sentimentos

A análise de sentimentos é uma ferramenta essencial para analistas de dados em grandes empresas, auxiliando na avaliação da opinião pública, pesquisas de mercado e monitoramento da reputação de marcas e produtos, proporcionando uma compreensão mais profunda das percepções dos clientes. Essa estratégia eficaz não apenas beneficia os clientes, mas também concede vantagens significativas aos fabricantes, permitindo uma adaptação estratégica fundamentada em dados acionáveis (GOMES, 2019).

Além disso, essa análise é aplicável a diversas estruturas de dados com conteúdo textual normalizado ou padronizado, sendo destacada pela Amazon (2023) como uma prática valiosa para a melhoria da tomada de decisões em diversos contextos de negócios, beneficiando clientes, fabricantes e *e-commerces*.

No dinâmico cenário do comércio online, a análise de sentimentos emerge como uma

ferramenta crucial, proporcionando benefícios significativos tanto para os *e-commerces* quanto para os consumidores. Segundo ressalta a Amazon (2023), ao aplicar ferramentas de análise de sentimentos no atendimento ao cliente, os *e-commerces* têm a capacidade de personalizar respostas, adaptando-as ao tom da conversa e identificando com eficiência tópicos urgentes, contribuindo diretamente para a satisfação do cliente.

No contexto do monitoramento da marca em ambientes digitais, a tecnologia de análise de sentimentos emerge como uma ferramenta estratégica, otimizando a compreensão e resposta a histórias relevantes e persistentes. Essa abordagem proporciona aos *e-commerces* uma visão inestimável para gerenciar eficazmente a reputação online de suas marcas, conforme destaca a pesquisa da Amazon (2023). Além disso, na esfera da pesquisa de mercado, os sistemas de análise de sentimentos desempenham um papel crucial na melhoria contínua das ofertas de produtos, permitindo uma abordagem orientada pelo cliente no desenvolvimento de novos produtos e estratégias.

A gestão da performance de campanhas publicitárias é outra área de impacto significativo identificada pela Amazon (2023). Profissionais de marketing em *e-commerces* utilizam ferramentas de análise de sentimentos para garantir respostas desejadas e ajustes imediatos com base em dados em tempo real, atendendo às expectativas dos clientes. Esses casos de uso destacam a essencialidade da análise de sentimentos para criar uma experiência de compra online mais informada e satisfatória, proporcionando ferramentas valiosas para os *e-commerces* otimizarem suas estratégias de marketing e interação com os consumidores.

Os benefícios no dinâmico cenário do comércio online são igualmente expressivos na perspectiva dos clientes. Ao acessarem *feedbacks* de outros consumidores, os compradores, conforme ressaltado pela Amazon (2023), obtêm *insights* imparciais e confiáveis sobre os produtos desejados. A análise de sentimentos permite uma compreensão mais profunda das experiências compartilhadas por outros consumidores, contribuindo assim para decisões de compra mais informadas.

A personalização do atendimento ao cliente, impulsionada pela análise de sentimentos,

oferece aos compradores respostas mais relevantes às suas dúvidas e preocupações específicas, proporcionando uma experiência de compra mais alinhada com suas expectativas. O monitoramento da marca nas mídias sociais fornece aos consumidores uma visão em tempo real das percepções de outros clientes sobre uma determinada marca ou produto, influenciando positivamente suas escolhas (AMAZON, 2023).

No contexto da pesquisa de mercado, a análise de sentimentos permite que os clientes identifiquem não apenas os pontos fortes, mas também as áreas de melhoria de um produto, como destacado pela Amazon (2023). O acesso a comentários sobre características específicas, funcionalidades ou desempenho oferece aos compradores uma compreensão holística, permitindo-lhes alinhar suas expectativas às realidades do produto.

Ao acompanhar a performance de campanhas publicitárias, os clientes podem, em tempo real, perceber como outros consumidores estão respondendo a produtos específicos, contribuindo para a formação de expectativas realistas e, assim, aumentando a confiança nas decisões de compra. Essa visibilidade proporcionada pela análise de sentimentos contribui para que os clientes estejam mais propensos a adquirir produtos que atendam às suas necessidades e expectativas individuais (AMAZON, 2023).

Em suma, a análise de sentimentos não apenas enriquece a experiência de compra online dos clientes, fornecendo informações valiosas e imparciais, mas também capacita-os a tomar decisões informadas, contribuindo para a construção de um ambiente de comércio eletrônico mais transparente e confiável.

3.2.3 Funcionamento da análise de sentimentos

A análise de sentimentos é uma aplicação da tecnologia de processamento de linguagem natural (PLN) que treina o software de computador para entender o texto de maneira semelhante à humana. A análise normalmente passa por vários estágios antes de produzir um resultado final (AMAZON, 2023).

a) Pré-processamento

Segundo Amazon (2023), no estágio de pré-processamento, a análise de sentimentos identifica palavras-chave para destacar as principais informações do texto.

Ou seja, o pré-processamento de texto envolve a conversão do banco de dados bruto em um formato mais tratável computacionalmente. Uma dessas etapas é retirar as *stop words*, que são palavras que não devem ser consideradas na análise, como "o", "e", "a", "de", "que", "do", etc. Essas palavras foram removidas porque não fornecem informações relevantes para a construção de modelos analíticos, apenas definem ou expressam ideias (GOMES, 2019).

b) *Stemming*

É uma técnica de redução dos termos aos seus radicais, retirando afixos e vogais sujeitas. Um de seus objetivos é reduzir o número de palavras no banco de dados. Reduzir o vocabulário aumenta a velocidade de processamento e reduz erros causados por combinações imprevistas (GOMES, 2019).

Exemplos:

- Frequentemente se torna "Frequent"
- Copiar se torna "Copi"

c) Pesquisa Semântica Contextual (CSS)

Para obter *insights* que produzam algum tipo de resultado real, é importante entender de que aspecto os usuários de uma marca específica estão discutindo (GOMES, 2019).

Para uma empresa como a Amazon, ela provavelmente gostaria de segregar mensagens relacionadas a entregas atrasadas, problemas de cobrança, consultas relacionadas a promoções e análises de produtos. Quanto à Starbucks, ela provavelmente gostaria de classificar as mensagens com base no comportamento dos funcionários, novos sabores de café, *feedback* sobre a higiene do ambiente,

pedidos online, nome e localização da loja, etc (GOMES, 2019).

A maneira tradicional de filtrar todas as mensagens relacionadas a "Preço" é fazer uma pesquisa por palavra-chave para esse e outros termos relacionados (por exemplo, preço, cobrança, \$, pago). No entanto, esta abordagem não é muito eficaz. É quase impossível encontrar todas as palavras-chave relevantes e suas variações que representam um determinado conceito (GOMES, 2019).

Para isso, é utilizado um algoritmo de busca inteligente chamado *Contextual Semantic Search* (pesquisa semântica contextual), também conhecido como *CSS* (GOMES, 2019).

O algoritmo recebe milhares de mensagens e um atributo (como "preço") como entrada. Em seguida, ele filtra todos os contextos semelhantes, mesmo que não sejam mencionadas variantes óbvias da palavra-chave (GOMES, 2019).

d) Polaridade

De acordo com Gomes (2019), a polaridade indica o quão positivo ou negativo é o texto. Geralmente é o resultado de uma análise e podendo ser binário (positivo ou negativo) ou ternário (positivo, negativo ou neutro), conforme exemplificado a seguir:

- "Carol está muito bonita hoje" é Positivo
- "Estou tendo um péssimo dia" é Negativo
- "O Natal é amanhã" é Neutro

e) Força do sentimento

A intensidade de um sentimento indica a força de um sentimento, geralmente um número variável entre (-1 e 1). Esse recurso é usado como indicador, pois ajuda a captar se o sentimento analisado em uma frase é positivo ou muito positivo, por

exemplo. Isso torna a análise mais eficiente (GOMES, 2019).

f) Análise de Emoções

Embora muitas vezes seja confundido com análise de sentimentos, eles não são sinônimos. A análise de emoções é um método utilizado para classificar emoções específicas, como raiva, surpresa e felicidade, presentes em uma mensagem (GOMES, 2019).

Por exemplo, o método *Emolex Lexicon* é usado para descrever sentimentos em sentenças em inglês. São eles: *anger, anticipation, disgust, fear, joy, sadness, surprise, trust* (GOMES, 2019).

Por último, além de identificar a emoção principal, é comum analisar as porcentagens de cada uma (GOMES, 2019).

g) Subjetividade vs. Objetividade

A capacidade de classificar sentenças de sentimentos predeterminados pode acabar ficando aquém dos resultados desejados. Em muitos desses casos, o sistema de análise pode misturar sentenças objetivas com sentenças subjetivas, que devem ser interpretadas de forma diferente (GOMES, 2019).

Em geral, textos objetivos não podem ser misturados com textos subjetivos para não prejudicar a análise, o que é mais comum em publicações em redes sociais. Mas claro, tudo depende do método utilizado para realizar a análise (GOMES, 2019).

3.2.4 Abordagens para a análise de sentimentos

De acordo com Cambria et al. (2013), a mineração de opiniões pode ser agrupada em quatro campos, na qual a análise pode ser realizada por meio de:

3.2.4.1 Com base em regras (palavras-chave e afinidade léxica)

Os métodos baseados em regras, conforme Amazon (2023) identificam, classificam e pontuam palavras-chave específicas de acordo com um léxico predeterminado. Um

léxico é uma coleção de palavras que representam as intenções, sentimentos e emoções de um autor. Os profissionais de marketing atribuem pontuações de sentimentos a dicionários positivos e negativos para refletir o peso emocional de diferentes expressões. Para determinar se uma frase é positiva, negativa ou neutra, o software busca palavras contidas em um léxico e soma uma pontuação de sentimento. A pontuação final é comparada a um limiar emocional para determinar a carga emocional geral.

Ou seja, esses métodos classificam o texto com base na presença de palavras inequívocas, como "feliz", "triste" e "medo". Além de detectar palavras óbvias, também atribui a outras palavras uma afinidade com um sentimento, bom ou ruim. Um aplicativo de exemplo é o SentiWordNet10 3.0, um recurso de dicionário para guiar aplicativos de mineração de opinião (BACCIANELLA; ESULI; SEBASTIANI, 2010).

3.2.4.2 Aprendizado de máquina/ML

Essa abordagem usa técnicas de *machine learning* (ML) e algoritmos de classificação de sentimentos para a classificação do texto, como *Naive Bayes*, *Support Vector Machines* (SVM) e redes neurais e aprendizado profundo para ensinar o software de computador a reconhecer sentimentos em texto. Nesse caso, além de analisar o texto em busca de frequência ou possibilidades de pontuação, o sistema considera outras palavras que podem ser chaves, além de aprender a importância de palavras-chave óbvias. O processo envolve a criação de um modelo de análise de sentimentos e o treinamento repetido do modelo em dados conhecidos para que ele possa adivinhar o sentimento em dados desconhecidos com alta precisão (AMAZON, 2023).

3.2.4.3 Orientação semântica

Esses métodos calculam a orientação semântica das palavras com base em sua coocorrência com palavras com a mesma orientação (por exemplo, para problemas binários em que a polaridade da palavra é conhecida). O principal trabalho propondo métodos para computar tais orientações semânticas é o algoritmo proposto por Turney (2002). O algoritmo *Pointwise Mutual Information and Information Retrieval* (PMI-IR) é usado para medir a similaridade de pares de palavras ou frases. A orientação é calculada comparando a semelhança de uma palavra com sentimentos positivos e negativos.

3.2.4.4 Baseado em conceitos

Neste método, realiza-se análises de texto usando ontologias ou redes de palavras-chave. Ou seja, pode-se analisar expressões que não têm uma emoção explícita, mas estão associadas a uma emoção implícita. No trabalho realizado por Kontopoulos et al. (2013), a fim de melhorar o desempenho da análise de sentimentos no Twitter, foi proposto o uso de ontologias.

Portanto, pode-se notar que existem diversas técnicas de análise de sentimentos, porém, o foco deste estudo é utilizar modelos de aprendizado de máquina juntamente com técnicas de extração de características para treinar e classificar um conjunto de opiniões.

3.2.5 Diferentes tipos de análise de sentimentos

A análise de sentimentos é utilizada por empresas em diferentes tipos, a fim de entender como seus clientes se sentem ao interagirem com produtos ou serviços. Conforme TIBCO (2022), os principais tipos que podem ser citados são os seguintes:

3.2.5.1 Pontuação refinada

A análise refinada de sentimentos decifra a polarização da opinião pública. Essa análise pode ser tão simples como uma distinção emocional binária gostar/não gostar ou positivo/negativo, ou pode ser mais complexa com normas mais profundas, como um sistema de classificação *Likert* de 1 a 7, medindo fortes sentimentos em relação a problemas comportamentais.

3.2.5.2 Com base em aspectos

A análise de sentimentos baseada em aspecto é um pouco mais profunda. Ele se concentra em descobrir o que os clientes pensam sobre um aspecto ou elemento específico de um produto (como a última atualização de software em um telefone). A análise baseada em faceta facilita o rastreamento de como os clientes percebem as atualizações e os pontos fortes ou fracos específicos das visualizações do cliente.

3.2.5.3 Com base em intenções

A análise de intenção é usada em serviços de suporte ao cliente para simplificar o fluxo de trabalho. Ou seja, identifica a intenção específica por trás da mensagem de alguém.

3.2.5.4 Detecção emocional

A análise de sentimentos baseada em sentimento detecta estados emocionais específicos presentes nas comunicações do cliente com base em algoritmos de aprendizado de máquina e linguagem. Os resultados determinam por que os clientes se sentem de determinada maneira em relação a um produto.

3.2.6 Desafios na análise de sentimentos

Apesar dos progressos alcançados nas tecnologias de processamento de linguagem natural (PLN), compreender a linguagem humana ainda representa um desafio para as máquinas, que podem falhar em capturar as sutilezas mais complexas da comunicação humana. Segundo abordagem de Mukherjee e Bhattacharyya (2013), os principais desafios para a análise de sentimento são:

3.2.6.1 Sentimentos implícitos e sarcasmo

Analisar o sentimento em frases que incorporam elementos de sarcasmo representa um desafio considerável para os computadores. As sentenças de um texto podem transmitir sentimentos de forma implícita, mesmo quando esses sentimentos não são expressos claramente por meio das palavras. Considerando o exemplo a seguir:

– "Sim, ótimo. Demorou três semanas para o meu pedido chegar."

Observa-se que essa sentença não manifesta explicitamente uma conotação negativa por meio de suas palavras, embora de fato contenha essa polaridade. Portanto, a menos que o computador analise a frase com um entendimento abrangente do contexto, ele classificará a experiência como positiva com base na palavra "ótimo".

3.2.6.2 Dependência de domínio

Neste contexto, é essencial compreender que o significado de algumas palavras pode ser alterado de acordo com o contexto em que são empregadas. Um bom exemplo disso seria a seguinte situação:

- "Leia o livro O Caçador de Pipas."

Essa frase possui uma conotação positiva quando considerada no âmbito da leitura de livros, uma vez que representa uma recomendação de leitura. No entanto, quando inserida no contexto de adaptações cinematográficas de obras literárias, essa mesma frase pode adquirir uma interpretação negativa, sugerindo que o diretor do filme está sujeito a críticas desfavoráveis, insinuando que ele precisa ler o livro para criar uma adaptação cinematográfica mais fiel.

3.2.6.3 Expectativas frustradas

Esse cenário envolve situações em que o autor introduz um contexto em sua mensagem, apenas para contrariá-lo no desfecho. Veja o exemplo a seguir:

- "Nossa, essa viagem deveria ser maravilhosa! Encontramos passagens aéreas baratas, conseguimos um excelente hotel e receberíamos um serviço turístico renomado na região. Porém, tudo acabou indo por água abaixo!"

A despeito das expressões agradáveis utilizadas na maior parte do texto, indicando um tom positivo, na realidade, a mensagem expressa um sentimento negativo em relação à situação. Isso ocorre porque a última sentença desempenha um papel crucial na definição da polaridade da mensagem.

3.2.6.4 Pragmática

A pragmática, um campo da linguística que se concentra na análise da linguagem em seu contexto de uso, ou seja, na comunicação, desempenha um papel fundamental e deve ser levada em consideração, uma vez que tem o potencial de transformar completamente a interpretação do interlocutor. Ilustrações dessas nuances podem ser vistas nos exemplos a seguir:

- "Caramba, meu time DESTRUIU o seu no jogo de ontem, hein?!"
- "Estou completamente destruído após um dia inteiro de trabalho!"

É evidente que o uso de letras maiúsculas no verbo "arrasar" na primeira frase reflete um sentimento positivo. Por outro lado, esse mesmo verbo assume um significado negativo na segunda frase, indicando exaustão.

3.2.6.5 Conhecimento de mundo

É fundamental incorporar algum nível de conhecimento contextual aos sistemas que realizam análises de sentimentos. Considere o exemplo a seguir:

- "Ela é uma verdadeira bruxa!"

Esta sentença expressa um sentimento negativo. No entanto, a compreensão desse sentimento requer a aplicação de um conhecimento prévio sobre o significado do termo "bruxa" no contexto da frase.

3.2.6.6 Detecção de subjetividade

Essa técnica envolve a identificação de textos que contêm opiniões e textos que não as contêm. Ela desempenha um papel fundamental na detecção de sentimentos, pois permite criar filtros que separam as mensagens objetivas do conjunto de dados a ser analisado. Considere os exemplos a seguir:

- "Comprei o novo smartphone da Motorola ontem, o Milestone 3."
- "Gostaria de comprar um celular que fosse leve."
- "Odiei o novo celular que ganhei!"

No primeiro exemplo, encontra-se uma declaração objetiva. O segundo exemplo apresenta uma declaração subjetiva, embora o autor não expresse uma opinião positiva ou negativa em relação a algo. Por fim, a terceira frase é subjetiva e expressa uma opinião negativa sobre o celular.

3.2.6.7 Identificação de entidade

Em algumas situações, uma única sentença pode envolver a presença de múltiplas entidades. Frases comparativas são exemplos clássicos desse cenário. Portanto, é crucial identificar a qual das entidades no contexto a opinião está direcionada. É importante considerar os seguintes exemplos:

- "O Motorola Razr é melhor que o iPhone 5."
- "O Vasco foi muito superior ao Flamengo no jogo do Campeonato Brasileiro de ontem."

Os exemplos acima refletem um sentimento positivo em relação ao Motorola Razr e ao Vasco, enquanto expressam um sentimento negativo em relação ao iPhone 5 e ao Flamengo.

3.2.6.8 Negação

Lidar com a presença de negações em uma frase é um dos principais desafios na análise de sentimentos. Isso ocorre porque as negações podem ser expressas de diversas formas, algumas vezes de maneira explícita e outras de forma implícita, tornando necessário avaliar o escopo da negação para determinar se ela de fato está presente.

Em resumo, a negação ocorre quando palavras de sentido negativo são utilizadas para inverter o significado de uma frase. Por exemplo, quando se diz: "Não diria que a assinatura era cara." Algoritmos de análise de sentimentos podem enfrentar desafios na interpretação dessas sentenças, especialmente quando a negação abrange duas frases, como em: "Pensei que a assinatura seria barata. Porém, não foi o caso."

3.2.7 Análise semântica

A análise semântica, na área da ciência da computação, desempenha um papel crucial na compreensão do significado das palavras em textos e informações. Essa abordagem utiliza técnicas de *machine learning* (ML) e processamento de linguagem natural (PLN) para decifrar a relação entre as palavras em um texto e garantir a correção gramatical

das frases. O processo envolve o estudo minucioso do contexto e da estrutura da linguagem, permitindo que os sistemas de IA compreendam o conteúdo e o sentido das informações textuais. Dessa forma, a análise semântica contribui para aprimorar a capacidade das máquinas em compreender e interagir com textos de maneira mais sofisticada e precisa (AMAZON, 2023).

3.2.7.1 Análise de sentimento versus análise semântica

No contexto da comparação entre análise de sentimento e análise semântica, é importante notar que uma solução de análise de sentimentos tem a função de categorizar o texto ao identificar a emoção subjacente. Isso é alcançado através do treinamento de algoritmos de *machine learning* com conjuntos de dados específicos ou da definição de léxicos baseados em regras. Por outro lado, a análise semântica lida com informações mais abrangentes e diversificadas (AMAZON, 2023).

Ambas essas tecnologias linguísticas podem ser integradas de forma sinérgica para auxiliar as empresas a obter uma compreensão mais profunda de seus clientes, proporcionando *insights* valiosos sobre suas necessidades, preferências e sentimentos (AMAZON, 2023).

Partindo para o pressuposto da aplicação da análise de sentimentos, uma das principais aplicações desse recurso é no marketing digital, onde as empresas utilizam a técnica para monitorar a percepção do público sobre suas marcas e produtos. De posse dessas informações, as empresas podem ajustar suas estratégias de marketing e comunicação para melhor atender às expectativas do público.

No entanto, é importante lembrar que a análise de sentimentos tem suas limitações. A interpretação das emoções expressas nos textos pode ser subjetiva e os algoritmos podem ter dificuldade em reconhecer sarcasmo, ironia e outras formas de linguagem figurativa. Além disso, a análise de sentimentos pode ser afetada por preconceitos e estereótipos, levando a resultados imprecisos.

Em conclusão, a análise de sentimentos é uma técnica valiosa para compreender as emoções expressas em textos e pode ser aplicada em vários contextos, desde o

marketing até a política. No entanto, é importante ter em mente suas limitações e usá-lo de forma complementar a outras técnicas analíticas e de pesquisa de mercado.

3.3 PROCESSAMENTO DE LINGUAGEM NATURAL

Embora o processamento de linguagem natural não seja uma ciência nova, a tecnologia está avançando rapidamente devido ao crescente interesse na comunicação homem-máquina, bem como à disponibilidade de *big data*, maior poder de computação e algoritmos aprimorados.

Os humanos contam a possibilidade de falar e escrever em idiomas diferentes como inglês, espanhol ou chinês. Mas a linguagem nativa dos computadores - conhecida como código de máquina ou linguagem de máquina - é bastante incompreensível para a maioria das pessoas. No nível mais profundo do seu dispositivo, a comunicação ocorre não por meio de palavras, mas por meio de milhões de 0s e 1s que produzem ações lógicas.

Os programadores usaram cartões perfurados para se comunicar com os primeiros computadores há 70 anos. Relativamente poucas pessoas entendem esse trabalhoso processo manual. Atualmente, por exemplo, é possível dizer "Alexa, eu gosto dessa música" e um dispositivo contido na própria casa do indivíduo diminuirá o volume e responderá com uma voz humana simulada: "Ok, classificação salva". Em seguida, ajusta o algoritmo para reproduzir essa música, junto com outras favoritas, na próxima vez que você ouvir essa estação.

Um dos componentes mais importantes para análise de sentimento de clientes, classificação de conteúdo e identificação de tendências, é o NLP. Em inglês, a sigla significa *Natural Language Processing*, que traduzido fica Processamento de Linguagem Natural. Algo que pode parecer simples para indivíduos que lidam bem com a tecnologia, mas que pode ser complexo para outras pessoas.

O Processamento de Linguagem Natural (PLN) é uma área da computação que tem como objetivo extrair representações e significados mais completos de textos livres escritos em linguagem natural (INDURKHYA; DAMERAU, 2010). Refere-se por "lingua-

gem natural", a forma mais comum de comunicação, que é utilizada para comunicações do dia-a-dia feitas por humanos, estes que estão acostumados com essa linguagem desde que nascem. Enquanto os computadores utilizam uma linguagem formal para processarem informações (linguagens de programação como Java, Python e Ruby são alguns exemplos), os humanos usam a linguagem natural (PRECIOSO, 2023).

Entende-se por Processamento de Linguagem Natural sendo um subcampo da Inteligência Artificial (IA) que visa permitir que as máquinas entendam, interpretem e respondam à linguagem humana de maneira natural e eficiente. É um campo crescente de pesquisa que usa aprendizado de máquina, mineração de texto e técnicas de análise semântica para extrair informações úteis de dados de texto (MAGALHÃES, 2022).

3.3.1 Como funciona o NLP

O NLP é treinado por meio do aprendizado de máquina, um ramo da inteligência artificial que utiliza grandes quantidades de dados e tempo para treinar algoritmos que podem fazer previsões sobre determinados tópicos. No caso da PLN, as máquinas são treinadas para entender a linguagem humana. Além disso, nenhum conhecimento é desperdiçado no processo e as máquinas ficam mais inteligentes com o tempo (MARTINS, 2022). Normalmente, o processo de NLP começa coletando e preparando texto não estruturado ou fala não estruturados de fontes como *data warehouses* em nuvem, pesquisas, e-mails ou aplicações internas de processos de negócios.

3.3.1.1 Pré-processamentos

A seguir estão as etapas comuns de pré-processamento que uma máquina NLP utiliza.

a) Tokenização

Conforme Premebida (2021), a tokenização é o processo de quebrar sentenças em palavras ou tokens individuais. No processo, a pontuação e os caracteres especiais são completamente removidos. É importante ressaltar que um token não precisa ser apenas uma palavra.

Quando existem palavras compostas, elas podem ter significados completamente

diferentes, como por exemplo: "beija-flor" e "segunda-feira". Em geral, a tokenização é o ato de simplificar um corpus e prepará-lo para as próximas etapas de processamento (PREMEBIDA, 2021).

b) Stemização e Lematização

Stemização e lematização são usados para reduzir as palavras à sua forma raiz, tornando o texto cada vez menos ruidoso. Dependendo de propósito de utilização da NLP, é uma técnica interessante. Um exemplo de utilização dessa técnica seria a aplicação perante palavras "ventania", "ventaninha" e "ventilador", pois todas podem ser convertidas para "vento" (PREMEBIDA, 2021).

3.3.1.2 Processamentos

Mediante o pré-processamento de um texto, torna-se possível realizar alguns processamento na linguagem natural, são eles:

a) Análise de Sentimento

A tarefa de avaliação do sentimento de uma frase, buscando encontrar a conotação do texto.

b) Classificação do Tópico

Ferramenta para encontrar o assunto do texto. A máquina de PNL consegue identificar os assuntos de um documento a partir de diversas técnicas.

c) Detecção de Intenção

O processo de detecção da intenção de um texto é fundamental para garantir uma melhor comunicação e confiança do cliente.

d) *Part-of-Speech Tagging*

O *tagging* ocorre em todas as palavras de um texto, cada token, assim chamado,

tem um valor de verbo, advérbio, substantivo, etc.

e) Reconhecimento de Fala

A conversão de fala para texto é uma das outras tarefas do PNL e só é possível com muito treinamento e aprendizado da máquina, já que temos muitas diferenças nas falas e várias entonações.

f) Reconhecimento de entidade nomeada

Aqui, a tarefa é fazer a máquina compreender qual o significado de uma entidade e até quais são palavras parecidas. Algo como, "Rainha" é parecido com "Menina".

g) Geração de linguagem natural

Como o próprio nome já indica, é a geração de texto automático a partir de uma máquina de processamento de linguagem natural. É usado majoritariamente em chats bots.

3.3.2 Pirâmide do Processamento de Linguagem Natural

As funções da PNL em várias atividades estão associadas a 4 estágios. Por meio deles, as máquinas são capazes de interpretar as comunicações humanas e usá-las para conduzir conversas de maneira eficaz (WENDLER, 2022). A seguir, pode-se destacar as seguintes etapas:

3.3.2.1 Pragmática

A análise lexical desempenha o papel de segmentar uma sentença em múltiplas palavras ou estruturas, tornando a interpretação por parte dos computadores mais acessível. Estas unidades são comumente referidas como "tokens" na área, e, quando interpretadas em conjunto, conferem significado à comunicação, assegurando um processamento eficiente (WENDLER, 2022).

Por exemplo, ao analisar a frase "Não é permitido entrar na loja de chapéu", ela pode ser interpretada de duas maneiras: na primeira, não se permite a entrada em uma loja

usando um chapéu, enquanto na segunda, não é permitida a entrada em uma loja que vende chapéus. Contudo, é evidente que uma das sentenças faz mais sentido do que a outra (MARTINS, 2022).

Portanto, pragmatismo desempenha um papel significativo na compreensão da linguagem natural, e cabe ao desenvolvedor aplicar princípios semelhantes ao desenvolver um software que depende dela.

3.3.2.2 Semântica

A semântica desempenha um papel fundamental na preservação do sentido de uma frase. Esta etapa tem a responsabilidade de estabelecer relações entre as estruturas sintáticas, abrangendo frases, orações e sentenças. Isso possibilita que a máquina analise a totalidade da escrita, discernindo os significados individuais presentes (WENDLER, 2022).

Nessa fase, o contexto em questão demanda que a palavra adira estritamente ao significado registrado no dicionário.

3.3.2.3 Sintaxe

Nesse contexto, é importante lembrar da sintaxe presente nas linguagens de programação, cujo significado é semelhante. No caso, é necessário que se tenha uma frase com a ordem correta das palavras, formando um contexto coerente (MARTINS, 2022).

Durante essa fase, ocorre a identificação e interpretação dos tokens, dando origem às frases utilizadas na comunicação. Para tanto, é implementada uma padronização das estruturas, garantindo que as expressões sejam organizadas em um nível hierárquico que estabelece conexões entre as diversas palavras (WENDLER, 2022).

3.3.2.4 Morfologia

Nesta etapa, há uma complexidade um pouco maior, porém, ela continua sendo crucial para o conjunto. Na entrada de uma palavra, tokens são gerados para seus parágrafos, frases e palavras. Por exemplo, uma palavra como "inquieto" pode ser dividida em dois

subconjuntos de tokens, como "in-quieto", em que "quieto" representa a raiz da palavra, e "in" constitui um afixo (MARTINS, 2022).

3.3.3 Os 4 tipos de abordagem do processamento NLP

O processamento de linguagem natural abrange muito mais do que simplesmente analisar o discurso. Na verdade, engloba uma ampla gama de abordagens para lidar com a linguagem humana. Neste contexto, exploraremos os quatro principais tipos de abordagem no processamento de linguagem natural (NLP).

3.3.3.1 Simbólica

A abordagem simbólica ou racionalista tem suas raízes no campo da Linguística e se baseia na manipulação de símbolos, significados e regras presentes em um texto. Um exemplo simples de método simbólico é o método de Brill (1992), que se aplica à análise léxica, ou seja, à identificação da classe gramatical das palavras em um texto. Tomando como exemplo a frase "João pintou a casa de branco" referenciada por Andreatta (2018), no método de Brill, as palavras são segmentadas e classificadas com base em um dicionário pré-definido, como segue na Figura 1:

Figura 1 – Categorização de Elementos de uma Frase Utilizando o Método Simbólico da NLP - Estado 1

Palavra	João	pintou	a	casa	de	branco
Classificação:		Verbo	Artigo	Substantivo	Preposição	Adjetivo

(ANDREATA, 2018)

É evidente que algumas palavras não puderam ser identificadas, como no caso de "João", e outras foram classificadas de forma incorreta, exemplificado por "branco". Portanto, o método de Brill incorpora duas regras adicionais para a classificação de palavras. A primeira destas regras atribui a classificação de substantivo a todas as palavras desconhecidas que começam com letra maiúscula, como é o caso de "João". A segunda regra associa a palavra desconhecida à mesma classe de palavras que compartilham as três últimas letras. Por exemplo, se a palavra "pintou" não estiver no dicionário, ela seria categorizada como um verbo, com base em outras palavras que terminam com o sufixo "tou".

Figura 2 – Categorização de Elementos de uma Frase Utilizando o Método Simbólico da NLP - Estado 2

Palavra	João	pintou	a	casa	de	branco
Classificação:	Substantivo	Verbo	Artigo	Substantivo	Preposição	Adjetivo

(ANDREATA, 2018)

Após realizar a classificação inicial, apresentado na Figura 2, o método procede a aplicar o conjunto subsequente de regras, que podem ser regras diretas ou derivadas a partir delas:

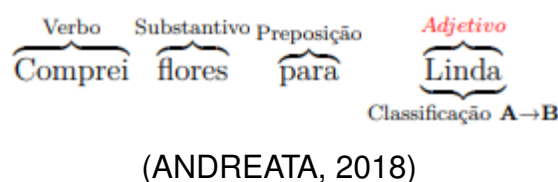
Caso uma palavra inicialmente classificada como **A** esteja inserida no contexto **C**, sua classificação deverá ser ajustada para **B**. A título de ilustração, se uma palavra inicialmente rotulada como **A** (como no caso de "branco" no exemplo) for um adjetivo, e uma das duas palavras anteriores for uma preposição (como "de" no contexto **C**), então a classificação dessa palavra será alterada para a de um substantivo, representada por **B**, como demonstrado na Figura 3.

Figura 3 – Categorização de Elementos de uma Frase Utilizando o Método Simbólico da NLP - Estado 3



Quando uma palavra é inicialmente categorizada como **A** e possui uma característica **P**, sua classificação será modificada para **B**. Por exemplo, se uma palavra **A**, como "Linda," é inicialmente classificada como um adjetivo e apresenta a característica **P** de começar com uma letra maiúscula, sua classificação será ajustada para **B**, indicando que se trata de um substantivo, mediante o apresentado na Figura 4.

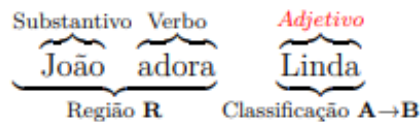
Figura 4 – Categorização de Elementos de uma Frase Utilizando o Método Simbólico da NLP - Estado 4



Se uma palavra é classificada como **A** e está localizada na região **R**, e se uma

palavra nas duas posições anteriores possui a propriedade **P**, então sua classificação será alterada para **B**. Por exemplo, se uma das duas palavras anteriores à palavra "Linda"(no contexto "João adora"na região **R**) começa com letra maiúscula (possuindo a propriedade **P**), a classificação da palavra "Linda"será modificada para substantivo, indicada como **B**, conforme ilustrado na Figura 5.

Figura 5 – Categorização de Elementos de uma Frase Utilizando o Método Simbólico da NLP - Estado 5



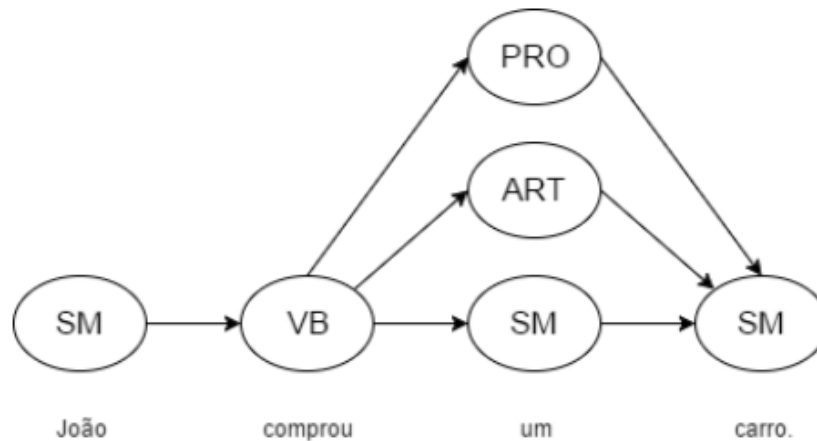
(ANDREATA, 2018)

3.3.3.2 Estatística

Um método estatístico utiliza uma grande quantidade de texto para identificar padrões e associações a modelos, independentemente de esses padrões estarem ou não relacionados a regras sintáticas ou semânticas (MANNING; SCHÜTZ, 1999). Essa abordagem estatística se baseia na implementação de um sistema de aprendizado supervisionado, onde a classificação é realizada com base em um conjunto de dados já previamente classificados, conhecido como conjunto de treinamento (training set). Um exemplo de método estatístico inclui o uso de Modelos de Markov em conjunto com o algoritmo de Viterbi.

Em um Modelo de Markov, conforme Andreatta (2018), a classificação de uma frase, como "João comprou um carro,"é executada utilizando um conjunto de treinamento que pode ser composto, por exemplo, por textos coletados de sites na web, onde as palavras já estão classificadas. Com base nesse conjunto de treinamento, as palavras "João", "comprou"e "carro"são categorizadas como substantivo, verbo e substantivo, respectivamente. No entanto, a palavra "um"apresenta ambiguidade, pois pode ser classificada como artigo (ART), substantivo (SM) ou pronome (PRO). A Figura 6 a seguir ilustra o conjunto de possibilidades gerado pelo classificador para obter uma classificação completa da frase:

Figura 6 – Classificação de Elementos Frasais por Meio do Método Estatístico da NLP - Estado 1



(ANDREATA, 2018)

O cerne da aplicação de Modelos de Markov reside na seleção, a partir dos diversos caminhos disponíveis (conforme ilustrado na figura anterior), daquele que detém a probabilidade mais elevada. Para atingir tal objetivo, é imperativo calcular a probabilidade associada a todos os possíveis caminhos dentro de um Modelo de Markov. Posteriormente, emprega-se o Algoritmo de Viterbi para determinar o caminho que apresenta a probabilidade máxima (MANNING; SCHÜTZ, 1999).

O Modelo de Markov, segundo Andreatta (2018) utilizará o conjunto de treinamento para deduzir a classificação da palavra "um". Para ilustrar, pode-se imaginar um conjunto de treinamento hipotético com as seguintes características:

- a) 10000 substantivos, onde 150 são a palavra "um", 10 são a palavra "João" e 50 são a palavra "carro".
- b) 20000 artigos, onde 500 são a palavra "um".
- c) 12000 verbos, onde 50 são a palavra "comprou".
- d) 15000 pronomes, onde 50 são a palavra "um".

Nesse cenário, a probabilidade de a palavra "um" ser classificada como substantivo é determinada pela Equação 2.1. Isso ocorre porque no conjunto de treinamento, há

150 ocorrências da palavra "um" classificadas como substantivo, de um total de 10.000 substantivos. Portanto, a probabilidade de "um" ser classificada como um substantivo é de 0,015. A Equação 2.1 também é empregada para calcular as probabilidades correspondentes a outras possíveis classes para a palavra "um", como pronome ou artigo. Por exemplo, a probabilidade de "um" ser classificada como um pronome seria 0,0033, enquanto a probabilidade de ser classificada como um artigo seria 0,025.

$$P(\text{palavra} \mid \text{classe}) = \frac{C(\text{classe}, \text{palavra})}{C(\text{classe})} \quad (1)$$

$$P(\text{um} \mid \text{SM}) = \frac{C(\text{SM}, \text{um})}{C(\text{SM})} = \frac{150}{10000} = 0,015. \quad (2)$$

O processo de cálculo de probabilidade é aplicado a todas as palavras presentes na sentença em processo de classificação. Os resultados correspondentes a cada palavra da frase "João comprou um carro" estão apresentados na seguinte tabela:

Tabela 1 – Classificação de Elementos Frasais por Meio do Método Estatístico da NLP - Estado 2

	João	comprou	um	carro
Substantivo	0.001	0	0.015	0.005
Verbo	0	0.0042	0	0
Artigo	0	0	0.025	0
Pronome	0	0	0.0033	0

(ANDREATA, 2018)

Além de avaliar a probabilidade de associação a uma classe específica, também é determinada a probabilidade de transição entre classes. Nesse contexto, é possível supor que o nosso conjunto de treinamento hipotético contenha as seguintes características:

- a) De 20000 frases, 2500 iniciam com um substantivo, 5000 iniciam com um verbo, 5000 iniciam com um artigo e 5000 iniciam com um pronome.
- b) De 10000 substantivos, os 10000 são seguidos por verbos.

- c) De 12000 verbos, 3000 são seguidos por um substantivo, 2000 são seguidos por um outro verbo, 5000 são seguidos por um artigo e 2000 são seguidos por um pronome.
- d) De 20000 artigos, os 20000 são seguidos por um substantivo.
- e) De 15000 pronomes, 10000 são seguidos por um substantivo e 5000 são seguidos por um verbo.

A probabilidade de transição de um verbo para um substantivo é expressa pela Equação 2.2, devido ao fato de o conjunto de treinamento conter 12.000 verbos, dos quais 3.000 são seguidos por um substantivo.

$$P(\text{transicao} \mid \text{classe}) = \frac{C(\text{classe}, \text{transicao})}{C(\text{classe})} \quad (3)$$

$$P(SM \mid VB) = \frac{C(VB, SM)}{C(VB)} = \frac{3000}{12000} = 0,25 \quad (4)$$

De maneira similar, o cálculo da probabilidade de transição é realizado para todas as outras classes. Por exemplo, a probabilidade de transição de um verbo para outro verbo é de 0,17, de um verbo para um artigo é de 0,42 e de um verbo para um pronome é de 0,17.

A Equação 2.2 também é empregada para calcular a probabilidade de a frase iniciar com uma classe específica. Os valores correspondentes à probabilidade de transição para todas as classes presentes no conjunto de treinamento de exemplo estão detalhados na próxima tabela.

Tabela 2 – Classificação de Elementos Frasais por Meio do Método Estatístico da NLP - Estado 3

	Substantivo	Verbo	Artigo	Pronome
Início	0.125	0.25	0.25	0.25
Substantivo	0.0	1.0	0.0	0.0
Verbo	0.25	0.17	0.42	0.17
Artigo	1.0	0.0	0.0	0.0
Pronome	0.67	0.33	0.0	0.0

(ANDREATA, 2018)

Utilizando as probabilidades obtidas por meio do Modelo de Markov, recorre-se ao algoritmo de Viterbi para determinar o caminho mais provável. Esse caminho é identificado com base na equação que virá a seguir, a qual é aplicada a todas as palavras na sentença. Nesta equação, os termos v_t , v_{t-1} , a_{ij} e $b_j(o_t)$ representam, respectivamente, o caminho mais provável atual, o caminho mais provável anterior, a probabilidade de transição e a probabilidade de associação. Por exemplo, para a palavra "João", tem-se que v_{t-1} é igual a 1, uma vez que é a primeira palavra da frase. A_{ij} (a probabilidade de transição entre "Início" e um substantivo) é igual a 0,125, conforme a tabela anterior, e $b_j(o_t)$ (a probabilidade de associação da palavra "João" com um substantivo) é igual a 0,001, conforme a primeira tabela. Assim, o valor de v_t para a palavra "João" é calculado da seguinte forma:

$$v_t(j) = v_{t-1} a_{ij} b_j(o_t) \quad (5)$$

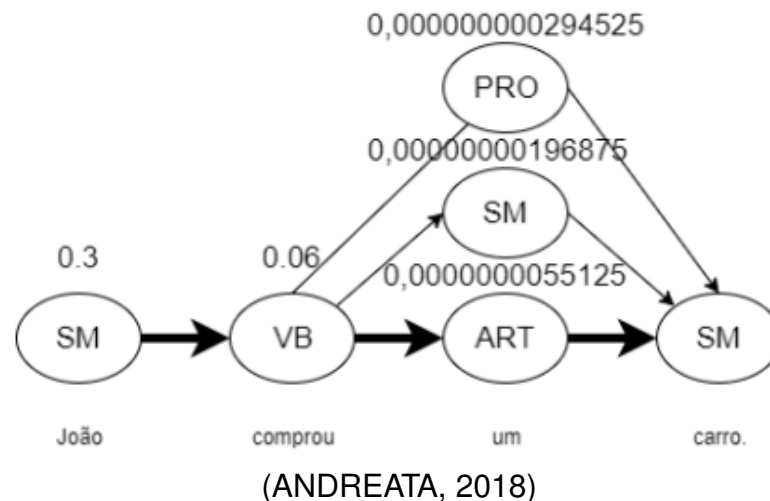
$$v_t(j) = 1 * 0,125 * 0.001 = 0,000125 \quad (6)$$

Quanto à palavra "comprou", obtêm-se 0.000125, que representa o caminho mais provável anterior, conforme a equação precedente. Enquanto isso, os valores 1 e 0.0042 correspondem às probabilidades de transição (conforme a segunda tabela) e de associação (de acordo com a primeira tabela), respectivamente.

$$v_i(j) = 0,000125 * 1 * 0,0042 = 0,000000525 \quad (7)$$

Após calcular as probabilidades de todos os caminhos, seleciona-se o caminho que possui a maior probabilidade, conforme ilustrado na Figura 7. Neste caso, o caminho com a maior probabilidade inclui a palavra "um" classificada como artigo. Esse caminho específico possui uma probabilidade de 0,0000000055125 (calculada como $0,000000525 * 0,42 * 0,025$). Em contraste, o caminho que classifica "um" como pronome tem uma probabilidade de 0,00000000294525 (calculada como $0,000000525 * 0,17 * 0,0033$), enquanto o caminho que considera "um" como substantivo apresenta uma probabilidade de 0,00000000196875 (calculada como $0,000000525 * 0,25 * 0,015$).

Figura 7 – Classificação de Elementos Frasais por Meio do Método Estatístico da NLP - Estado 4



3.3.3.3 Conexionista

A abordagem conexionista, no contexto do processamento de linguagem natural, representa um campo fascinante que se apoia tanto no aprendizado estatístico quanto nas teorias de representação do conhecimento. Conforme Antonio (2021), essa combinação singular de abordagens tem como objetivo criar modelos de linguagem genéricos, proporcionando às máquinas a capacidade de compreender, transformar e inferir a partir de textos de maneira eficaz.

Por meio do aprendizado estatístico, o conexionismo permite que as máquinas identifiquem padrões e relações dentro dos textos. Isso implica a capacidade de extrair informações relevantes de grandes volumes de dados textuais, o que é crucial em tarefas como a análise de sentimentos, tradução automática e processamento de informações em geral. Ao combinar esses insights estatísticos com as teorias de

representação do conhecimento, o conexionismo estabelece uma base sólida para a compreensão da linguagem humana (ANTONIO, 2021).

Um dos resultados notáveis dessa abordagem é a capacidade das máquinas de realizar inferências a partir dos textos. Isso significa que, com base nas informações disponíveis em um texto, as máquinas podem deduzir conclusões lógicas e fazer suposições contextualmente relevantes. Isso é particularmente útil em cenários de busca de informações, onde as máquinas podem aprimorar sua capacidade de resposta com base no contexto e nas informações encontradas no texto (ANTONIO, 2021).

Além disso, segundo Amaral (2021) o conexionismo oferece uma base sólida para a manipulação de texto. Isso inclui tarefas como resumo automático, geração de texto e até mesmo a criação de *chatbots* capazes de responder de maneira coerente a perguntas e interações humanas. Essa capacidade de manipulação de texto torna a abordagem conexionista uma parte vital do desenvolvimento de sistemas de processamento de linguagem natural.

Em resumo, o conexionismo, ao combinar o aprendizado estatístico com as teorias de representação do conhecimento, proporciona às máquinas a capacidade de compreender, inferir e manipular textos, o que é essencial para a evolução contínua da tecnologia de processamento de linguagem natural e sua aplicação em uma ampla gama de domínios, desde assistentes virtuais até análise de dados textuais em larga escala.

3.3.3.4 Híbrida

A abordagem híbrida no campo do processamento de linguagem natural representa um avanço significativo, pois é uma síntese das abordagens mencionadas anteriormente. Essa abordagem combina elementos do processamento simbólico, conexionista e estatístico, proporcionando uma flexibilidade e eficácia notáveis na resolução de uma ampla gama de desafios relacionados ao PLN (AMARAL, 2021).

Ao unir o melhor de todas essas abordagens, a abordagem híbrida se destaca na flexibilidade de lidar com questões linguísticas complexas. Ela permite que as má-

quinas compreendam e processem textos de maneira mais completa e contextual, considerando tanto as regras gramaticais e semânticas quanto as nuances presentes na linguagem natural (AMARAL, 2021).

A eficácia da abordagem híbrida é evidenciada pelo seu poder de lidar com tarefas variadas, desde a análise de sentimentos em redes sociais até a tradução automática de idiomas. Ela representa um avanço notável no campo do PLN, capacitando sistemas a compreender e gerar linguagem de maneira mais próxima do entendimento humano.

Em síntese, a abordagem híbrida no processamento de linguagem natural é uma combinação de abordagens pré-existentes, o que a torna altamente flexível e eficaz na resolução de uma ampla variedade de desafios linguísticos (ANTONIO, 2021). Essa síntese representa um marco importante no avanço da tecnologia de PLN e na sua aplicação em diversos domínios, proporcionando soluções mais sofisticadas e contextualmente relevantes.

3.3.4 Vantagens e desafios do NLP

A ciência do Processamento de Linguagem Natural (NLP) está em constante evolução, mas enfrenta desafios significativos. De acordo com Stoque (2022), um dos principais obstáculos é a análise semântica, pois as expressões da linguagem humana possuem uma variedade de significados, e simplesmente examinar as palavras em uma frase não é suficiente para compreendê-las. O contexto em que as palavras são usadas desempenha um papel fundamental nessa compreensão.

Mesmo com o avanço das técnicas de *deep learning*, a criação de modelos genéricos de NLP que possam ser aplicados a diferentes idiomas continua sendo um desafio. Isso ocorre, em grande parte, devido à falta de dados em certas línguas, o que limita a capacidade de desenvolver soluções amplamente aplicáveis (STOQUE, 2022).

Além disso, a especificidade de cada idioma e as variações na forma de comunicação das pessoas, como velocidade, ênfase, entonação, sotaque e aplicação da gramática, também apresentam desafios para a NLP (STOQUE, 2022).

O campo do Processamento de Linguagem Natural desempenha um papel essencial

na otimização da Engenharia de Dados, aprimorando consideravelmente a capacidade de lidar com dados não estruturados. Dentro desse âmbito, as aplicações do NLP variam desde a avaliação de sentimentos e extração de informações até a criação de assistentes virtuais. Apesar dos desafios que se apresentam, o Processamento de Linguagem Natural oferece vantagens notáveis, conforme menciona Stoque (2022), como:

- Análise eficaz de grandes volumes de dados textuais.
- Estruturação de dados para facilitar a compreensão e a extração de informações.
- Otimização de processos e aumento da produtividade.
- Redução de custos na análise e interpretação de dados.
- Capacidade de obter *insights* em tempo real, o que é valioso em ambientes de negócios e pesquisa.

Portanto, embora a NLP enfrente desafios em seu desenvolvimento contínuo, as recompensas em termos de eficiência e insights valiosos a tornam uma área de pesquisa e aplicação promissora.

3.3.5 Principais termos do NLP

No amplo cenário do Processamento de Linguagem Natural (NLP), há diversos conceitos essenciais que desempenham um papel fundamental na análise, compreensão e geração de linguagem natural. Vamos explorar esses termos-chave: Corpus, Tokenização, Normalização, *N-grams*, *Lexicons* e Pré-processamento e limpeza dos dados. Cada um desses elementos desempenha um papel crucial no desenvolvimento de modelos e sistemas de NLP, tornando possível a interpretação e a interação eficaz com textos.

3.3.5.1 Corpus

A palavra "corpus," derivada do latim, tem como tradução "corpo." Nesse contexto, ela se refere a um conjunto de textos, seja escritos ou orais, abrangendo um ou mais idiomas.

Para representar tal conjunto de textos, utiliza-se o termo "corpora," que representa o plural de "corpus." Essa coleção de textos pode abordar tópicos específicos ou ser de natureza mais abrangente. Alguns exemplos de textos frequentemente utilizados no campo do Processamento de Linguagem Natural (PLN) incluem resenhas de filmes, comentários na internet, avaliações de cursos, críticas de aplicativos online, e-mails, entre outros (PREMEBIDA, 2021).

Um corpus é, essencialmente, uma compilação de documentos, nos quais cada documento pode variar em extensão, representando uma unidade de texto. Esses documentos podem ser tão curtos quanto uma única frase ou tão longos quanto um texto completo. Na prática, um documento pode assumir várias formas, como um comentário em uma rede social, uma ata de reunião ou mesmo um artigo de blog, entre muitos outros exemplos (PRATES, 2019).

3.3.5.2 Tokenization ou Tokenização

A tokenização, às vezes referida como segmentação de palavras, consiste em dividir o texto em partes individuais, identificando os pontos de separação entre palavras, ou seja, onde uma palavra termina e a próxima começa. No contexto da linguística computacional, essas unidades identificadas são comumente denominadas tokens (PALMER, 2010). Nesse procedimento, pontuações e caracteres especiais são inteiramente eliminados.

Há uma distinção fundamental entre abordagens de tokenização em linguagens com espaços delimitadores e aquelas em linguagens não segmentadas. Nas linguagens delimitadas por espaços, como a maioria das línguas europeias, os limites entre palavras são geralmente indicados pela inserção de espaços em branco. No entanto, as sequências de caracteres delimitadas por espaços nem sempre correspondem diretamente aos tokens necessários para processamento adicional, devido à ambiguidade inerente aos sistemas de escrita e à variedade de convenções de tokenização exigidas por diferentes aplicações, como apontado por Palmer em 2010. Por outro lado, nas línguas não segmentadas, como o chinês e o tailandês, as palavras são escritas em sequência, sem indicação clara dos limites entre elas. Portanto, a tokenização em linguagens não segmentadas requer informações léxicas e morfológicas adicionais (PALMER, 2010).

Vale destacar que os tokens não se limitam necessariamente a palavras isoladas. Em casos de palavras compostas, por exemplo, como "beija-flor" e "segunda-feira", as quais podem possuir significados completamente distintos, a tokenização desempenha um papel fundamental. De maneira geral, a tokenização é o ato de simplificar o corpus, tornando-o apto para as etapas subsequentes de processamento (PREMEBIDA, 2021).

Quando se trata da tokenização em linguagens delimitadas por espaço, as principais ambiguidades surgem em relação ao uso de sinais de pontuação, como pontos finais, vírgulas, aspas, apóstrofes e hifens. Isso ocorre porque um mesmo sinal de pontuação pode desempenhar diferentes funções em uma única sentença (PALMER, 2010).

3.3.5.3 *Normalization* ou Normalização

A normalização de textos tem como objetivo aprimorar a precisão da análise, garantindo uma uniformidade no uso de letras maiúsculas e minúsculas. Geralmente, essa normalização é realizada após o processo de tokenização, permitindo a combinação de frases semelhantes, independentemente das variações (PREMEBIDA, 2021).

Um exemplo disso é a distinção entre as palavras "Terra" e "terra", que possuem significados distintos, referindo-se ao planeta e ao solo, respectivamente. No processo de normalização, é possível combinar esses termos em "terra". Cada fase do processo apresenta vantagens e desvantagens. Embora melhore a correspondência durante pesquisas, a alteração entre letras maiúsculas e minúsculas pode impactar a confiabilidade geral da aplicação (PREMEBIDA, 2021).

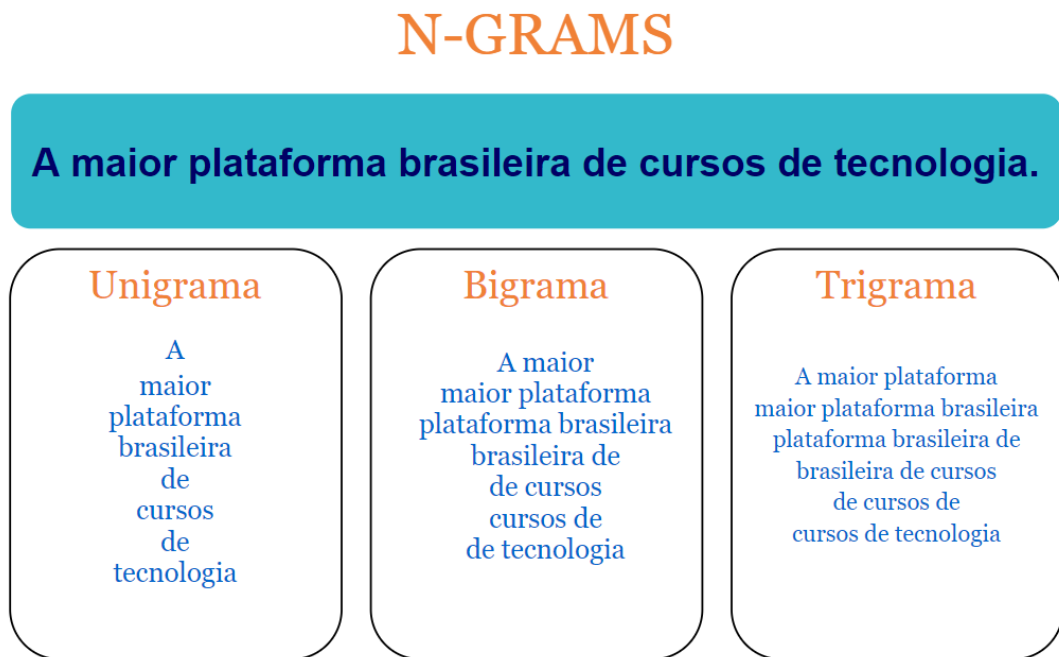
Em resumo, etapa de normalização no campo do Processamento de Linguagem Natural envolve a conversão de todas as palavras para letras minúsculas, removendo acentos e cedilhas. Isso garante que os algoritmos de análise não considerem palavras idênticas como diferentes simplesmente por conta de variações na letra inicial em maiúscula ou erros de acentuação, por exemplo (PRATES, 2019).

3.3.5.4 *N-grams*

N-grams representam modelos probabilísticos utilizados para a previsão do próximo elemento de uma sequência, empregando um modelo de Markov. No contexto linguístico,

esses *n-grams* referem-se a sequências de *n* palavras. Conforme ilustrado na Figura 8 nos exemplos a seguir: "Estudando" é um exemplo de *1-gram* (ou unigrama); na frase "Estudando NLP," observa-se um *2-gram* (ou bigrama); e em "Estudando *Machine Learning*," é possível identificar um *3-gram* (ou trigrama) (PREMEBIDA, 2021).

Figura 8 – *N-grams*



(PREMEBIDA, 2021)

N-grams podem ser úteis para entender relações entre palavras em um texto, identificar padrões e até mesmo ajudar a prever palavras ou preencher lacunas em tarefas de linguagem natural. Eles são uma técnica fundamental em NLP e muitas vezes são usados em conjunto com modelos de aprendizado de máquina para melhorar o desempenho em várias tarefas.

3.3.5.5 *Lexicons* ou Léxicos

Durante a execução de tarefas relacionadas ao Processamento de Linguagem Natural (NLP), é comum que se observe a necessidade de considerar não apenas a própria linguagem, mas também outros fatores. Nesse contexto, os léxicos desempenham um papel significativo. Na esfera da Linguística e no âmbito do NLP, eles constituem uma parte integral da gramática, direcionando o foco para a interpretação do significado das

palavras. Quando palavras são utilizadas em contextos diversos, como na análise de sentimentos em *tweets*, a escolha de expressões coloquiais para descrever situações e objetos pode exercer uma influência substancial sobre os resultados finais. Ignorar tais nuances coloquiais pode comprometer a acurácia e relevância das análises realizadas (PREMEBIDA, 2021).

3.3.5.6 Pré processamento e limpeza dos dados

Antes de prosseguir para quaisquer outras fases, é imperativo realizar o pré-processamento do texto que se pretende utilizar. Nessa etapa, o propósito principal é purificar o texto, eliminando quaisquer elementos disruptivos, como pontuações, caracteres especiais, entre outros, bem como identificando e suprimindo palavras repetidas e aquelas que contribuem com pouco valor à base de dados. Essa fase de aprimoramento do texto é frequentemente conduzida empregando técnicas fundamentais, incluindo o uso de *stop words*, lematização e stemização (*stemming*) (PREMEBIDA, 2021).

Esses termos são fundamentais para quem explora o NLP, já que desempenham um papel crucial na análise, compreensão e geração de linguagem natural, permitindo a construção de modelos e sistemas capazes de interpretar e interagir com textos de forma mais eficaz e precisa.

3.3.6 Principais técnicas usadas no NLP

O Processamento de Linguagem Natural (NLP) é uma área em constante evolução que engloba diversas técnicas e abordagens para compreender, analisar e gerar texto em linguagem humana. Nesse âmbito, um conjunto de técnicas fundamentais desempenha um papel vital. Essas técnicas incluem:

3.3.6.1 *Stop Words*

As "*stop words*", ou palavras de parada, são aquelas que não contribuem significativamente para o conteúdo do texto, uma vez que sua função é principalmente promover a coesão e fornecer contexto, carecendo de significado específico quando consideradas individualmente (PRATES, 2019).

No contexto do Processamento de Linguagem Natural (NLP), uma técnica crucial

consiste na eliminação de ruídos de texto que se mostram menos evidentes do que as pontuações, como os conectivos "que,o,a,de,"e outros semelhantes. Praticamente todos os textos em português contêm tais conectivos e palavras comuns que carecem de significado relevante para o modelo em uso, portanto, é prática comum removê-los (PREMEBIDA, 2021).

Dessa forma, conforme Prates (2019) em análises que empregam o Processamento de Linguagem Natural, é essencial proceder com a exclusão das "*stop words*", as quais são amplamente prevalentes no idioma, mas desempenham apenas um papel de conexão entre frases ou palavras. Estas palavras de parada não acrescentam valor semântico adicional ao texto.

3.3.6.2 *Stemming e Lemmatization*

Dentro do domínio do Processamento de Linguagem Natural (NLP), o *Stemming* e a *Lemmatization* desempenham papéis distintos na redução de palavras para suas formas raiz, com o objetivo de diminuir o ruído nos textos.

O *Stemming* visa a simplificar palavras, reduzindo-as à sua forma raiz. Essa técnica pode ser valiosa, dependendo dos objetivos da aplicação de NLP. Por exemplo, palavras como "correr,correndo"e "correu"podem ser reduzidas a "corre"pela técnica de *stemming* (PREMEBIDA, 2021).

Por outro lado, a *Lemmatization* foca em reduzir palavras à sua forma lematizada, garantindo que a forma resultante seja uma palavra válida e gramaticalmente correta. Para exemplificar, as palavras "melhor,melhores"e "melhorar"seriam todas lematizadas para "bom"(PREMEBIDA, 2021).

Ambas as técnicas têm seu lugar no NLP, sendo escolhidas com base nos requisitos específicos da tarefa em questão.

3.3.6.3 *Bag of words*

No contexto da análise de texto, é uma das técnicas mais amplamente empregadas para realizar a extração de características, com o objetivo de organizar as informações de

forma estruturada. No entanto, os textos não se apresentam como dados estruturados, não seguindo uma organização fixa, como uma tabela. Portanto, é necessário converter os textos em representações numéricas (PREMEBIDA, 2021).

A técnica do "*bag of words*" possibilita a representação do texto com base na frequência de ocorrência de cada palavra, sem considerar a ordem das palavras ou sua estrutura no texto. É como se todas as palavras fossem coletadas e colocadas em um saco, o que permite uma análise numérica das informações textuais (PREMEBIDA, 2021).

3.3.6.4 TF-IDF

A sigla TF-IDF, que representa "*Term Frequency - Inverse Document Frequency*" (Frequência do Termo - Inverso da Frequência no Documento, em tradução livre), é uma técnica estatística que avalia a importância de uma palavra em um texto. Ela guarda semelhanças com a técnica do "*bag of words*", mas apresenta distinções significativas (PREMEBIDA, 2021).

A medida de Frequência do Termo (TF) quantifica com que frequência um termo ocorre em um documento específico, enquanto a medida de Inverso da Frequência no Documento (IDF) avalia a importância desse termo em relação a todo o corpus. Ambas as métricas são consideradas na determinação da relevância de uma palavra. Em termos simples, quanto mais frequente uma palavra é em um documento específico, mais relevante ela tende a ser nesse contexto (PREMEBIDA, 2021).

O TF-IDF é usado em várias aplicações de NLP, como extração de palavras-chave, classificação de texto, pesquisa de informações e recomendação de conteúdo, para destacar e avaliar a relevância de palavras ou termos em um contexto específico.

3.3.6.5 *Word Embeddings*

A incorporação de palavras, um dos principais avanços da aprendizagem profunda no processamento de linguagem natural, revoluciona a representação de palavras e documentos. Essa abordagem mapeia palavras com significados semelhantes para representações similares, facilitando a compreensão do contexto (BROWNLEE, 2019).

Nessa técnica, as palavras são representadas como vetores em um espaço vetorial predefinido, em contraste com as representações esparsas, como a codificação *one-hot*. Cada palavra é associada a um vetor de dezenas ou centenas de dimensões, e essa representação densa e distribuída é aprendida com base no uso de palavras, refletindo naturalmente seus significados (BROWNLEE, 2019).

A base teórica para essa abordagem é a "hipótese distributiva" de Zellig Harris, que sugere que palavras com contextos semelhantes tendem a ter significados semelhantes (BROWNLEE, 2019).

No contexto brasileiro, essa técnica é conhecida como "incorporação de palavras." Ela consiste em representar tokens como vetores numéricos antes de serem usados em modelos de aprendizado de máquina, permitindo que máquinas compreendam a linguagem. Essa técnica enfatiza a semelhança e as relações de sentido entre palavras, permitindo a representação numérica e a incorporação em vetores (PREMEBIDA, 2021).

É importante notar que esses vetores variam de acordo com as características das palavras, mantendo consistência, como distinção entre gêneros. Além disso, a técnica pode ser aplicada a caracteres e frases, e o Word2Vec, desenvolvido pelo Google, é um exemplo de algoritmo eficaz para criar incorporações de palavras, otimizando o treinamento de redes neurais (PREMEBIDA, 2021).

Além do Word2Vec, existem outras diversas técnicas para criar incorporações de palavras, incluindo GloVe (*Global Vectors for Word Representation*), FastText e modelos de linguagem pré-treinados, como o BERT. Essas técnicas ajudam a melhorar o desempenho de tarefas de NLP, tornando-as mais eficazes na compreensão e geração de texto.

3.3.6.6 *Named Entity Recognition (NER)*

O Reconhecimento de Entidades Nomeadas (NER), uma subfunção importante no campo do Processamento de Linguagem Natural (PLN), desempenha um papel crucial na classificação de entidades nomeadas em categorias predefinidas, como nomes de

peças, organizações, locais, códigos médicos, expressões de tempo, quantidades, valores monetários e muito mais (AWAN, 2023). A compreensão dessas entidades é essencial em muitas aplicações do PLN, uma vez que frequentemente contêm as informações mais relevantes nos textos.

No contexto do Processamento de Linguagem Natural, a análise de textos é fundamental, e o Reconhecimento de Entidades Nomeadas (NER) desempenha um papel crucial na extração de informações de um texto e na atribuição de rótulos a categorias pré-definidas, como pessoas, lugares, datas e outros (PREMEBIDA, 2021). O uso do NER aprimora a precisão das análises textuais, garantindo que informações relevantes não se percam quando se depende exclusivamente de técnicas computacionais, como o representado na Figura 9.

Figura 9 – *Named Entity Recognition* (NER)

Local Serviço
A maior plataforma brasileira de cursos de tecnologia.

(PREMEBIDA, 2021)

3.3.6.7 *Parts-of-speech (POS) Tagging*

A marcação de parte da fala (POS), conhecida como "etiquetar partes da fala" em português, é um processo essencial no campo do processamento de linguagem natural (PLN). Envolve a atribuição de rótulos de classe gramatical, como substantivos, verbos, adjetivos e outras categorias, a cada palavra em um texto (HARSHA, 2022).

Essa técnica desempenha um papel fundamental em várias tarefas de PLN, incluindo extração de informações, reconhecimento de entidades nomeadas e tradução automática. Além disso, permite a análise da estrutura gramatical de uma sentença e a desambiguação de palavras que possuem múltiplos significados (HARSHA, 2022).

Normalmente, a marcação de POS é realizada por meio de algoritmos de aprendizado de máquina, que são treinados em grandes conjuntos de dados de texto anotado.

Esses algoritmos aprendem a prever a classe gramatical correta para uma palavra com base no contexto em que ela é usada (HARSHA, 2022).

A saída desse processo consiste em múltiplas tuplas, cada uma contendo a palavra e a sua etiqueta gramatical, como verbo, adjetivo, substantivo, entre outras, o que enriquece a compreensão das estruturas linguísticas no texto (PREMEBIDA, 2021).

Portanto, conforme Premebida (2021), ao realizar o *POS tagging* na frase "A maior plataforma brasileira de cursos de tecnologia", o resultado consistiria em:

Quadro 3.1 – Parts-of-speech

1	<pre>[A, [maior, adjetivo], [plataforma, substantivo], [brasileira, adjetivo], de, [cursos, substantivo], de, [tecnologia, substantivo]]</pre>
---	--

Observa-se que apenas as palavras principais da frase recebem uma *tag*, uma vez que, no processo de eliminação de palavras de parada, o artigo "A" e as duas preposições "de" contidas na frase são removidos (PREMEBIDA, 2021).

Essas técnicas desempenham um papel central na análise e interpretação de texto no contexto do NLP. À medida que a tecnologia avança, a combinação e aplicação dessas técnicas continua a impulsionar o desenvolvimento de aplicações em várias áreas, tornando a interação entre humanos e máquinas mais natural e eficaz.

3.3.7 As 5 principais aplicações do NLP

Embora a inteligência artificial (IA) e o Processamento de Linguagem Natural (NLP) tenham alimentado a imaginação de muitos com inúmeras possibilidades futurísticas, já existem várias aplicações práticas no cotidiano. O NLP, em particular, desempenha um papel crucial em tornar a interação com a tecnologia mais intuitiva e eficaz. Aqui estão cinco das principais aplicações do NLP que já fazem parte do dia-a-dia das pessoas:

a) Filtros de e-mail

Com seu início nos filtros de spam, o Processamento de Linguagem Natural (PNL)

demonstrou seu propósito. Através do aprendizado de palavras e frases específicas, as caixas de entrada conseguem identificar quais mensagens são potenciais *spams* ou não. Um exemplo amplamente conhecido de filtragem de spam é o sistema de e-mail do Google, o Gmail. Nesse sistema, aprimora-se o uso da PNL com maior sofisticação na análise de palavras, frases, endereços de e-mail, e outros aspectos (MARTINS, 2022).

Com base nessa tecnologia, o sistema consegue categorizar e-mails em três divisões distintas: principal, social e promocional, tudo isso por meio da compreensão do conteúdo das mensagens (MARTINS, 2022).

Esses filtros desempenham um papel crucial na organização da caixa de entrada, separando mensagens importantes das que não são relevantes. Eles contribuem para reduzir a sobrecarga de e-mails e aumentar a eficiência na comunicação eletrônica, tornando a experiência de gerenciamento de e-mails mais suave e produtiva.

b) Busca de resultados

A busca de resultados com o auxílio do Processamento de Linguagem Natural (NLP) refere-se à aplicação de técnicas de NLP para melhorar a precisão e a relevância dos resultados obtidos em mecanismos de busca, como motores de busca na web.

A NLP desempenha um importante papel nos mecanismos de busca atuais. Sua principal função é analisar o fluxo de pesquisas com termos semelhantes e sugerir palavras relacionadas para outros usuários (MARTINS, 2022).

O Google, a partir das palavras digitadas pelos usuários, inicia a interpretação e tenta criar hipóteses sobre o que a pessoa possa estar procurando. Além disso, ao analisar frases completas, o mecanismo de busca é capaz de reconhecer com alta precisão a intenção por trás da pesquisa do usuário (MARTINS, 2022).

Em resumo, a aplicação de técnicas de NLP na busca de resultados visa aprimorar a experiência do usuário, garantindo que os resultados sejam mais relevantes,

precisos e adaptados às necessidades individuais. Isso torna a pesquisa na web mais eficaz e satisfatória.

c) Texto preditivo

Mais uma vez, o NLP demonstra excelente desempenho no que diz respeito ao manuseio de palavras e rótulos, permitindo a geração de previsões de grande utilidade. Os sistemas de texto preditivo baseados em Processamento de Linguagem Natural (NLP) são uma inovação tecnológica que visa facilitar e agilizar a comunicação escrita. Eles são projetados para prever e sugerir palavras, frases e até mesmo sentenças inteiras enquanto o usuário digita. Conforme Martins (2022), exemplos comuns dessa capacidade incluem recursos de preenchimento automático, correção ortográfica e até mesmo sugestões de texto encontrados em *smartphones*.

A base para essa funcionalidade é a capacidade de aprender tanto com o usuário como a partir da coleta de dados e informações de diversas pessoas em todo o mundo. Basicamente, o algoritmo aprende com o usuário e é capaz de realizar correções automáticas em palavras e frases (MARTINS, 2022).

Os benefícios dos textos preditivos NLP são muitos. Eles economizam tempo, evitam erros de digitação, ajudam na escrita de mensagens mais claras e, em geral, tornam a comunicação escrita mais eficaz. Além disso, são particularmente úteis em dispositivos móveis, onde a digitação em teclados menores pode ser desafiadora.

Essa tecnologia está presente em uma variedade de aplicativos, incluindo teclados de *smartphones*, aplicativos de mensagens, e-mails e até mesmo em assistentes virtuais. Ela desempenha um papel importante na melhoria da experiência do usuário ao tornar a escrita mais rápida, precisa e acessível.

d) Tradução de idiomas

Quando se discute inteligência artificial, geralmente é associada à tradução de idiomas, uma vez que os tradutores desempenham um papel fundamental no desenvolvimento da PNL.

O princípio é simples: utilizando palavras como rótulos e conjuntos de dados para entradas e saídas, o algoritmo inicia o processo de treinamento e passa a compreender as necessidades dos usuários. Ou seja, ao traduzir um texto de um idioma para outro, é possível incorporar sugestões e receber *feedback* construtivo. Esse processo, por sua vez, contribui para o aprimoramento geral da linguagem utilizada em *websites* (MARTINS, 2022).

Tais sistemas têm tido um impacto significativo na eliminação de barreiras linguísticas e na promoção da compreensão global. Eles são amplamente empregados em serviços online, aplicativos de viagem, comércio internacional e em muitos outros contextos para facilitar a comunicação e a compreensão em diversos idiomas.

e) Análise de dados

A análise de dados baseada em Processamento de Linguagem Natural (NLP) é uma aplicação da inteligência artificial que envolve a extração e interpretação de informações de texto escrito ou falado. Ela permite que as organizações compreendam e utilizem dados de texto de forma eficaz, tornando-os mais acessíveis e úteis para a tomada de decisões.

Profissionais analistas de dados, por exemplo, fazem uso cotidiano dos fluxos integrados de dados proporcionados pelo NLP. O resultado é simples: uma tela com gráficos que exibe todas as informações coletadas (MARTINS, 2022).

Essa integração não apenas aprimora a acessibilidade e a análise em organizações, mas também beneficia o mercado de trabalho e os desenvolvedores de software. Um exemplo disso é que os dados e números apresentados visualmente são considerados mais inteligíveis, proporcionando uma semântica atraente para os usuários. Como resultado, surgem oportunidades adicionais para explorar dados e visualizar históricos de informações relevantes (MARTINS, 2022).

Essa capacidade de extrair *insights* de dados de texto é amplamente utilizada em campos como marketing, atendimento ao cliente, pesquisa de mercado, análise de redes sociais, análise de sentimentos, mineração de dados, entre outros. A análise

de dados NLP permite que as organizações tomem decisões informadas com base em informações anteriormente não estruturadas.

Essa capacidade de extrair *insights* de dados de texto é amplamente utilizada em campos como marketing, atendimento ao cliente, pesquisa de mercado, análise de redes sociais, análise de sentimentos, mineração de dados, entre outros. A análise de dados NLP permite que as organizações tomem decisões informadas com base em informações anteriormente não estruturadas.

Essas são apenas algumas das aplicações do NLP que já fazem parte da vida de muitas pessoas. À medida que a tecnologia avança, espera-se ver mais inovações que tornarão a interação homem-máquina ainda mais natural e eficaz.

3.3.8 3 exemplos de NLPs na prática

A aplicação prática do Processamento de Linguagem Natural (NLP) tem se expandido significativamente nos últimos anos, transformando a maneira como interagimos com a tecnologia e como as empresas coletam *insights* valiosos a partir do texto. Aqui estão três exemplos impressionantes de como o NLP é utilizado no mundo real:

a) Pesquisa inteligente com Klevu

Com a utilização da PNL, as empresas têm à disposição uma alternativa ao preenchimento automático para melhorar suas funcionalidades de pesquisa em seus sites. A empresa Klevu é um exemplo de provedor de pesquisa inteligente impulsionado pela NLP e caracterizado por sua capacidade de aprendizado autodidata. Essa solução é particularmente eficaz no setor de *e-commerce*, pois aprimora sua performance ao observar como os compradores interagem com as funcionalidades de pesquisa na loja (MARTINS, 2022).

Além de oferecer a funcionalidade essencial de pesquisa com preenchimento automático, o Klevu também incorpora automaticamente sinônimos contextualmente relevantes ao catálogo, resultando em uma melhora de até três vezes na profundidade dos resultados da pesquisa. Além disso, o software proporciona uma

experiência de pesquisa personalizada, sugerindo produtos com base na interação anterior dos clientes ou destacando produtos populares (MARTINS, 2022).

Ao utilizar a pesquisa inteligente com Klevu NLP, as empresas podem melhorar a experiência de compra online de seus clientes, tornando mais fácil encontrar os produtos desejados e, assim, aumentar a conversão e a satisfação do cliente. Essa tecnologia é particularmente valiosa para empresas de *e-commerce*, pois ajuda a impulsionar as vendas e a fidelização do cliente por meio de uma busca mais inteligente e eficiente.

b) Tradução de textos com Lilt

A tradução de textos com o Lilt NLP envolve a utilização da tecnologia de Processamento de Linguagem Natural (NLP) oferecida pela plataforma Lilt para aprimorar a qualidade e a eficiência das traduções de texto.

O Lilt é uma ferramenta de tradução que se integra a diversas plataformas, incluindo software de suporte como o Zendesk. Essa integração torna a comunicação através de barreiras linguísticas mais rápida e econômica em comparação com o uso exclusivo de tradutores humanos (MARTINS, 2022).

Desenvolvida por dois ex-engenheiros que contribuíram para o Google Tradutor, a ferramenta não é completamente automatizada, mas opera em conjunto com tradutores humanos, aprimorando sua eficácia ao longo do tempo (MARTINS, 2022).

Empresas e profissionais que buscam traduzir textos de forma eficiente e precisa podem aproveitar o Lilt NLP para atender às suas necessidades de tradução. Essa abordagem híbrida, que combina a automação com a intervenção humana, visa fornecer traduções de alta qualidade para uma ampla gama de idiomas e contextos.

c) Alexa Skills

As "Alexa Skills" são funcionalidades personalizadas e interativas desenvolvidas para a assistente de voz da Amazon, a Alexa. Quando se trata de NLP (Processamento

de Linguagem Natural), as "Alexa Skills" podem incorporar essa tecnologia para melhorar a capacidade da assistente de compreender e responder às solicitações dos usuários de maneira mais natural e contextual.

A Alexa opera de forma semelhante aos bots de mensagens, oferecendo um vasto conjunto de habilidades possíveis. Por exemplo, uma "Alexa Skill" que usa NLP pode ser capaz de responder a perguntas complexas, entender comandos mais elaborados e até mesmo manter uma conversa contínua com o usuário, melhorando a experiência de interação com a assistente de voz. Empresas podem aproveitar essa funcionalidade ao desenvolver suas próprias habilidades, que podem ser integradas aos seus produtos ou serviços baseados em nuvem (MARTINS, 2022).

Em resumo, "Alexa Skills" com NLP elevam a capacidade da Alexa de se comunicar de forma mais natural e aprimoram a usabilidade e a utilidade da assistente em uma variedade de aplicativos e cenários.

A PLN vem evoluindo com o uso de técnicas de aprendizado de máquina, como redes neurais profundas, que permitem o desenvolvimento de modelos de processamento de linguagem natural mais precisos e robustos. Esses modelos são capazes de aprender com grandes conjuntos de dados, o que lhes permite melhorar continuamente a compreensão do texto e os recursos de geração.

Embora os PLNs tenham feito grandes avanços nos últimos anos, ainda há muitos desafios a serem superados. Um dos desafios é entender a linguagem natural em um contexto específico, como o jargão comercial ou o dialeto regional. Outro desafio é entender o sarcasmo e a ironia, que podem levar a interpretações errôneas.

Em conclusão, o processamento de linguagem natural é um campo promissor da inteligência artificial que tem o potencial de mudar a forma como os humanos interagem com as máquinas. À medida que as técnicas de aprendizado de máquina continuam a evoluir, espera-se que a precisão e a eficiência do PLN continuem a melhorar, tornando-o cada vez mais importante para aplicações práticas em vários campos.

4 METODOLOGIA

Este capítulo desempenha um papel crucial ao revelar os bastidores do projeto, esclarecendo as escolhas tecnológicas e delineando os passos que levaram à materialização do sistema proposto.

Neste projeto, desenvolveu-se um sistema avançado que utiliza técnicas de Processamento de Linguagem Natural (PLN) e um modelo de aprendizado profundo para identificar e classificar sentimentos expressos em comentários de redes sociais, com foco especial em avaliações de produtos. A proposta tem como objetivo fornecer recomendações mais personalizadas, considerando não apenas características objetivas, mas também as emoções e opiniões expressas pelos usuários em suas interações sociais. Essa abordagem visa não apenas auxiliar os clientes na realização de compras mais acertivas, mas também permite aos fabricantes identificar tendências e adaptar suas estratégias de produtos de acordo com as preferências do mercado.

A metodologia apresentada não é apenas um conjunto de passos; é uma narrativa estratégica para facilitar a compreensão e oferecer um guia para os interessados no desenvolvimento tecnológico. O foco está na apresentação das tecnologias, padrões cruciais e no processo de construção do software.

Antes de entrar nos detalhes da metodologia, é importante entender os objetivos do sistema. O projeto tem como objetivo auxiliar tanto os clientes, proporcionando-lhes uma experiência de compra mais informada e acertiva, quanto os *e-commerces*, fornecendo ferramentas para a identificação de tendências e insights de mercado. Isso inclui aprimorar a compreensão sobre a aceitação de produtos pelo público, permitindo adaptações estratégicas por parte dos fabricantes para atender às demandas e preferências do mercado de forma mais eficaz.

Para atingir essas metas, foi delineado um caminho estruturado em etapas interdependentes, desde a coleta inicial de dados até a classificação eficiente de novas informações. Destacam-se etapas como coleta de dados, pré-processamento, análise de sentimentos, extração de recursos, treinamento do modelo, avaliação e classificação

de dados emergentes.

Ao descrever essas fases, o objetivo é fornecer um guia claro para a implementação prática e estabelecer um alicerce para a compreensão das escolhas metodológicas. Este capítulo é uma jornada conceitual que busca conectar logicamente os elementos fundamentais para a criação de um sistema robusto.

4.1 FERRAMENTAS E RECURSOS UTILIZADOS

4.1.1 Google API e YouTube Data API

No contexto deste projeto, a obtenção de dados do YouTube é concretizada por meio da implementação da Google API e da YouTube Data API. Essas interfaces, essenciais para a extração de informações relevantes, desempenham um papel crucial ao possibilitar a coleta de dados específicos de vídeos, comentários e outras interações na plataforma.

Para assegurar a legitimidade e a segurança do acesso, é adotada uma chave de API do Google, desempenhando um papel crucial em autenticação e autorização. Esse procedimento é indispensável para verificar a identidade do solicitante, garantindo que apenas usuários autorizados tenham acesso aos dados do YouTube. Essa camada de segurança assume importância especial devido à natureza sensível dos dados e às políticas de privacidade associadas.

Ao integrar de maneira coordenada as APIs da Google e do YouTube, o projeto não apenas busca eficiência no acesso a dados, mas também assegura a conformidade com as diretrizes rigorosas de segurança e privacidade estipuladas pelas plataformas. Nesse contexto, a chave de API do Google atua como um componente central, garantindo a legitimidade das solicitações, controlando o acesso e assegurando que apenas usuários autorizados possam realizar consultas e coletar informações relevantes.

Além disso, a implementação específica das APIs do YouTube Data API adota uma abordagem mais centrada, permitindo a extração direcionada de dados relacionados a avaliações de produtos em vídeos e comentários associados. Esse enfoque não apenas contribui para a precisão e relevância dos dados coletados, concentrando-se

nas informações cruciais para a análise de sentimentos, recomendação e análise de produtos, mas também aprimora a eficácia do processo.

Em síntese, a integração operacional das APIs da Google e do YouTube, juntamente com a utilização da chave de API, constitui a infraestrutura técnica fundamental que facilita o acesso e a coleta de dados essenciais para o desenvolvimento deste projeto. A autenticação robusta proporcionada pela chave de API emerge como um componente vital, assegurando a segurança e legitimidade no processo de obtenção de dados do YouTube. Isso consolida uma base confiável para a análise de sentimentos e o aprimoramento das recomendações e análises de produtos, consolidando, assim, uma base confiável para a análise de sentimentos e o aprimoramento das recomendações e avaliações de produtos.

4.1.2 Preparação dos Dados

Para a análise e recomendação de produtos com base em sentimentos, o conjunto de dados IMDB-pt foi escolhido e adaptado para atender às necessidades específicas dessa tarefa. Originalmente focado em avaliações de filmes, o IMDB-pt foi ajustado e expandido para incluir uma variedade mais ampla de produtos, possibilitando uma análise mais abrangente da experiência do consumidor. O mesmo foi organizado da seguinte forma, conforme Figura 10, para o treinamento do modelo:

Figura 10 – Modelo IMDB-PT

	text_en	text_pt	sentiment
id			
1	Once again Mr. Costner has dragged out a movie...	Mais uma vez, o Sr. Costner arrumou um filme p...	neg

Fonte: Autor

É importante destacar que o conjunto de dados empregado para o treinamento consiste em aproximadamente 49.459 (quarenta e nove mil quatrocentos e cinquenta e nove) registros/frases.

4.1.3 Modelo BERT para Análise de Sentimentos

Dentro do escopo deste projeto, adota-se o modelo BERT (*Bidirectional Encoder Representations from Transformers*) como peça central para a análise de sentimentos em

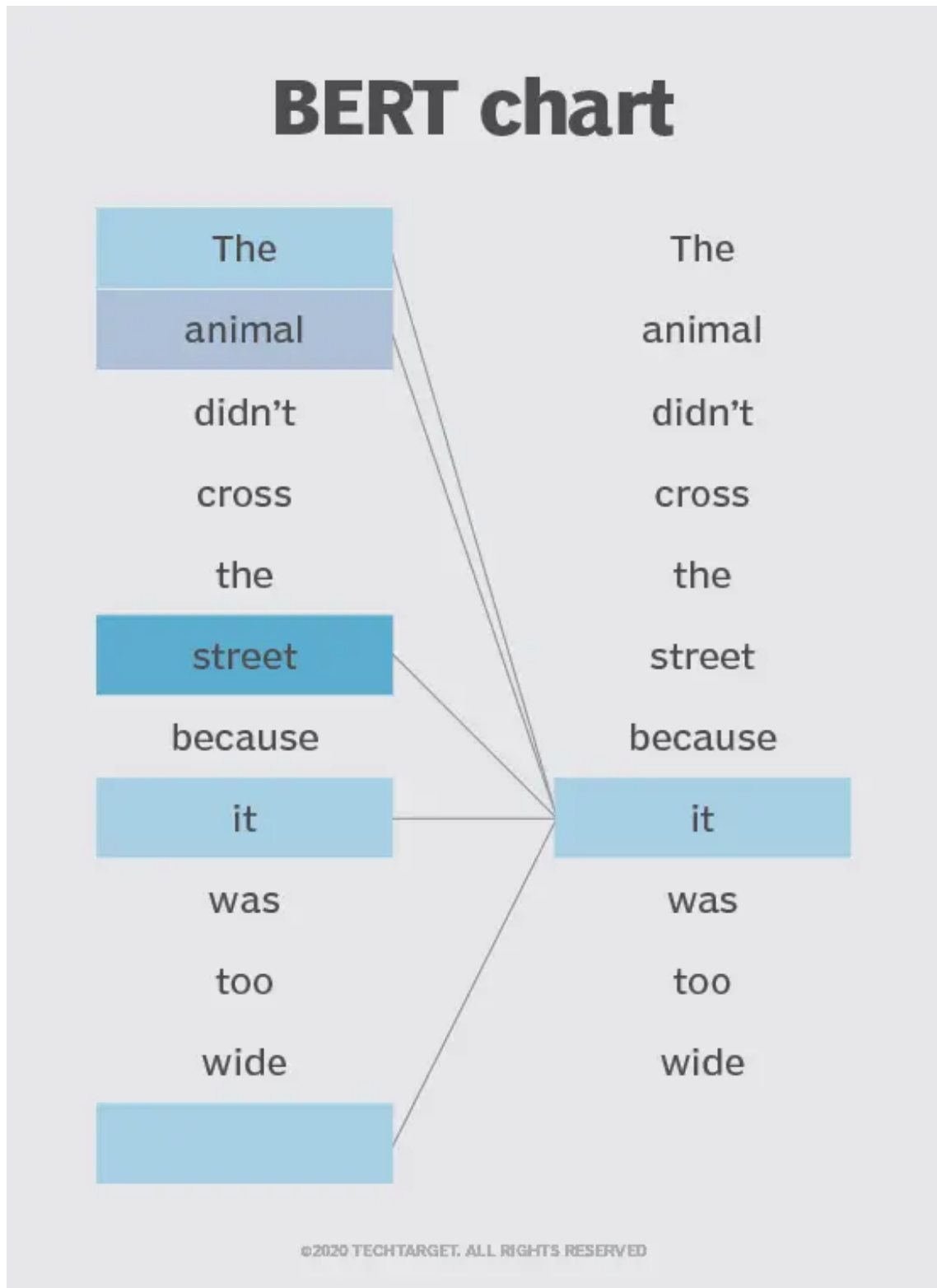
comentários. O BERT, sendo um modelo pré-treinado, demonstra uma eficácia notável ao lidar com a complexidade da língua portuguesa, fornecendo uma compreensão profunda do contexto semântico presente nas postagens textuais analisadas.

O BERT (*Bidirectional Encoder Representations from Transformers*) é um modelo de processamento de linguagem natural desenvolvido pelo Google, notável por sua abordagem bidirecional em relação a outros sistemas unidirecionais. Enquanto modelos tradicionais contextualizam palavras considerando apenas termos à esquerda ou à direita no texto, o BERT analisa o contexto em ambas as direções, proporcionando uma compreensão mais profunda das relações entre palavras e frases (LACERDA, 2020).

Além disso, o fato de o modelo BERT ser pré-treinado em língua portuguesa confere uma vantagem significativa, pois permite que o modelo compreenda a estrutura e a gramática específicas desse idioma. Isso resulta em uma capacidade de análise de sentimentos mais precisa e contextualmente informada, uma vez que o modelo foi ajustado para as particularidades da língua portuguesa.

O BERT é também a primeira técnica de Processamento de Linguagem Natural (PLN) a depender exclusivamente do mecanismo de autoatenção, viabilizado pelos Transformadores bidirecionais no cerne do design do BERT. Isso é notável porque, frequentemente, uma palavra pode mudar de significado à medida que uma frase se desenvolve. Cada palavra adicionada amplifica o significado global da palavra em foco para o algoritmo de PLN. À medida que mais palavras estão presentes em cada sentença ou expressão, a palavra de foco torna-se mais ambígua. O BERT considera esse significado amplificado ao ler bidirecionalmente, levando em conta o efeito de todas as outras palavras em uma sentença sobre a palavra de foco e eliminando o ímpeto da esquerda para a direita que envia as palavras em direção a um significado específico conforme a sentença progride, demonstrado na Figura 11 (LUTKEVICH, 2020).

Figura 11 – Gráfico BERT



(LUTKEVICH, 2020)

O diferencial do BERT também inclui a capacidade de construir um modelo de linguagem eficaz com um corpus de texto relativamente pequeno. Enquanto outros modelos

dependem de grandes volumes de dados para treinamento, a abordagem bidirecional do BERT permite um treinamento mais preciso com menos dados. Após o treinamento inicial em um corpus, como a Wikipédia, o modelo passa por um "ajuste fino" (*fine-tuning*), adaptando-se a tarefas específicas, como perguntas e respostas ou análise de sentimento (LACERDA, 2020).

No contexto do Google, o BERT é aplicado no sistema de busca para compreender as intenções dos usuários e os conteúdos indexados. Ao contrário do RankBrain, o BERT não necessita analisar consultas passadas para entender as intenções dos usuários, pois possui uma compreensão mais holística de palavras, frases e conteúdos. Vale ressaltar que o BERT é uma parte do algoritmo, focando na compreensão semântica, enquanto o restante do algoritmo realiza tarefas como associar buscas às páginas do índice, selecionar os melhores resultados e classificá-los por relevância para os usuários (LACERDA, 2020).

Ao empregar o BERT para a análise de sentimentos em comentários, busca-se não apenas categorizar as expressões emocionais de maneira binária (positiva ou negativa), mas também compreender as nuances e gradações presentes nas opiniões dos usuários. Essa abordagem mais refinada contribui para a criação de um sistema de análise de sentimentos mais sofisticado, capaz de capturar a riqueza emocional presente nas interações online.

Em resumo, a escolha estratégica do modelo BERT nesta aplicação específica busca elevar a qualidade da análise de sentimentos em comentários, proporcionando uma compreensão mais profunda e contextualizada das emoções expressas pelos usuários em língua portuguesa. Essa abordagem robusta é crucial para aprimorar a precisão e relevância da análise de sentimentos, contribuindo para a eficácia global do projeto no contexto da análise personalizada de produtos.

4.1.4 Flask

No desenvolvimento da aplicação web, ressalta-se a utilização do *framework* Flask como uma escolha estratégica. Este busca criar uma interface intuitiva para proporcionar uma interação eficaz com a aplicação. Conhecido por sua flexibilidade e eficiência,

o Flask oferece uma base sólida para o desenvolvimento de uma interface que facilita a interação dos usuários. Sua característica de ser um *framework* leve e modular adiciona uma abordagem pragmática ao desenvolvimento web, permitindo a construção ágil de aplicações robustas.

Destacando-se como a escolha central para o desenvolvimento da aplicação web, o Flask é reconhecido por proporcionar um ambiente de desenvolvimento amigável. Sua flexibilidade é especialmente destacada, possibilitando a adaptação às exigências específicas do sistema. Através do Flask, a implementação eficiente de funcionalidades interativas é facilitada, contribuindo para a criação de uma experiência envolvente e acessível aos usuários.

A versatilidade do Flask não se limita apenas à eficiência do desenvolvimento. Sua integração harmoniosa com outras tecnologias e bibliotecas amplia as oportunidades de aprimoramento da aplicação. Seja na incorporação de elementos visuais, na gestão eficiente de rotas e URLs, ou na integração com bancos de dados, o Flask demonstra ser uma estrutura adaptável que simplifica a expansão e manutenção contínua da aplicação ao longo do tempo.

O foco na criação de uma interface intuitiva para interação com o sistema de recomendação é central para proporcionar aos usuários uma experiência de navegação eficaz e transparente. Isso engloba não apenas a apresentação clara dos resultados de análise de produtos e serviços, mas também a implementação de recursos interativos que facilitam a personalização das preferências, proporcionando uma experiência altamente adaptável e centrada no usuário.

Em resumo, a adoção do *framework* Flask no desenvolvimento da aplicação web visa não apenas garantir uma implementação técnica eficiente, mas também realçar o compromisso em proporcionar uma experiência de usuário aprimorada. Integrando tecnologia e usabilidade, a interface resultante, dinâmica e intuitiva, oferece aos usuários uma interação fluida e agradável com as avaliações personalizadas de produtos.

4.2 COLETA DE DADOS

4.2.1 Busca por Vídeos de Avaliações de Produtos

No decorrer deste procedimento, a exploração dos vídeos associados a avaliações de produtos no YouTube é viabilizada por meio da eficiente YouTube Data API. Essa estratégia proativa possibilita uma busca minuciosa, levando em consideração não apenas os temas gerais, mas também as preferências específicas dos usuários, as quais são expressas durante as consultas de pesquisa.

A execução dessa busca meticulosa por meio da YouTube Data API oferece não apenas uma amplitude de resultados, mas também uma especificidade afinada de acordo com as intenções dos usuários. Ao considerar as preferências expressas durante as pesquisas, o sistema não apenas coleta uma variedade de vídeos relacionados a avaliações de produtos, mas também ajusta os resultados para garantir uma correspondência mais precisa e relevante com as expectativas dos usuários.

Além disso, a YouTube Data API proporciona uma interface dinâmica e adaptável, possibilitando não apenas a busca eficiente, mas também a filtragem e ordenação inteligentes dos resultados. Esse nível de flexibilidade é crucial para fornecer aos usuários uma experiência de pesquisa mais personalizada e alinhada com suas expectativas individuais.

Dentro desse contexto, a utilização estratégica da YouTube Data API não apenas amplia as possibilidades de coleta de dados, mas também eleva a precisão e relevância das informações obtidas. Essa busca por vídeos relacionados a avaliações de produtos torna-se, assim, uma etapa fundamental no processo de análise de sentimentos, proporcionando uma base sólida e abrangente para a eficácia do sistema.

4.2.2 Captura de Comentários

Neste ponto do processo, a captura e o subsequente processamento dos comentários associados aos vídeos identificados desempenham um papel crucial na análise de sentimentos. A implementação estratégica dessa etapa envolve não apenas a coleta de comentários, mas também a seleção criteriosa daqueles expressos em português

brasileiro, visando uma análise mais refinada e contextualmente relevante.

A etapa de captura dos comentários é um componente essencial para a compreensão das reações e opiniões dos usuários em relação aos produtos em questão. Ao centrar-se nos comentários vinculados aos vídeos identificados, o sistema busca obter uma amostra representativa das interações dos usuários, abrangendo uma variedade de perspectivas e experiências.

A decisão de concentrar a análise nos comentários em português brasileiro destaca a consideração cuidadosa das características linguísticas específicas. Isso não apenas aprimora a precisão da análise de sentimentos, mas também garante uma compreensão mais aprofundada das nuances culturais e linguísticas associadas aos comentários, contribuindo para uma interpretação mais precisa das emoções expressas pelos usuários.

Ao abordar a captura e o processamento de comentários de maneira estratégica e contextualizada, essa etapa emerge como uma fonte valiosa de *insights*, proporcionando dados cruciais para a análise de sentimentos e, por conseguinte, para o aprimoramento do sistema de análise personalizada.

4.3 PROCESSAMENTO DE DADOS

4.3.1 Limpeza e Tratamento de Comentários

Na fase crítica do processo de análise de sentimentos, foi adotada uma abordagem abrangente de pré-processamento para otimizar a qualidade dos dados e elevar a precisão da interpretação das emoções expressas pelos usuários. Esse meticuloso processo inclui a remoção de caracteres especiais, a eliminação de espaços adicionais e a normalização, visando purificar os comentários e torná-los mais consistentes e tratáveis.

A remoção de elementos irrelevantes, como caracteres especiais e espaços adicionais, é crucial para assegurar a qualidade da análise, permitindo que a interpretação dos sentimentos seja baseada em dados limpos e coesos. Além disso, a normalização padroniza a estrutura dos comentários, facilitando uma análise uniforme e consistente.

Após esse processo inicial, a etapa de detecção de idioma é implementada com precisão, garantindo que apenas os comentários em português brasileiro sejam considerados para análise subsequente. Essa abordagem aprimorada assegura uma análise de sentimentos contextualmente relevante, concentrando-se nas interações dos usuários nesse idioma específico.

No âmbito deste projeto, a etapa crucial de tokenização é conduzida usando o modelo BERT (*Bidirectional Encoder Representations from Transformers*). Essa escolha é baseada na capacidade avançada do BERT de compreender o contexto à esquerda e à direita de cada palavra, proporcionando uma representação mais rica e precisa das relações semânticas presentes no texto.

A tokenização com o modelo BERT prioriza a captura de nuances e sutilezas nas postagens, essenciais para uma análise de sentimentos mais precisa e abrangente. Essa abordagem visa superar desafios comuns em modelos tradicionais, permitindo uma melhor compreensão de expressões idiomáticas, sarcasmo e outras complexidades linguísticas que influenciam a análise de sentimentos.

Além disso, a preparação cuidadosa dos conjuntos de dados é imperativa, não apenas envolvendo a tokenização, mas também a manipulação adequada dos dados para evitar vieses e garantir uma distribuição equilibrada das classes de sentimentos. Essa etapa é crucial para promover a generalização e a eficácia do modelo diante de dados de entrada diversos.

Assim, ao concluir a etapa de preparação dos dados, são estabelecidas as bases sólidas para uma análise de sentimentos robusta e orientada por contextos, capaz de interpretar de maneira mais fiel as nuances emocionais presentes nas postagens textuais.

4.4 ANÁLISE DE SENTIMENTOS

4.4.1 Classificação com BERT

Em uma etapa crucial do fluxo de trabalho, os comentários previamente processados são submetidos a um processo de classificação de sentimentos, utilizando o modelo

BERT como ferramenta central dessa análise. Esse modelo, renomado por sua eficácia em compreender contextos semânticos complexos, emerge como uma escolha estratégica para decifrar as nuances presentes nos comentários.

A classificação de sentimentos é conduzida de maneira abrangente, onde cada comentário é avaliado em termos de sua probabilidade de ser positivo ou negativo. O modelo BERT, com sua capacidade única de considerar o contexto bidirecional das palavras em uma sentença, possibilita uma análise mais profunda e contextualmente informada das emoções expressas pelos usuários. Esse processo não apenas categoriza os comentários de maneira binária, mas também leva em consideração as gradações e complexidades presentes nas expressões emocionais.

A determinação da probabilidade de um comentário ser positivo ou negativo representa uma medida quantitativa da intensidade do sentimento expresso. Essa abordagem fornece uma compreensão mais refinada das opiniões dos usuários, permitindo uma análise mais granular das emoções veiculadas nos comentários processados. O resultado desse processo não apenas classifica os comentários, mas também fornece uma base sólida para a geração de *insights* sobre a polaridade das opiniões e sentimentos dos usuários.

Ao integrar o modelo BERT nessa etapa crítica da análise de sentimentos, o sistema busca não apenas categorizar os comentários, mas também interpretar as sutilezas e complexidades presentes nas expressões dos usuários. Essa abordagem mais avançada contribui significativamente para a precisão e relevância da análise de sentimentos, proporcionando uma visão mais aprofundada das emoções presentes nos comentários processados.

4.4.2 Limiar de Classificação

Nesta fase, a definição de um limiar de probabilidade torna-se crucial para a categorização dos comentários como positivos ou negativos. Esse limiar estabelece uma fronteira quantitativa, onde os comentários com uma probabilidade acima desse limiar são considerados positivos, enquanto aqueles abaixo são categorizados como negativos. A flexibilidade inerente a essa abordagem permite ajustes conforme necessário,

proporcionando uma adaptação contínua às nuances e dinâmicas das expressões dos usuários.

O estabelecimento do limiar de probabilidade é uma decisão estratégica que influencia diretamente a interpretação das polaridades nos comentários processados. A sensibilidade desse limiar pode ser ajustada para atender aos requisitos específicos do projeto, levando em consideração as preferências e metas da análise de sentimentos. Essa personalização possibilita a adaptação do sistema às particularidades das interações dos usuários, garantindo uma categorização mais alinhada com as nuances expressas nos comentários.

A flexibilidade incorporada na definição do limiar permite uma abordagem dinâmica e adaptativa, onde ajustes podem ser feitos conforme necessário. Essa capacidade de personalização é fundamental para aprimorar a precisão da categorização e garantir que as análises de sentimentos permaneçam alinhadas com as expectativas e requisitos em constante evolução. Assim, a definição do limiar de probabilidade não apenas orienta a classificação dos comentários, mas também oferece uma margem de ajuste que potencializa a eficácia e relevância da análise de sentimentos.

4.5 INTERFACE DE USUÁRIO

4.5.1 Formulário de Pesquisa

Em uma dimensão prática e interativa, a interface do sistema desempenha um papel fundamental, oferecendo aos usuários a capacidade de iniciar o processo de recomendação de maneira intuitiva. A concepção dessa interface foi orientada pela necessidade de proporcionar aos usuários uma experiência de interação eficiente e personalizada. Ao adentrar essa interface, os usuários têm à disposição a funcionalidade de inserir termos de pesquisa e palavras-chave relacionadas a produtos, dando início a um processo de recomendação ágil e centrado nas suas preferências individuais.

A inserção de termos de pesquisa e palavras-chave representa a porta de entrada para a personalização das recomendações. Essa abordagem permite que os usuários expressem suas intenções de forma detalhada, indo além de simples categorias pré-definidas. Ao capacitar os usuários a definir suas preferências de maneira mais granular,

a interface promove uma experiência de recomendação mais precisa e alinhada com as expectativas individuais.

Além disso, a interface foi desenvolvida visando não apenas a eficiência operacional, mas também a clareza e a acessibilidade. A simplicidade na inserção de termos de pesquisa e palavras-chave busca eliminar barreiras à entrada, garantindo que os usuários possam iniciar o processo de recomendação de forma descomplicada e direta. Essa abordagem centrada no usuário contribui para uma interação mais fluida e intuitiva, priorizando a usabilidade e a experiência positiva do usuário.

Em síntese, a interface do sistema assume um papel central ao oferecer aos usuários a capacidade de iniciar o processo de recomendação de maneira fácil e personalizada. Através da inserção de termos de pesquisa e palavras-chave, a interface proporciona um ponto de entrada flexível e informativo, capacitando os usuários a moldar suas recomendações de acordo com suas preferências individuais. Essa abordagem estratégica, alinhada com a usabilidade e a clareza, visa potencializar a interação dos usuários com o sistema de avaliação de produtos.

4.5.2 Apresentação de Resultados

Na fase conclusiva do processo, os resultados obtidos são apresentados de maneira intuitiva na interface do sistema, cuidadosamente projetada para proporcionar uma visão compreensível dos insights gerados. Essa abordagem transparente destaca vídeos relevantes, análise de sentimentos e avaliação de produtos, promovendo uma experiência informativa e acessível para os usuários. A apresentação visual clara e organizada não apenas comunica os *insights* eficazmente, mas também facilita a usabilidade da interface, permitindo que os usuários explorem e compreendam os resultados de maneira eficiente.

Ao finalizar a análise de sentimentos, os resultados são compilados de forma cuidadosa, oferecendo uma visão abrangente da polaridade expressa nos comentários de cada vídeo. Essa compilação quantitativa destaca a proporção entre comentários positivos e negativos, proporcionando *insights* valiosos sobre as percepções dos usuários em relação aos produtos analisados nos vídeos. Essa abordagem não só destaca a

polaridade geral, mas também fornece uma compreensão mais refinada das nuances nas reações dos usuários, contribuindo para a identificação de tendências e padrões na percepção dos produtos. A apresentação dos resultados, tanto visual quanto numericamente, é crucial para uma interpretação rápida e aprofundada dos *insights* gerados pela análise de sentimentos.

4.6 TREINAMENTO DO MODELO

Este tópico oferece uma visão aprofundada do processo essencial no desenvolvimento do projeto em questão. Aqui, o foco recai sobre o treinamento de modelos BERT para análise de sentimentos em textos em português, uma etapa crucial que molda a capacidade do sistema em compreender as nuances emocionais presentes nas expressões linguísticas.

Desde a escolha e preparação cuidadosa do conjunto de dados específico até a avaliação meticulosa do desempenho do modelo, cada passo é delineado de maneira acessível. O código fornecido, baseado nas melhores práticas e na biblioteca Transformers da Hugging Face, serve como um guia prático, refletindo não apenas o caminho técnico, mas também a lógica por trás das escolhas feitas durante o desenvolvimento.

Abaixo estão os detalhes sobre como o código está utilizando e treinando o modelo BERT:

4.6.1 Instalação de Pacotes

Quadro 4.1 – Instalação de Pacotes

1

```
!pip install matplotlib numpy pandas plotly torch scikit-learn  
transformers tqdm
```

O código começa instalando as bibliotecas necessárias, incluindo o PyTorch, Transformers (Hugging Face), e outras bibliotecas para processamento e visualização de dados.

4.6.2 Importação de Bibliotecas

Quadro 4.2 – Importação de Bibliotecas

```

1     import math
2     import os
3     import pickle
4     from dataclasses import dataclass
5
6     import matplotlib.pyplot as plt
7     import numpy as np
8     import pandas as pd
9     import plotly.express as px
10    import torch
11    import torch.nn.functional as F
12    from torch.utils.data import DataLoader
13    from sklearn import metrics
14    from sklearn.model_selection import train_test_split
15    from torch.nn.utils.rnn import pad_sequence
16    from torch.utils.data.dataset import Dataset
17    from transformers import AdamW, BertForSequenceClassification,
18        BertTokenizer, DataCollatorWithPadding
19    from tqdm import tqdm
20    from tqdm.notebook import tqdm as tqdm_notebook

```

Importação de bibliotecas necessárias para manipulação de dados, visualização, PyTorch, Transformers (BERT), etc.

4.6.3 Carregamento e Exploração do *Dataset*

Quadro 4.3 – Carregamento e Exploração do Dataset

```

1     data = pd.read_csv('imdb-reviews-pt-br.csv', index_col=0)

```

Carrega um conjunto de dados CSV contendo avaliações do IMDb em português.

4.6.4 Tokenização do *Dataset*

Quadro 4.4 – Tokenização do Dataset

```

1     tokenizer = BertTokenizer.from_pretrained('neuralmind/bert-base
        -portuguese-cased')

```

```

2     train_dataset = ImdbPt(tokenizer, train['text_pt'], (train['
        sentiment'] == 'pos').astype(int))
3     dev_dataset   = ImdbPt(tokenizer, dev['text_pt'], (dev['
        sentiment'] == 'pos').astype(int))
4     test_dataset  = ImdbPt(tokenizer, test['text_pt'], (test['
        sentiment'] == 'pos').astype(int))

```

Usa o *tokenizer* BERT para tokenizar as sentenças do conjunto de treinamento, desenvolvimento e teste, e cria *datasets* apropriados.

4.6.5 Definição do Modelo BERT

Quadro 4.5 – Definição do Modelo BERT

```

1     model = BertForSequenceClassification.from_pretrained('
        neuralmind/bert-base-portuguese-cased')

```

Inicializa um modelo de sequência de classificação baseado no BERT para análise de sentimentos em português.

4.6.6 Treinamento do Modelo

Quadro 4.6 – Treinamento do Modelo

```

1     optimizer = AdamW(model.parameters(), lr=5e-6)
2     scheduler = torch.optim.lr_scheduler.ExponentialLR(optimizer,
        0.9997)

```

Configura otimizador e programador de taxa de aprendizado.

Quadro 4.7 – Treinamento do Modelo

```

1     for epoch in epoch_bar:
2         for idx, inputs in batch_bar:
3             # ...
4             loss, logits = model(**inputs)[:2]
5             # ...
6         # ...
7     model.save_pretrained('/working/checkpoints/epoch'+str(epoch))

```

Itera sobre épocas e *batches* para treinar o modelo. Salva *checkpoints* periodicamente.

4.6.7 Avaliação do Modelo no Conjunto de Desenvolvimento

Quadro 4.8 – Avaliação do Modelo no Conjunto de Desenvolvimento

```

1     def evaluate(model, dev_loader, device):
2         # ...
3         for inputs in dev_loader:
4             # ...
5         # ...

```

Função para avaliar o desempenho do modelo no conjunto de desenvolvimento durante o treinamento.

4.6.8 Avaliação do Modelo e Visualização das Métricas de Desempenho

Neste estágio, o modelo desenvolvido será submetido a uma avaliação rigorosa utilizando conjuntos de validação e teste. Esta etapa crítica tem como objetivo não apenas medir a performance do modelo, mas também sua eficácia em um conjunto de dados amostrais.

Ao avaliar o modelo nos conjuntos de validação e teste, serão analisadas métricas específicas projetadas para a análise de sentimentos. Essas métricas incluirão, entre outras, a acurácia, a precisão, a revocação e a pontuação F1, proporcionando uma visão abrangente do desempenho em diferentes aspectos da tarefa de classificação de sentimentos. A consideração cuidadosa dessas métricas é fundamental para compreender não apenas a capacidade geral do modelo, mas também sua habilidade de discernir nuances e complexidades emocionais nos dados.

Quadro 4.9 – Avaliação do Modelo e Visualização das Métricas de Desempenho

```

1     # ...
2     with torch.no_grad():
3         # ...
4         fig = px.scatter(
5             # ...
6         fig.show()

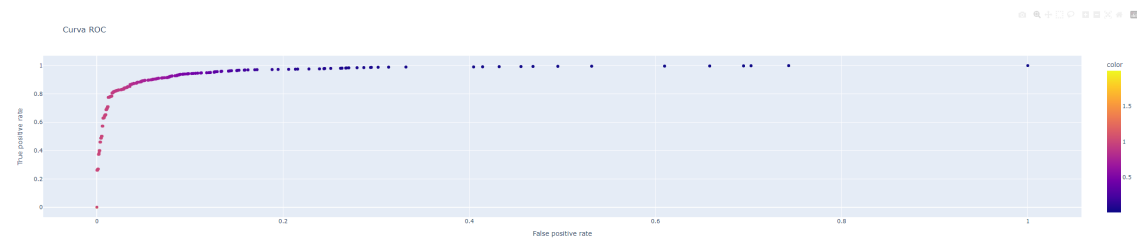
```

7

. . .

Avaliação final do modelo no conjunto de desenvolvimento e visualização da curva ROC (Figura 12), esta que é uma representação gráfica que ilustra o desempenho de um sistema classificador binário à medida que o seu limiar de discriminação varia.

Figura 12 – Curva ROC



Fonte: Autor

E, a seguir, na Figura 13, a acurácia obtida no treinamento:

Figura 13 – Acurácia



Fonte: Autor

A acurácia serve como um indicador abrangente do desempenho do modelo de inteligência artificial encarregado da classificação, permitindo que se avalie com precisão a proporção de classificações corretas em relação ao total de publicações. Com o treinamento adequado, a notável acurácia de 92% na precisão da classificação de sentimentos em frases foi alcançada, conforme evidenciado por STILINGUE (2023), um patamar considerado altamente satisfatório para essa finalidade.

4.6.9 Salvando o Modelo Treinado

Quadro 4.10 – Salvando o Modelo Treinado

1

```
model.save_pretrained(model_save_path)
```

Salva o modelo treinado para uso futuro.

4.6.10 Importação e Uso do Modelo Treinado

Quadro 4.11 – Importação e Uso do Modelo Treinado

```

1     model = BertForSequenceClassification.from_pretrained(
        model_path)
2     tokenizer = BertTokenizer.from_pretrained('neuralmind/bert-base
        -portuguese-cased')
3     # ...

```

Importa o modelo treinado e o *tokenizer* para realizar a classificação de sentimentos em novos textos.

4.6.11 Classificação de Sentimentos em um Novo Texto

Quadro 4.12 – Classificação de Sentimentos em um Novo Texto

```

1     # ...
2     prob_pos = F.softmax(scores, dim=1)[: , 1]
3     # ...
4     print(f"Texto classificado como: {classe_resultante}")
5     print(f"Probabilidade da classe positiva: {prob_pos.item()}")

```

Exemplo de como usar o modelo treinado para classificar o sentimento em um novo texto e calcular a probabilidade associada.

4.7 UTILIZAÇÃO DO MODELO BERT TREINADO

A seguir, será apresentado os detalhes sobre como o modelo BERT treinado é utilizado para análise de sentimentos em vídeos do YouTube:

4.7.1 Importação de Bibliotecas

Quadro 4.13 – Importação de Bibliotecas

```

1     import os
2     import googleapiclient.discovery
3     from langdetect import detect, LangDetectException
4     from googleapiclient.errors import HttpError
5     import re

```



```

6     import torch
7     import torch.nn.functional as F
8     from transformers import BertForSequenceClassification,
        BertTokenizer
9     from flask import Flask, request, render_template

```

Importa as bibliotecas necessárias, incluindo as ferramentas do Google API, a biblioteca Transformers para BERT, e o Flask para criar uma aplicação web.

4.7.2 Carregamento do Modelo BERT Pré-treinado e Tokenizador

Quadro 4.14 – Carregamento do Modelo BERT Pré-treinado e Tokenizador

```

1     model_path = 'Treinamento'
2     model = BertForSequenceClassification.from_pretrained(
        model_path)
3     tokenizer = BertTokenizer.from_pretrained('neuralmind/bert-base
        -portuguese-cased')

```

Carrega o modelo BERT treinado previamente e o tokenizador associado.

4.7.3 Definição de Parâmetros

Quadro 4.15 – Definição de Parâmetros

```

1     max_seq_length = 128

```

Define o comprimento máximo da sequência para o processamento do modelo.

4.7.4 Classe para Representar Resultados de Vídeos

Quadro 4.16 – Classe para Representar Resultados de Vídeos

```

1     class VideoResult:
2         def __init__(self, video_title, video_description,
        total_positivos, total_negativos):
3             self.video_title = video_title
4             self.video_description = video_description
5             self.total_positivos = total_positivos
6             self.total_negativos = total_negativos

```

Define uma classe para armazenar os resultados de cada vídeo, incluindo título, descrição, total de classificações positivas e total de classificações negativas.

4.7.5 Função para Buscar Vídeos de Avaliações de Produtos no YouTube

Quadro 4.17 – Função para Buscar Vídeos de Avaliações de Produtos no YouTube

```

1  def search_product_review_videos(youtube, search_query,
2      maxResults=1000):
3      # ...
      return video_info

```

Utiliza a API do YouTube para buscar vídeos relevantes com base em uma consulta de pesquisa.

4.7.6 Função para Processar e Salvar Comentários

Quadro 4.18 – Função para Processar e Salvar Comentários

```

1  def process_and_save_comments(youtube, video_id,
2      product_keywords, output_file, model, tokenizer,
3      max_seq_length, **kwargs):
4      # ...
      return next_page_token

```

Itera sobre os comentários de um vídeo, processa-os, classifica-os usando o modelo BERT, e salva os resultados em um arquivo.

4.7.7 Configuração da Aplicação Web com Flask

Quadro 4.19 – Configuração da Aplicação Web com Flask

```

1  app = Flask(__name__)

```

Inicializa a aplicação web usando o *framework* Flask.

4.7.8 Rotas da Aplicação Web

Quadro 4.20 – Rotas da Aplicação Web

```

1  @app.route('/', methods=['GET', 'POST'])
2  def index():

```

```

3     # ...
4     return render_template('index.html')
5
6 @app.route('/search', methods=['POST'])
7 def search():
8     # ...
9     return render_template('results.html', video_results=
        video_results, search_query=search_query,
        product_keywords=product_keywords,
        overall_total_positivos=overall_total_positivos,
        overall_total_negativos=overall_total_negativos)

```

Define rotas para as páginas web, incluindo a página inicial e a página de resultados.

4.7.9 Execução da Aplicação Web

Quadro 4.21 – Execução da Aplicação Web

```

1     if __name__ == '__main__':
2         app.run(debug=True)

```

Inicia a aplicação web Flask quando o *script* é executado.

4.7.10 Integração do Modelo BERT no Processamento de Comentários

Quadro 4.22 – Integração do Modelo BERT no Processamento de Comentários

```

1     # ...
2     with torch.no_grad():
3         model.eval()
4         scores = model(**inputs)[0]
5         prob_pos = F.softmax(scores, dim=1)[: , 1]
6     # ...

```

Utiliza o modelo BERT treinado para classificar os comentários e determinar a probabilidade da classe positiva.

4.7.11 Aplicação da Lógica de Classificação

Quadro 4.23 – Aplicação da Lógica de Classificação

```

1   # ...
2   if prob_pos > limiar_probabilidade:
3       classe_resultante = "Positivo"
4   else:
5       classe_resultante = "Negativo"
6   # ...

```

Aplica uma lógica simples para determinar se a probabilidade da classe positiva é suficiente para classificar o comentário como positivo.

4.7.12 Atualização dos Resultados e Exibição na Aplicação Web

Quadro 4.24 – Atualização dos Resultados e Exibição na Aplicação Web

```

1   # ...
2   video_results.append(video_result)
3   # ...
4   return render_template('results.html', video_results=
       video_results, search_query=search_query, product_keywords=
       product_keywords, overall_total_positivos=
       overall_total_positivos, overall_total_negativos=
       overall_total_negativos)

```

Atualiza os resultados e exibe na página da aplicação web.

4.8 SCRIPTS DE PROCESSAMENTO

A utilização dos *scripts* neste código específico está relacionada à implementação da funcionalidade de pesquisa na interface web.

4.8.1 Evento de Submissão do Formulário

- a) `document.getElementById('searchForm').addEventListener('submit', function (event) ...`: O código está associado ao evento de submissão do formulário com o ID 'searchForm'. A função será executada quando o formulário for submetido.

4.8.2 Prevenção do Envio Padrão

- a) `event.preventDefault();`: Impede o envio padrão do formulário, que geralmente resultaria no recarregamento da página.

4.8.3 Captura da Consulta de Pesquisa

- a) `const searchQuery = document.getElementById('searchQuery').value;`: Obtém o valor inserido pelo usuário no campo de pesquisa com o ID 'searchQuery'.

4.8.4 Envio da Consulta para o Servidor

- a) `fetch('/search', ...)`: Utiliza a função `fetch` para enviar uma requisição POST para o endpoint `/search`.
- b) O cabeçalho `Content-Type` é configurado para indicar que o corpo da requisição está no formato JSON.
- c) O corpo da requisição é um objeto JSON contendo a consulta de pesquisa.

4.8.5 Processamento da Resposta do Servidor

- a) `.then(response => response.json())`: Processa a resposta do servidor, assumindo que ela está no formato JSON.

4.8.6 Exibição dos Resultados na Página

- a) `const resultsDiv = document.getElementById('results');`: Obtém o elemento HTML onde os resultados serão exibidos.
- b) `resultsDiv.innerHTML = ''`: Limpa quaisquer resultados anteriores.
- c) `data.forEach(video => ...)`: Itera sobre os resultados recebidos do servidor.

- d) Para cada vídeo, um novo elemento HTML é criado dinamicamente, contendo o título e a descrição do vídeo.
- e) Esse elemento é então anexado ao contêiner de resultados na página.

Em resumo, este script permite uma interação dinâmica do usuário ao enviar consultas de pesquisa para o servidor, processar os resultados e exibi-los instantaneamente na página sem a necessidade de recarregamento.

4.9 UTILIZAÇÃO DA INTERFACE WEB - DETALHAMENTO

Inicialmente, será apresentada uma análise detalhada do arquivo "index.html", o qual desempenha um papel crucial na exibição do formulário de busca ao usuário:

4.9.1 Meta Tags e Título

Quadro 4.25 – Meta Tags e Título

```

1 <meta charset="UTF-8">
2 <meta name="viewport" content="width=device-width, initial-
   scale=1.0">
3 <title>Pesquisa no YouTube</title>

```

Define a codificação da página, configura a visualização para dispositivos móveis e atribui um título à página.

4.9.2 Bootstrap e Estilo

Quadro 4.26 – Bootstrap e Estilo

```

1 <link rel="stylesheet" href="https://stackpath.bootstrapcdn.com
   /bootstrap/4.5.2/css/bootstrap.min.css">
2 <script src="https://code.jquery.com/jquery-3.5.1.slim.min.js"
   ></script>
3 <script src="https://cdn.jsdelivr.net/npm/@popperjs/core@2.5.4/
   dist/umd/popper.min.js"></script>
4 <script src="https://stackpath.bootstrapcdn.com/bootstrap
   /4.5.2/js/bootstrap.min.js"></script>

```

Inclui os arquivos CSS e JavaScript do Bootstrap para estilização e funcionalidades responsivas.

4.9.3 Cor de Fundo e Layout Responsivo

Quadro 4.27 – Cor de Fundo e Layout Responsivo

```
1 <body style="background-color: #f5f5f5;">
2 <div class="container mt-5">
```

Define uma cor de fundo e utiliza o Bootstrap para criar um *container* responsivo com uma margem superior.

4.9.4 Formulário de Pesquisa

Quadro 4.28 – Formulário de Pesquisa

```
1 <form action="/search" method="POST">
2     <!-- Campos do formulário -->
3 </form>
```

Cria um formulário que envia dados para a rota `/search` usando o método POST.

4.9.5 Campos do Formulário

Quadro 4.29 – Campos do Formulário

```
1 <div class="form-group">
2     <!-- Campos de entrada -->
3 </div>
```

Adiciona campos de entrada para o termo de pesquisa, palavras-chave do produto e a quantidade desejada de vídeos.

4.9.6 Botão de Pesquisa

Quadro 4.30 – Botão de Pesquisa

```
1 <button type="submit" class="btn btn-primary d-flex mx-auto">
    Pesquisar </button>
```

Cria um botão de envio estilizado com o Bootstrap.

4.9.7 Estilização Adicional

Quadro 4.31 – Estilização Adicional

```
1 <h1 class="text-center mb-4">Pesquisa no YouTube</h1>
```

Como último passo desse arquivo, é adicionado um título centrado à página para indicar o propósito da aplicação.

Em seguida, apresenta-se a análise detalhada do arquivo "results.html", incumbido de exibir os resultados da aplicação:

4.9.8 Estilo CSS Personalizado

Quadro 4.32 – Estilo CSS Personalizado

```
1 <link rel="stylesheet" href="seu-estilo.css">
```

Inclui um arquivo de estilo CSS personalizado, caso desejado.

4.9.9 Bibliotecas Externas

Quadro 4.33 – Bibliotecas Externas

```
1 <script src="https://cdn.plot.ly/plotly-latest.min.js"></script>
>
```

Inclui referências para as bibliotecas Plotly e Matplotlib.

4.9.10 Dados Dinâmicos e Bootstrap

Quadro 4.34 – Dados Dinâmicos e Bootstrap

```
1 <div class="container mt-5">
2   <h1 class="text-center mb-4">Resultados da Pesquisa</h1>
```

Utiliza dados dinâmicos do Flask e o Bootstrap para criar um layout responsivo.

4.9.11 Informações da Pesquisa

Quadro 4.35 – Informações da Pesquisa

```
1 <div class="mb-4">
```



```

2      <p><strong>Termo de Pesquisa:</strong> {{ search_query }}</
      p>
3      <p><strong>Palavras-chave do Produto:</strong> {{
      product_keywords }}</p>
4  </div>

```

Exibe informações do termo de pesquisa e palavras-chave do produto.

4.9.12 Gráfico Geral

Quadro 4.36 – Gráfico Geral

```

1      <div class="mb-5">
2          <!-- Gráfico geral -->
3          <div id="overall_sentiment_chart"></div>
4      </div>

```

Prepara um espaço para o gráfico geral.

4.9.13 Recomendação Geral

Quadro 4.37 – Recomendação Geral

```

1      <div class="alert {{ 'alert-success' if overall_total_positivos
      >= overall_total_negativos else 'alert-danger' }}" role="
      alert">
2          <!-- Condicional de recomendação -->
3      </div>

```

Exibe uma mensagem de recomendação com base nos totais de comentários positivos e negativos.

4.9.14 Gráficos por Vídeo

Quadro 4.38 – Gráficos por Vídeo

```

1      <h2 class="text-center">Gráficos por Vídeo</h2>
2      <ul class="list-group mt-3">
3          <!-- Loop para resultados de vídeos -->
4          <li class="list-group-item">

```

```

5         <!-- Informa es do v deo e gr fico de barras -->
6     </li>
7 </ul>

```

Utiliza um *loop* para exibir informações e gráficos de barras para cada vídeo.

4.9.15 Scripts para Gráficos

Quadro 4.39 – Scripts para Gráficos

```

1 <script>
2     // Scripts para gr ficos de barras por v deo
3 </script>

```

Inclui *scripts* para criar gráficos de barras dinâmicos usando Plotly.

4.9.16 Script para Gráfico Geral

Quadro 4.40 – Script para Gráfico Geral

```

1 <script>
2     // Script para o gr fico geral
3 </script>

```

Inclui um *script* para criar o gráfico geral de barras.

Essa estrutura HTML utiliza Flask para renderizar dinamicamente os resultados da pesquisa, incluindo informações gerais, gráficos por vídeo e uma recomendação geral com base nos comentários analisados.

Por fim, o código completo dessa aplicação está disponível no GitHub para acesso público e exploração. O repositório pode ser encontrado no link <<https://github.com/williamdstccifes/Avaliacao-de-Produtos-com-Analise-de-Sentimentos-em-Redes-Sociais>>. Essa disponibilidade visa promover a transparência, compartilhar conhecimento e incentivar colaborações na área de análise de sentimentos aplicada à avaliação de produtos em ambientes online. Interessados são encorajados a explorar, fornecer *feedback* ou contribuir para aprimorar ainda mais essa ferramenta valiosa.

5 RESULTADOS

O projeto, desde sua concepção até a implementação, culminou em uma aplicação poderosa e inovadora. Esta proporciona aos usuários a capacidade de analisar a recepção de produtos específicos por meio de avaliações e opiniões expressas em vídeos do YouTube. Neste capítulo, são apresentados os principais resultados obtidos, destacando o valor agregado proporcionado pela análise de sentimentos.

5.1 RESUMO DOS RESULTADOS

5.1.1 Análise de Sentimentos Eficiente

A aplicação demonstrou uma análise de sentimentos eficaz e precisa, permitindo aos usuários obterem uma compreensão profunda da percepção geral sobre produtos específicos. A utilização do modelo BERT pré-treinado em português mostrou-se crucial para interpretar nuances linguísticas e fornecer uma avaliação precisa.

5.1.2 Classificação de Comentários

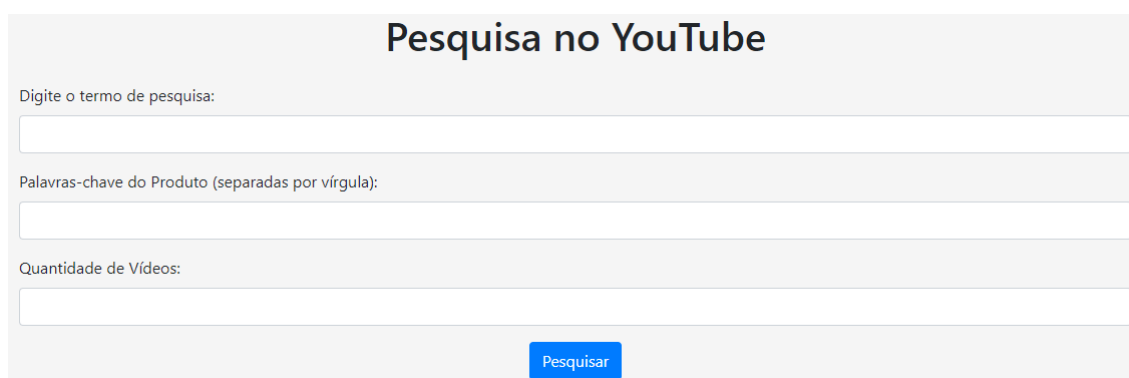
Os algoritmos implementados permitiram a classificação eficiente de comentários como positivos ou negativos, proporcionando uma visão rápida sobre a aceitação ou críticas em relação ao produto pesquisado. Isso possibilita uma tomada de decisão mais informada para empresários e gestores de comércio eletrônico, além de oferecer auxílio aos clientes na avaliação dos produtos antes de efetuarem suas compras.

5.1.3 Interface Intuitiva

A interface web desenvolvida simplificou o processo de pesquisa, tornando-o acessível até mesmo para usuários não técnicos. A facilidade de uso da aplicação amplia sua utilidade, atingindo um público mais amplo.

A seguir, na Figura 14, é apresentada uma imagem ilustrativa que representa a interface da tela de pesquisa conforme exibida ao usuário:

Figura 14 – Formulário de Pesquisa



Pesquisa no YouTube

Digite o termo de pesquisa:

Palavras-chave do Produto (separadas por vírgula):

Quantidade de Vídeos:

Pesquisar

Fonte: Autor

Conforme evidenciado, o usuário é solicitado a inserir informações neste formulário, incluindo o produto desejado para pesquisa, palavras-chave como filtros de busca e, por último, a quantidade de vídeos a serem capturados para análise de comentários.

5.2 CONTRIBUIÇÕES E IMPACTO

5.2.1 Auxílio na Tomada de Decisão

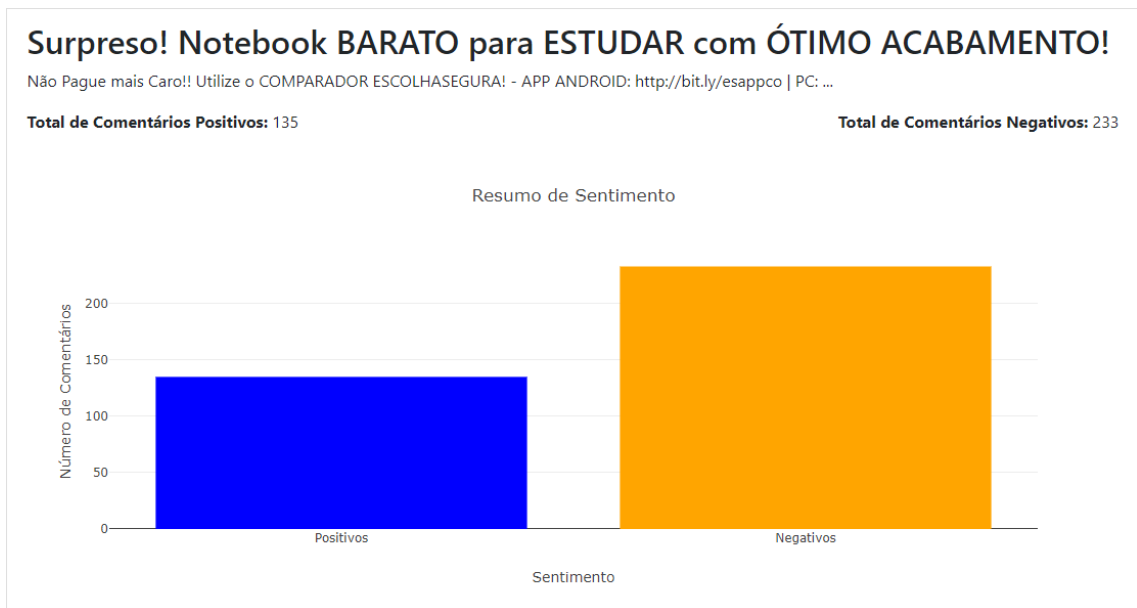
Em seguida, serão apresentadas imagens, Figura 15, Figura 16, Figura 17 e Figura 18, que exemplificam como a avaliação de um produto/serviço é disponibilizada ao usuário após o processamento do formulário enviado anteriormente. Estas imagens incluem os resultados gerais da pesquisa, bem como os resultados específicos de cada vídeo analisado individualmente.

Figura 15 – Resultados Gerais de Pesquisa para o Termo "Notebook Positivo"



Fonte: Autor

Figura 16 – Resultados Específicos de um Vídeo de Pesquisa para o Termo "Notebook Positivo"



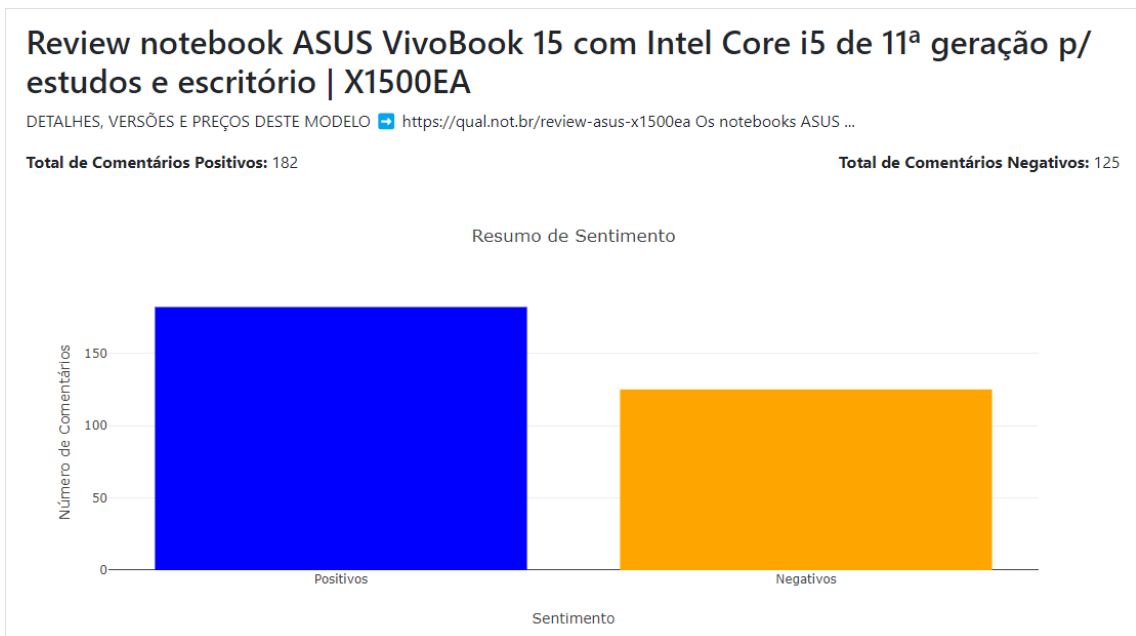
Fonte: Autor

Figura 17 – Resultados Gerais de Pesquisa para o Termo "Notebook Asus"



Fonte: Autor

Figura 18 – Resultados Específicos de um Vídeo de Pesquisa para o Termo "Notebook Asus"



Fonte: Autor

Como pode ser observado, ao fornecer uma visão clara das avaliações de produtos,

a aplicação apoia empresários na tomada de decisões estratégicas. A capacidade de identificar rapidamente se um produto é bem recebido ou não pode ser crucial para o sucesso de uma estratégia de vendas, proporcionando, assim, uma tomada de decisão mais informada para empresários e gestores de comércio eletrônico. Além disso, oferece auxílio aos clientes, permitindo uma compreensão rápida e eficaz sobre a aceitação ou críticas em relação ao produto pesquisado.

5.2.2 Otimização do Investimento e Tomada de Decisão para Empresários, Consumidores e Fabricantes

A análise de sentimentos não apenas oferece *insights* valiosos para empresários, permitindo a otimização de investimentos ao direcionar recursos para produtos mais propensos a obter uma resposta positiva no mercado, mas também beneficia os consumidores. Ao utilizar o sistema, os clientes podem tomar decisões mais informadas antes de comprar produtos, baseando-se nas avaliações e sentimentos expressos por outros usuários. Além disso, as fabricantes de produtos também colhem benefícios, pois a análise de sentimentos fornece informações cruciais sobre a aceitação de seus produtos, permitindo ajustes estratégicos e aprimoramentos contínuos para atender às expectativas do mercado. Dessa forma, a análise de sentimentos contribui para uma tomada de decisão mais eficiente tanto para empresários quanto para consumidores, promovendo a maximização dos investimentos e a satisfação do cliente.

5.2.3 Adaptabilidade a Diferentes Segmentos

A aplicação não se limita a setores específicos, sendo adaptável a uma variedade de produtos e serviços. Isso amplia sua utilidade em diversos nichos de mercado.

5.3 CONCLUSÕES E PERSPECTIVAS FUTURAS

O desenvolvimento bem-sucedido da aplicação de análise de sentimentos marca um avanço significativo na compreensão do mercado. Este capítulo destaca não apenas os resultados obtidos, mas também sugere oportunidades para futuras melhorias e expansões, visando a contínua inovação no cenário do comércio eletrônico.

6 CONCLUSÃO

Ao longo dessa pesquisa, o foco principal foi o desenvolvimento de uma aplicação capaz de avaliar produtos com base na análise de sentimentos extraídos de comentários em vídeos do YouTube. O objetivo era oferecer aos usuários *insights* valiosos por meio de uma abordagem avançada de processamento de linguagem natural e aprendizado de máquina, com destaque para o uso do modelo BERT pré-treinado.

Os resultados obtidos destacam a eficácia da abordagem proposta. A aplicação permite aos usuários realizar buscas por avaliações de produtos específicos, apresentando de forma visual a análise de sentimentos dos comentários associados. Essa classificação oferece uma compreensão clara da recepção do produto pela comunidade, proporcionando aos usuários uma ferramenta valiosa para tomar decisões informadas.

Esta pesquisa contribui significativamente para a área de avaliação de produtos. A integração da análise de sentimentos em comentários de vídeos no YouTube fornece *insights* únicos sobre a percepção do público, impactando positivamente empreendedores, profissionais de marketing, consumidores e fabricantes. A ferramenta oferece uma visão prática e acessível, simplificando a análise da opinião coletiva expressa em plataformas de redes sociais.

É fundamental reconhecer as limitações do estudo. Apesar da eficácia demonstrada, a precisão da análise de sentimentos pode ser aprimorada com ajustes contínuos. Além disso, a aplicação atualmente se concentra em comentários de vídeos no YouTube, sugerindo oportunidades de expansão para outras plataformas de redes sociais.

Em resumo, a aplicação desenvolvida representa uma contribuição significativa para a análise de sentimentos aplicada à avaliação de produtos em ambientes online. A combinação de técnicas avançadas com a acessibilidade de uma interface web a torna uma ferramenta valiosa para diversas áreas, evidenciando a utilidade prática da abordagem proposta. Este estudo não apenas destaca a aplicação, mas também aponta para futuras oportunidades de pesquisa na análise de sentimentos em avaliações de produtos online.

REFERÊNCIAS

- ALBUQUERQUE, C. L. da S. Análise de sentimento sobre comentários em sites de e-commerce no idioma português/br fortaleza 2022. *Sistemas de Informação - Trabalhos de Conclusão de Curso*, 2022.
- ALBUQUERQUE, K. *E-commerce deve crescer 56% no Brasil até 2024*. 2022. Disponível em: <<https://olhardigital.com.br/2022/04/18/pro/e-commerce-crescer-54-brasil-2024/>>.
- ALONSO, G. I.; TONIOLO, L. P.; LOPES, F. S. Análise de sentimento: um estudo de performance entre comentários e vendas de produtos em um e-commerce pautado em text mining. *Sistemas de Informação - TCC - FCI Higienópolis*, v. 85, 2020.
- AMARAL, B. *O que é NLP: guia sobre o processamento de linguagem natural*. 2021. Disponível em: <<https://weni.ai/blog/o-que-e-nlp/>>.
- AMAZON. *O que é análise de sentimentos?* 2023. Disponível em: <<https://aws.amazon.com/pt/what-is/sentiment-analysis/>>.
- ANDREATA, G. H. S. O uso de processamento de linguagem natural para a análise de sentimentos na rede social reddit. 2018.
- ANTONIO, A. M. *Inteligência Artificial: O que é Processamento de Linguagem Natural (PLN)?* 2021. Disponível em: <<https://amartins-antonio.medium.com/intelig%C3%Aancia-artificial-o-que-%C3%A9-processamento-de-linguagem-natural-pln-f7904e68af1b>>.
- ARANHA, C.; PASSOS, E. A tecnologia de mineração de textos. *RESI-Revista Eletrônica de Sistemas de Informação*, n. 2, 2006.
- AUGUSTINE, E. et al. Outage detection via real-time social stream analysis: leveraging the power of online complaints. *WWW '12 Companion: Proceedings of the 21st International Conference on World Wide Web*, p. 13–22, 2012.
- AWAN, A. A. *What is Named Entity Recognition (NER)? Methods, Use Cases, and Challenges*. 2023. Disponível em: <<https://www.datacamp.com/blog/what-is-named-entity-recognition-ner>>.
- BACCIANELLA, S.; ESULI, A.; SEBASTIANI, F. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. *Proceedings of the International Conference on Language Resources and Evaluation (LREC'10)*, p. 2200–2204, 2010.
- BARION, E. C. N.; LAGO, D. A tecnologia de mineração de textos. *Revista de Ciências Exatas e Tecnologia*, v. 3, n. 3, p. 123–140, 2008.
- BRILL, E. A simple rule-based part of speech tagger. *ANLC '92: Proceedings of the third conference on Applied natural language processing*, p. 152–155, 1992.
- BROWNLEE, J. *What Are Word Embeddings for Text?* 2019. Disponível em: <<https://machinelearningmastery.com/what-are-word-embeddings/>>.

- CAMBRIA, E. et al. New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems*, v. 28, n. 2, p. 15–21, 2013.
- CASTRO, M. Redes sociais forçam empresas a responder rápido. 2013.
- CHRISTINO, C. *Satisfação de clientes: como realizar seu monitoramento?* 2021. Disponível em: <<https://blog.softexpert.com/satisfacao-clientes/>>.
- DAVIDOV, D.; TSUR, O.; RAPPOPORT, A. Enhanced sentiment learning using twitter hashtags and smileys. *COLING '10: Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, p. 241–249, 2010.
- E-COMMERCE. *A importância das Redes Sociais para e-commerce*. 2023. Disponível em: <<https://blog.saninternet.com/redes-sociais-para-e-commerce>>.
- EVANGELISTA, T. R.; PADILHA, T. P. P. Monitoramento de posts sobre empresas de e-commerce em redes sociais utilizando análise de sentimentos. *BRAZILIAN WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING (BRASNAM)*, v. 3, p. 152–163, 2014.
- GOMES, P. C. T. *Análise de Sentimentos com Machine Learning*. 2019. Disponível em: <<https://www.datageeks.com.br/analise-de-sentimentos/>>.
- GRASSI, M. et al. Sentic web: A new paradigm for managing social media affective information. *Cognitive Computation*, v. 3, n. 3, p. 480–489, 2011.
- HARSHA, A. *Understanding Part-of-Speech Tagging in NLP: Techniques and Applications*. 2022. Disponível em: <<https://www.shiksha.com/online-courses/articles/pos-tagging-in-nlp/>>.
- HASSAN, A.; QAZVINIAN, V.; RADEV, D. What's with the attitude?: identifying sentences with attitude in online discussions. *EMNLP '10: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, p. 1245–1255, 2010.
- INDURKHYA, N.; DAMERAU, F. J. *Handbook of Natural Language Processing*. [S.l.]: Chapman Hall/CRC, 2nd edition, 2010.
- KONTOPOULOS, E. et al. Ontology-based sentiment analysis of twitter posts. *Expert Systems with Applications*, v. 40, n. 10, p. 4065–4074, 2013.
- LACERDA, L. *Tudo sobre o BERT: o novo algoritmo do Google que promete revolucionar as SERPs*. 2020. Disponível em: <<https://rockcontent.com/br/blog/bert/#5>>.
- LIU, B. Sentiment analysis and subjectivity. *Handbook of Natural Language Processing*, v. 2, n. 2010, p. 627–666, 2010.
- LUTKEVICH, B. *BERT language model*. 2020. Disponível em: <<https://www.techtarget.com/searchenterpriseai/definition/BERT-language-model>>.
- MAGALHÃES, W. *Processamento de linguagem natural: entenda o que é PLN*. 2022. Disponível em: <[https://www.remessaonline.com.br/blog/processamento-de-linguagem-natural/#:~:text=natural%20\(PLN\)%3F-,O%20processamento%20de%20linguagem%20natural%20\(PLN\)%20%C3%A9%20uma%20sub%20%C3%A1rea%20da,estruturados%20e%20baseados%20em%20textos.>](https://www.remessaonline.com.br/blog/processamento-de-linguagem-natural/#:~:text=natural%20(PLN)%3F-,O%20processamento%20de%20linguagem%20natural%20(PLN)%20%C3%A9%20uma%20sub%20%C3%A1rea%20da,estruturados%20e%20baseados%20em%20textos.>)>

MANNING, C.; SCHÜTZE, H. *Foundations of Statistical Natural Language Processing*. [S.l.]: The MIT Press, 1999.

MARTINS, V. *NLP: O guia completo sobre Processamento de Linguagem Natural!* 2022. Disponível em: <<https://blog.betrybe.com/tecnologia/nlp-processamento-de-linguagem-natural/>>.

MENA, D. et al. Estratégias de marketing digital em empresas e-commerce. *593 Digital Publisher CEIT, ISSN-e 2588-0705*, v. 4, n. Extra 5-1, p. 108–122, 2019.

MUKHERJEE, S.; BHATTACHARYYA, P. Sentiment analysis : A literature survey. *Indian Institute of Technology, Bombay*, 2013.

NASCIMENTO, P.; OSIEK, B.; XEXÉO, G. Análise de sentimento de tweets com foco em notícias. *Handbook of Natural Language Processing*, v. 14, n. 2, p. 627–666, 2012.

OLIVEIRA, F. R. de. Modelos lineares e não lineares aplicados à análise de sentimentos de consumidores de e-commerce no Brasil. *Research Gate*, 2020.

OLIVEIRA, F. R. de. Modelos lineares e não lineares aplicados à análise de sentimentos de consumidores de e-commerce no Brasil. *Technical Report*, 2020.

PALMER, D. D. Text preprocessing. *Handbook of Natural Language Processing, Second Edition*, p. 9–30, 2010.

PRATES, W. R. *Introdução ao Processamento de Linguagem Natural (NLP)*. 2019. Disponível em: <<https://cienciaenegocios.com/processamento-de-linguagem-natural-nlp/>>.

PRECIOSO, V. *Linguagem natural: entenda o que é e a importância na era de chatbots*. 2023. Disponível em: <<https://www.cedrotech.com/blog/linguagem-natural-entenda-o-que-e-e-a-importancia-na-era-de-chatbots/>>.

PREMEBIDA, S. M. *Guia de NLP - conceitos e técnicas*. 2021. Disponível em: <<https://www.alura.com.br/artigos/guia-nlp-conceitos-tecnicas>>.

RODRIGUES, M. B.; FLÁVIO, V. V. S.; NETO, G. H. Desenvolvimento e implementação do workflow para realizar a análise de sentimentos como uma possível estratégia comercial. *Revista Eletrônica de Computação Aplicada*, v. 2, n. 1, p. 185–204, 2021.

SANTOS, L. M. Protótipo para mineração de opinião em redes sociais: Estudo de casos selecionados usando o twitter. *Departamento de Ciência da Computação, Universidade Federal de Lavras*, 2010.

STILINGUE. *Como funciona a análise de sentimento da STILINGUE?* 2023. Disponível em: <<https://help.blip.ai/hc/pt-br/articles/19066118945175-Como-funciona-a-an%C3%A1lise-de-sentimento-da-STILINGUE->>.

STOQUE. *Processamento de Linguagem Natural (NLP): o que é e como pode ser aplicada aos negócios*. 2022. Disponível em: <<https://stoque.com.br/dpa/processamento-de-linguagem-natural/>>.

SULZ, P. *O guia completo de Redes Sociais: saiba tudo sobre as plataformas de mídias sociais!* 2020. Disponível em: <<https://rockcontent.com/br/blog/tudo-sobre-redes-sociais/>>.

TIBCO. *O que é análise de sentimentos?* 2022. Disponível em: <<https://www.tibco.com/pt-br/reference-center/what-is-sentiment-analysis>>.

TURNEY, P. D. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. *ACL '02: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, p. 417–424, 2002.

WENDLER, S. *O que é NLP (Processamento de Linguagem Natural), seus usos e como utilizar.* 2022. Disponível em: <<https://rockcontent.com/br/blog/o-que-e-nlp/>>.

ZENDESK. *Qual a importância das redes sociais para as empresas e como seus resultados?* 2022. Disponível em: <<https://www.zendesk.com.br/blog/qual-a-importancia-das-redes-sociais-para-as-empresas/>>.