

UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ENGENHARIA DE SÃO CARLOS
DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO

VICTOR ROSSIN TAVARES

**Utilização de Análise de Sentimentos Baseada em Aspectos não
supervisionada para a captura da percepção de valor pelo cliente: uma
aplicação no contexto de smartphones**

São Carlos
2021

VICTOR ROSSIN TAVARES**Utilização de Análise de Sentimentos Baseada em Aspectos não supervisionada para a captura da percepção de valor pelo cliente: uma aplicação no contexto de smartphones**

Trabalho de Conclusão de Curso (TCC) apresentado à Escola de Engenharia de São Carlos da Universidade de São Paulo, como requisito para a obtenção do título de Engenheiro de Produção.

Orientador: Prof. Titular Luiz Cesar Ribeiro Carpinetti

Coorientador: Lucas Gabriel Zanon

São Carlos

2021

AUTORIZO A REPRODUÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pela Biblioteca Prof. Dr. Sérgio Rodrigues Fontes da EESC/USP com os dados inseridos pelo(a) autor(a).

Rossin, Victor

R835u Utilização de Análise de Sentimentos Baseada em Aspectos não supervisionada para a captura da percepção de valor pelo cliente: uma aplicação no contexto de smartphones / Victor Rossin; orientador Luiz Cesar Ribeiro Carpinetti; coorientador Lucas Gabriel Zanon. São Carlos, 2021.

Monografia (Graduação em Engenharia de Produção) -- Escola de Engenharia de São Carlos da Universidade de São Paulo, 2021.

1. Costumer Value. 2. Análise de Sentimentos. 3. Análise de Sentimentos Baseada em Aspectos. 4. Redes Sociais. 5. Twitter. 6. Smartphones. 7. Tomada de Decisão. I. Título.

FOLHA DE APROVAÇÃO

Candidato: Victor Rossin Tavares
Título do TCC: Utilização de Análise de Sentimentos Baseada em Aspectos não supervisionada para a captura da percepção de valor pelo cliente: uma aplicação no contexto de smartphones
Data de defesa: 14/12/2021

Comissão Julgadora	Resultado
Professor Titular Luiz Cesar Ribeiro Carpinetti (orientador)	APROVADO
Instituição: EESC - SEP	
Professor Associado Daniel Capaldo Amaral	APROVADO
Instituição: EESC - SEP	
Doutorando Lucas Gabriel Zanon	APROVADO
Instituição: EESC - SEP	

Presidente da Banca: **Professor Titular Luiz Cesar Ribeiro Carpinetti**

RESUMO

ROSSIN, V. Utilização de Análise de Sentimentos Baseada em Aspectos não supervisionada para a captura da percepção de valor pelo cliente: uma aplicação no contexto de smartphones. 2021. 198 f. Monografia (Trabalho de Conclusão de Curso) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2021.

A evolução das redes sociais, impulsionada pela internet e pela difusão dos *smartphones*, cada vez mais tem mostrado a relevância da opinião dos consumidores como uma ferramenta para a tomada de decisões estratégicas em grandes empresas. Os conceitos de cliente no centro e valor do cliente já são citados na literatura desde o século XX. Com o expressivo volume de opiniões desses usuários sendo publicadas constantemente nas redes sociais, as empresas têm interesse em interpretar *streaming big data* para captar tendências e percepções. A Análise de Sentimentos é uma forma rápida e eficiente de processar esses dados continuamente gerados e proporcionar a obtenção de conhecimento útil a partir deles. A Análise de Sentimentos consiste em atribuir um rótulo emocional ao texto, os quais indicam uma polaridade se o texto expressa algo positivo ou negativo de forma geral. Por sua vez, a Análise de Sentimentos Baseada em Aspectos permite maior granularidade, a nível de aspectos e não no nível do texto com um todo. Diante disso, o objetivo desse trabalho é desenvolver um código em Python capaz de minerar dados do Twitter, pré-processá-los, realizar a extração dos aspectos e em sequência aplicar a análise de sentimento baseada nos aspectos extraídos. O modelo computacional desenvolvido utiliza uma abordagem não supervisionada, ou seja, não é necessário um conjunto de dados pré-rotulados, ampliando as possibilidades de aplicação. A partir desse código foi feita uma aplicação piloto para analisar a percepção de valor sobre dois *smartphones* de empresas concorrentes, sendo eles Iphone 13 da Apple e Galaxy S21 da Samsung. Através dos resultados obtidos foi realizada uma análise comparando as percepções dos clientes entre os celulares, cruzando-as com as especificações técnicas dos produtos e com as estratégias adotadas pelas empresas. Além disso, foi realizada uma análise SWOT com o propósito de prover *insights* para a tomada de decisão, entendendo melhor os pontos fortes e fraquezas da atual geração de *smartphones*, ademais identificar possíveis oportunidades e ameaças.

Palavras-chave: Costumer Value, Análise de Sentimentos, Análise de Sentimentos Baseada em Aspectos, Redes Sociais, Twitter, *Smartphones*, Tomada de Decisão.

ABSTRACT

ROSSIN, V. **Applying Aspect Based Sentiment Analysis using a Lexicon-Based Approach to capture customer perceived value: An application in the smartphone's context.** 2021. 198 f. Monography (Undergraduate Final Work) – São Carlos School of Engineering, University of São Paulo, São Carlos, 2021.

The social media evolution has been boosted by the internet and by the diffusion of smartphones, which reveals the relevance of consumer's opinions as an instrument to help make strategic decisions inside companies. The concepts of Client-Centered and Customer Value are found at the literature since last century. The expressive volume of user's opinions being posted frequently on the social medias, drives the companies to long for process streaming big data and capture perceptions and trends. The sentiment analysis is a fast and effective method to process these streaming data and allows the acquirement of useful knowledge. The Sentiment Analysis is a way to classify texts with a sentiment tag, which designate if the entire text has a positive-polarity or a negative polarity. On the Other hand, the Aspect Based Sentiment Analysis offers a bigger granularity, reaching the aspect level instead of text level. Given this fact, the goal of this work is developing a Python code capable to mining Twitter data, pre-processing this data, extract the aspects, and then use the sentiment analysis to tag the polarity of them. The computational model uses the lexicon-based approach, which means that isn't necessary the use of pre-labeled data and expand the application possibilities. From this code it was made a pilot application to analyze the perception value of two competitors smartphones, iPhone 13 by Apple and Galaxy S21 by Samsung. Through the results was made an analysis comparing the client's perception of the mobile phones, matching with the technical specifications and the company's strategies. Moreover, was made a SWOT analysis aiming good insights to guide the decision make, understand better the strengths and weakness, besides identify opportunities and threats.

Keywords: Costumer Value, Sentiment Analysis, Aspect Based Sentiment Analysis, Social Media, Twitter, Smartphones, Business Intelligence.

SUMÁRIO

1.	INTRODUÇÃO	13
1.1.	CONTEXTUALIZAÇÃO E LACUNAS DE PESQUISA.....	13
1.2.	OBJETIVOS.....	15
1.2.1.	OBJETIVO GERAL	15
1.2.2.	OBJETIVOS ESPECÍFICOS	16
2.	METODOLOGIA	17
2.1.	ENTENDIMENTO INICIAL DA PESQUISA.....	18
2.2.	FUNDAMENTAÇÃO TEÓRICA	18
2.3.	DESENVOLVIMENTO DO MODELO COMPUTACIONAL	19
2.4.	APLICAÇÃO PILOTO.....	19
2.5.	ANÁLISE DE RESULTADOS.....	19
3.	REVISÃO BIBLIOGRÁFICA.....	20
3.1.	SENTIMENT ANALYSIS	20
3.2.	ASPECT BASED SENTIMENT ANALYSIS	22
3.3.	TEXT BLOB	24
3.4.	CUSTOMER VALUE E CUSTOMER PERCEIVED VALUE.....	25
3.5.	MÍDIAS SOCIAIS E TWITTER	27
4.	DESENVOLVIMENTO DO MODELO COMPUTACIONAL.....	29
4.1.	MINERAÇÃO	30
4.2.	PRÉ-PROCESSAMENTO.....	31
4.3.	EXTRAÇÃO DE ASPECTOS.....	33
4.4.	ANÁLISE DE SENTIMENTO	37
5.	APLICAÇÃO PILOTO.....	38
5.1.	MINERAÇÃO	39
5.2.	PRÉ-PROCESSAMENTO.....	41
5.3.	EXTRAÇÃO DE ASPECTOS.....	42
5.4.	ANÁLISE DE SENTIMENTO	43
5.5.	SUMARIZAÇÃO DOS DADOS	45
6.	ANÁLISE DE RESULTADOS E DISCUSSÕES	52

6.1.	DISCUSSÕES GERAIS.....	52
6.2.	ANÁLISE DA CONFIABILIDADE	57
6.3.	SWOT	58
6.3.1.	STRENGTHS.....	59
6.3.2.	WEAKNESS	59
6.3.3.	OPPORTUNITY.....	59
6.3.4.	THREATS.....	60
7.	CONCLUSÃO	61
8.	REFERÊNCIAS	63

1. INTRODUÇÃO

Este capítulo apresenta a contextualização e lacunas de pesquisa (seção 1.1) e o enunciado do problema e os objetivos da pesquisa (seção 1.2).

1.1. CONTEXTUALIZAÇÃO E LACUNAS DE PESQUISA

O conceito de “Cliente no Centro” não é algo novo e na metade do século passado já estava registrada no livro *The Practice of Management*: “O Cliente é quem vai determinar o que é o negócio, o que irá produzir e se irá prosperar.” (DRUCKER, 1954). Ou seja, as firmas devem focar em atender as necessidades dos clientes e não na venda de produtos (LEVITT, 1960). Sendo assim, um relatório do Gartner Group (2003) sinalizou que as empresas que investissem mais no conceito de Cliente no Centro alcançariam retornos sobre o investimento consideravelmente superiores aos seus pares no futuro.

Sabendo da importância do Cliente, vem a questão de como é possível identificar suas necessidades e dores, de uma forma geral, o que agrega valor a esse usuário. O entendimento desse valor gerado e a sua gestão ao longo do tempo tornou-se reconhecida como um ponto essencial nas estratégias de negócio (PORTER, 1998). A Análise de Valor do Cliente (Customer Value Analysis) é uma vertente dentro da Gestão do Valor do Cliente (Customer Values Management) que permite a identificação de *drivers* de valor para o cliente, por exemplo qualidade e preço que são dois fatores determinantes para os compradores (PRAHALAD; RAMASWAMY, 2000).

As redes sociais são ambientes ideais para capturar essa voz do cliente. Atualmente, dados do Statista (2020) mostram que o Twitter possui cerca de 325 milhões de usuários ativos por mês e tem servido como um grande meio de engajamento de usuários. Um estudo da SMG Twitter Social TV Lab (2013) concluiu que a rede se tornou um ambiente onde os usuários expressam opiniões, críticas e debatem diversos assuntos de forma interativa com a TV, tanto relacionado à programação televisionada como às propagandas. Essa vasta quantidade de dados e o fato de serem públicos, além de disponível acesso através de uma API (Application Programming Interface) do próprio Twitter, fazem dessa rede social uma ótima oportunidade para captar a percepção de valor dos clientes.

Com a proliferação de opiniões e outras formas de expressão on-line, a opinião se transformou em uma espécie de moeda virtual para empresas que desejam comercializar os seus produtos, identificar novas oportunidades e gerenciar suas reputações, onde os clientes fazem o papel de coprodutores de valor (PRAHALAD; RAMASWAMY, 2000). As empresas estão cada vez mais interessadas em automatizar o processo de pesquisa sobre o que está sendo falado sobre elas e compreender as conversas nas mídias sociais, a fim de identificar conteúdo relevante, o que tem levado as empresas a buscarem soluções de análise de sentimentos (CHAE, 2015).

Diante desse grande volume de dados, o maior desafio é a forma de interpretá-los e gerar informação a partir deles. A Análise de Sentimentos é a tarefa de identificar se a opinião que foi expressa em um determinado texto é positiva ou negativa (STINE, 2018). Porém, somente essa classificação da polaridade do texto de forma geral não é suficiente para a maioria das frases já que um único texto costuma apresentar múltiplos aspectos que são expressos de forma independente, sendo assim, uma opção mais específica é a Análise de Sentimentos Baseada em Aspectos (*Aspect Based Sentiment Analysis*), uma ferramenta que classifica o sentimento relativo a um determinado aspecto dentro do texto (NOH; PARK; PARK, 2019).

Um exemplo de aplicação da *Aspect Based Sentiment Analysis* (ABSA) é: “O celular tem uma ótima aparência. A câmera é muito estável. O alto-falante é de ótima qualidade. Porém, a bateria é péssima.”. A frase no geral apresenta muitos aspectos positivos, mas isso não representa toda a ideia do texto (PANDITA, 2021).

Esse valor da mídia social para o gerenciamento dos negócios tem sido estudado em diversas perspectivas multidisciplinares. Por exemplo, em Gestão da Informação, revisões feitas pelo Ngai, Tao e Moon (2015) e Alavi e Denford (2011) sugerem que mídias sociais podem aprimorar a eficiência no compartilhamento de informação e gestão do conhecimento. Na Gestão de Recursos Humanos, a revisão de Poba-Nzaou et al. (2016) identificou desafios na utilização de redes sociais dentro das organizações.

A aplicação da Análise de Sentimentos também pode ser vista sendo aplicada em diversos casos. Tumasjan (2010) utilizou o Twitter para prever os resultados de eleição. Asur e Huberman (2010), Joshi (2010) e Sadikov, Parameswaran e Venetis (2009) utilizaram dados do Twitter, revisões de filmes e blogs para prever receitas de bilheteria para filmes. E Bollen e Zeng (2011) utilizaram as emoções expressas no Twitter para prever o mercado de ações. Carosia, Coelho e Silva (2021) também utilizaram análise de sentimentos para estudar o

mercado de ações, porém focado no Brasil. Schröder e Tieben (2021) utilizaram análise de sentimentos para a gestão de risco.

Porém, não foram encontrados trabalhos que objetivassem utilizar a mineração de dados do Twitter para capturar a percepção de valor pelo cliente através da Análise de Sentimentos Baseada em Aspectos em uma aplicação comparativa considerando dois aparelhos celulares concorrentes.

Portanto, a questão de pesquisa que motiva a execução do presente trabalho é: Como implementar Análise de Sentimento Baseada em Aspectos para extrair informações valiosas para a tomada de decisão focada na percepção de valor do cliente de grandes conjuntos de dados minerados do Twitter?

Sendo assim, o presente trabalho desenvolverá uma aplicação computacional não supervisionada utilizando a linguagem de programação Python com a finalidade de obter um código que possibilita a mineração de dados do Twitter, Tratamento dos Dados, Extração de Aspectos e Classificação de Sentimentos, os quais contemplam todo o escopo do ABSA. Os resultados da Análise de Sentimento Baseada em Aspectos serão interpretados, estudados e terão como entrega final a identificação de aspectos de valor para o cliente e a comparação entre dois produtos no contexto de dois aparelhos celulares lançamentos no mercado, sendo eles Iphone 13 e Galaxy S21. Tais produtos foram escolhidos por apresentarem *features* bem definidas, por serem novidades para os usuários e as opiniões nas redes sociais estão mais aquecidas, além de serem produtos amplamente utilizados pela população mundial, em especial nos Estados Unidos.

O código não se limita ao assunto produtos de varejo e tampouco aos modelos de celulares. A aplicação ilustrativa é executada nesse trabalho para demonstrar como se dá o funcionamento do modelo proposto e para guiar futuras aplicações reais do mesmo em outros contextos. Isso se justifica pelo fato de que não foi encontrado nenhum outro registro, na literatura, da proposição de um modelo similar nesse contexto.

1.2. OBJETIVOS

1.2.1. OBJETIVO GERAL

O estudo em questão tem como objetivo a mineração de dados do Twitter para capturar a percepção de valor pelo cliente através da Análise de Sentimentos Baseada em Aspectos utilizando uma abordagem não supervisionada. Será realizada uma aplicação em dois produtos escolhidos, sendo eles os aparelhos celulares Iphone 13 e o Samsung Galaxy S21. Assim, extrair informação desses dados, analisá-las e obter conclusões e *insights* para tomada de decisão, além de uma comparação entre os dois produtos escolhidos com a finalidade de identificar os componentes que agregam valor ao cliente e são decisivos no momento de compra dos consumidores. Ademais entender as relações entre as opiniões dos usuários com as estratégias tomadas pelas empresas.

1.2.2. OBJETIVOS ESPECÍFICOS

O objetivo geral desdobra-se nos seguintes objetivos específicos:

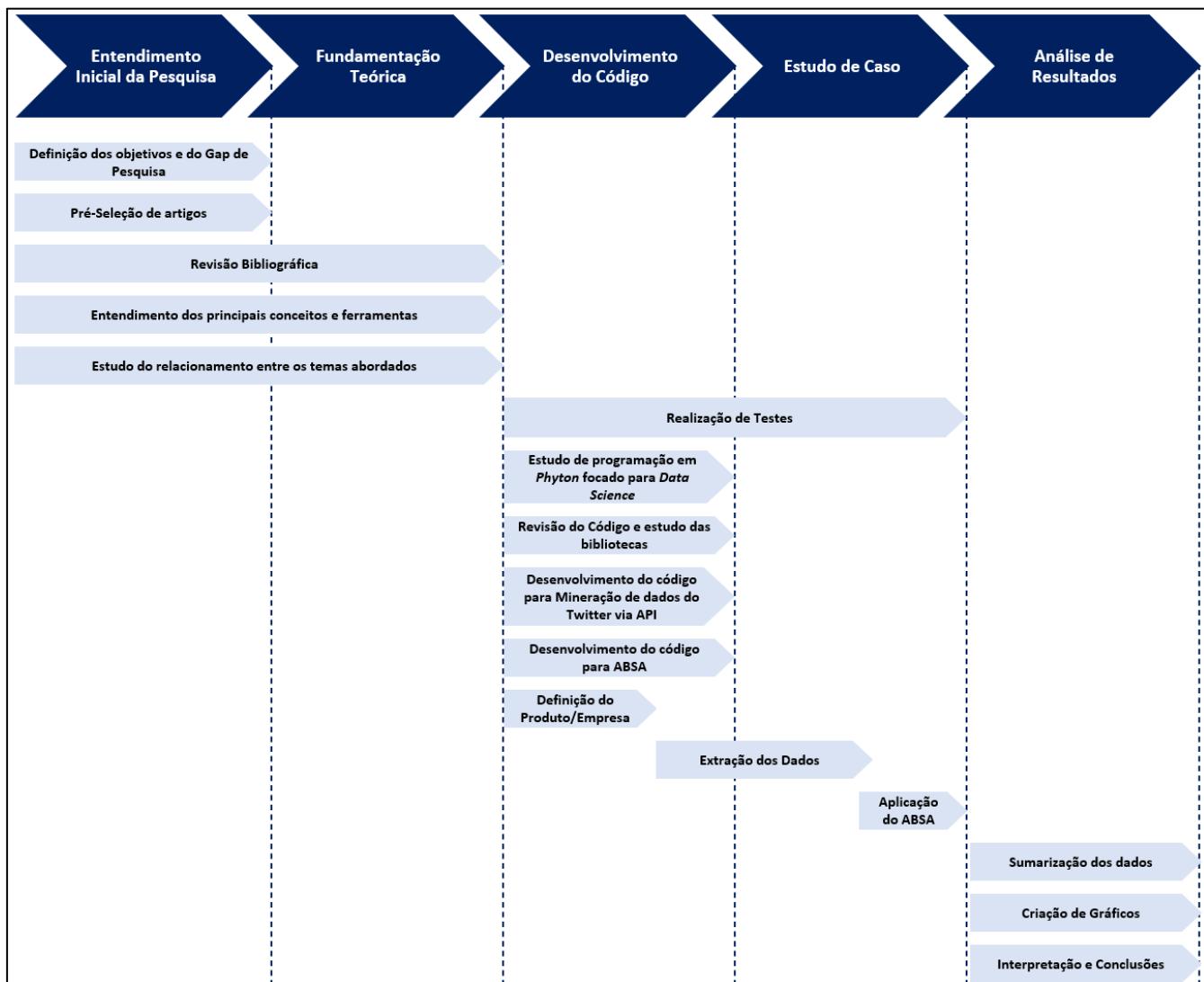
- I. Revisar a literatura sobre valor percebido pelo cliente e *Aspect Based Sentiment Analysis*.
- II. Desenvolver uma aplicação computacional utilizando a linguagem de programação Python com a finalidade de obter um código que possibilita a mineração de dados do Twitter, Tratamento dos Dados, Extração de Aspectos e Classificação de Sentimentos.
- III. Aplicação do modelo em um caso real utilizando como base dois smartphones concorrentes.
- IV. Analisar resultados sobre potencial utilização dos dados para obtenção de insights que auxiliem na tomada de decisão e comparação entre as percepções dos produtos de marcas distintas.

2. METODOLOGIA

De acordo com Bertrand e Fransoo (2002), a presente pesquisa pode ser classificada como quantitativa, axiomática e normativa, pois propõe um modelo a partir das lacunas identificadas na revisão de literatura e das opiniões de especialistas, integrando os conceitos abordados. Segundo Cauchick et al. (2010), a pesquisa quantitativa axiomática normativa baseia-se em modelos que prescrevem uma decisão para determinado problema.

A metodologia que guiou o desenvolvimento do presente trabalho consiste nos seguintes passos, conforme ilustra a Figura 1: Entendimento Inicial da Pesquisa, Fundamentação Teórica, Desenvolvimento do código, Estudo de Caso e Análise de Resultados.

Figura 1 – Etapas e Sub-etapas utilizadas na Metodologia



Fonte: Elaboração própria (2021).

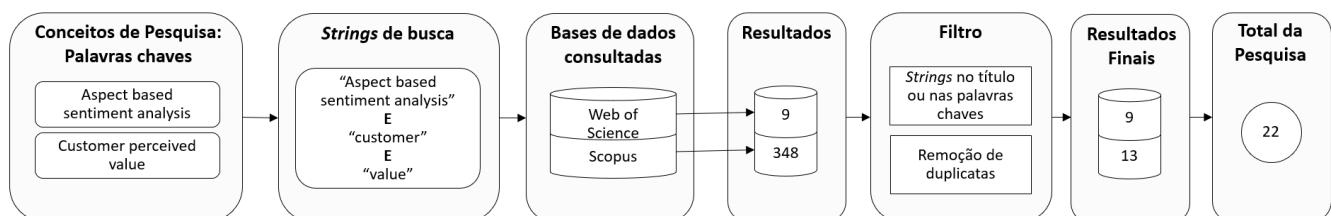
2.1. ENTENDIMENTO INICIAL DA PESQUISA

Essa etapa resume-se no estudo dos temas relacionados à pesquisa. Foi realizada uma pré-seleção e leitura dos principais de artigos relacionados ao tema de interesse, além do entendimento dos principais conceitos e ferramentas. Também, identificação do Gap de Pesquisa e definição dos objetivos da pesquisa.

2.2. FUNDAMENTAÇÃO TEÓRICA

Os textos utilizados durante a revisão bibliográfica foram lidos de forma íntegra e utilizados para aprofundamento na teoria relacionada à pesquisa. Essa fase teve como principal objetivo o embasamento teórico para conseguir desenvolver e aplicar o código da melhor forma possível, já possuindo o conhecimento prévio das ferramentas e formas de aplicação que têm sido utilizadas na literatura. Foi realizada uma breve revisão na literatura para encontrar o estado da arte no contexto de análise de sentimentos baseada em aspectos aplicada para a identificação de valor percebido pelo cliente. Para isso, a partir desses dois conceitos de pesquisa, foram definidas *strings* de busca que foram utilizadas em pesquisas nas bases de dados Web of Science e Scopus. Para reduzir o número de resultados encontrados e possibilitar a leitura dos documentos, foi feito um filtro adicional, onde as *strings* de busca foram procuradas apenas nos títulos ou nas palavras-chaves. Com essa revisão concluiu-se que nos 22 artigos resultantes não foi encontrado nenhum trabalho que objetivasse utilizar a mineração de dados do Twitter para capturar a percepção de valor pelo cliente através da Análise de Sentimentos Baseada em Aspectos em uma aplicação comparativa considerando dois aparelhos celulares concorrentes. O passo a passo descrito acima está ilustrado na Figura 2.

Figura 2 – Metodologia da revisão da literatura



Fonte: Elaboração própria (2021).

2.3. DESENVOLVIMENTO DO MODELO COMPUTACIONAL

Nessa etapa foi desenvolvido o código escrito em Python, através da plataforma Google Colab. O código divide-se em quatro principais etapas sendo elas a Mineração de Dados do Twitter via API pública, Pré-processamento, Extração dos aspectos e, por fim, Análise de Sentimentos Baseada em Aspectos em cima dos dados obtidos. Durante o processo de desenvolvimento foram realizados testes periodicamente para validação do código, utilizando uma lista de frases prontas como exemplo para aplicação. O capítulo 4 detalhará essa atividade.

2.4. APLICAÇÃO PILOTO

Para o estudo de caso foram escolhidos os aparelhos celulares Iphone 13 e Galaxy S21. Os motivos que levaram a escolha desses produtos são: Produtos com *features* variadas e bem definidas, lançamento recente e alta repercussão nas redes sociais. Além disso, a aplicação de ABSA é bem consolidada dentro do contexto de *mobile-phone reviews*. Esta fase compreende as etapas de mineração dos dados reais disponibilizados pelo Twitter, utilizando filtros relacionados aos produtos e *features*, pré-processamento desses dados, extração dos aspectos e a classificação dos sentimentos expressados, assim como a subjetividade, através do uso do Text Blob. O estudo de caso será apresentado no capítulo 5.

2.5. ANÁLISE DE RESULTADOS

Essa etapa explicita as contribuições que o código pode fornecer através da análise dos resultados obtidos durante a aplicação real do código. A partir das informações obtidas durante o estudo de caso, os resultados foram organizados em tabelas, gráficos e foi utilizada uma matriz *SWOT*, com a finalidade de obter insights úteis que auxiliam na tomada de decisão, como a identificação dos principais aspectos na visão do consumidor em relação aos produtos escolhidos. Além disso, uma comparação entre os produtos destacando suas vantagens e desvantagens entre si.

3. REVISÃO BIBLIOGRÁFICA

Este capítulo caracteriza os conceitos de *Sentiment Analysis*, *Aspect Based Sentiment Analysis*, *Customer Value* e Mídias Sociais, além de apresentar diversas definições presentes na literatura. Discorre também sobre o Twitter mais especificamente e sobre a biblioteca utilizada para aplicação da análise de sentimentos.

3.1. SENTIMENT ANALYSIS

A análise de sentimentos consiste em atribuir um rótulo emocional ao texto, rótulos os quais indicam uma polaridade se o texto expressa algo positivo ou negativo (STINE, 2019). A análise de sentimentos pode ser dividida em 3 níveis sendo o primeiro deles o *Document Level* que aponta apenas uma polaridade de todo o documento, o *Sentence Level* que também aponta uma polaridade, porém no nível das sentenças, e por fim o *Feature Level* que entra na granularidade do sentimento em relação a cada aspecto dentro do texto (PANDITA, 2021).

Sempre que decisões precisam ser tomadas, opiniões de outras pessoas são levadas em consideração, sendo que antigamente era comum que as pessoas perguntassem para família e amigos. Porém, com o crescimento explosivo das redes sociais, pessoas e organizações passaram a utilizar os conteúdos desses aplicativos para auxiliar na tomada de decisões, já que é algo muito mais abrangente e barato em relação aos métodos mais antigos, como por exemplo realizar uma pesquisa de mercado através de ligações telefônicas (LIU, 2012).

Com a relevância das mídias sociais na internet, análise de sentimentos tornou-se uma abordagem essencial para analisar uma vasta quantidade de dados fluídos (SUN; LUO; CHEN, 2017). Essas técnicas referem-se ao ato de manusear e analisar opiniões e emoções dos usuários de determinadas marcas ou produtos (LIU, 2012). O objetivo é detectar e classificar opiniões e sentimentos de forma automática através da combinação entre mineração de dados, tratamento dos dados, estruturação da informação, processamento de linguagem natural e, em alguns casos, aprendizagem de máquina (BALAHUR; JACQUET, 2015).

A aplicação da Análise de Sentimentos pode ser vista em diversos casos. Tumasjan (2010) utilizou o Twitter para prever os resultados de eleição. Asur e Huberman (2010), Joshi (2010) e Sadikov, Parameswaran e Venetis (2009) utilizaram dados do Twitter, revisões de

filmes e blogs para prever receitas de bilheteria para filmes. E Bollen e Zeng (2011) utilizaram as emoções expressas no Twitter para prever o mercado de ações. Carosia, Coelho e Silva (2021) também utilizaram análise de sentimentos para estudar o mercado de ações, porém focado no Brasil. Schröder e Tieben (2021) utilizaram análise de sentimentos para a gestão de risco. Pode-se observar diversas aplicações atuais e em contextos distintos.

Big data analytics, o que é definido como uma coleção de informações, ferramentas de análises, algoritmos computacionais e técnicas para obter *insights* relevantes e padrões a partir de uma extensa coleção de dados, tem o potencial de ajudar gestores a entregar valor contínuo, melhorias de negócio e vantagem competitiva (KAMBLE; GUNASEKARAN, 2020). No contexto em questão o Twitter abriga a coleção de informações e o código desenvolvido, juntamente com a análise dos dados obtidos, representa as ferramentas e algoritmos para a obtenção de *insights*.

Big data stream computing depende de programas que processam um fluxo de dados continuamente, provendo rápido processamento de uma grande quantidade de data, tornando-se a mais rápida e eficiente solução para obtenção de conhecimento útil a partir de uma *Big data*, permitindo as organizações reagirem rapidamente aos problemas que surgirem e realizarem previsões de novas tendencias (SUN et al., 2015). Tal ponto fortalece a utilização da Análise de Sentimentos para tomada de decisão rápida e mais assertiva.

Um processo de mineração usual de opinião consiste na seguinte série de passos definidos: obtenção dos dados, pré-processamento do texto, classificação da opinião, por fim sumarização e visualização dos resultados (SUN; LUO; CHEN, 2017).

Balazs e Velásquez (2016) argumentam que a relevância dessa técnica se deve ao fato de as opiniões representarem o principal fator no processo de tomada de decisão de indivíduos e organizações, já que elas influenciam as atitudes e crenças. Os autores também pontuam que tamanho o interesse em performance na detecção e interpretação de opiniões de forma automática fez desse campo de pesquisa, atualmente, um dos mais populares nas comunidades de *Natural Language Processing* e *Computer Science*.

Porém, a tradicional análise de sentimentos focada em classificar o sentimento expresso em um texto de forma abrangente e sem especificar sobre o que o sentimento se refere não é suficiente para textos que expressam diferentes sentimentos para diferentes aspectos (HOANG; BIHORAC; ROUCES, 2019). A identificação de sentimentos associadas à um específico

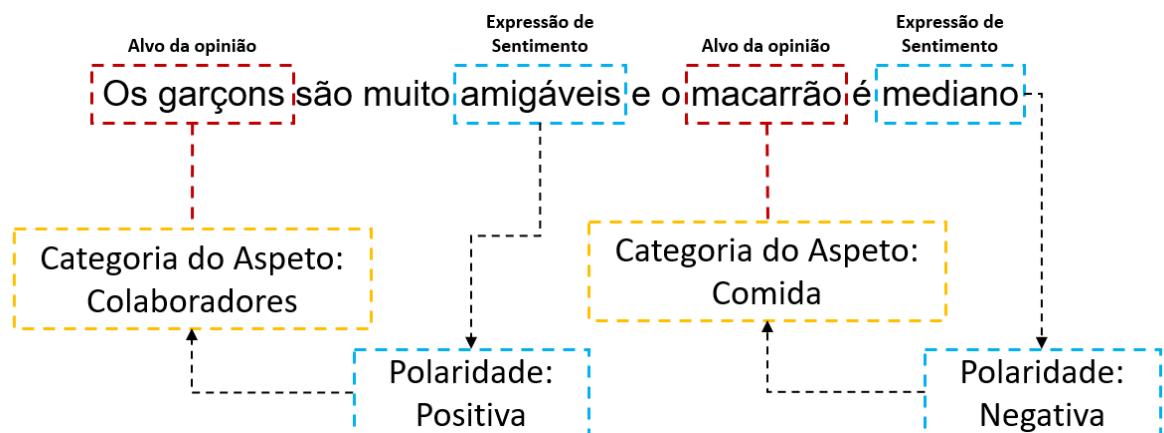
aspecto no texto é uma tarefa mais complexa e é conhecida como Análise de Sentimentos Baseada em Aspectos (ABSA). A principal vantagem do ABSA é que além de permitir analisar o conteúdo das sentenças em larga escala, é possível identificar qual aspecto está sendo tratado na sentença e qual o sentimento atribuído ao aspecto em questão.

3.2. ASPECT BASED SENTIMENT ANALYSIS

A Análise de Sentimentos Baseada em Aspectos resume-se em atribuir diversas análises de sentimento em um nível alto de granularidade (LIU, 2021). Para isso, é utilizada a identificação dos termos e a classificação do sentimento relacionados a esses termos (PENG et al., 2020). Em geral os textos possuem múltiplos aspectos sendo eles expressos de forma independente, por isso, para uma análise de sentimentos mais profunda, o conteúdo deve ser examinado no nível dos aspectos e não no nível do texto (NOH; PARK; PARK, 2019).

Um exemplo é a sentença “Os garçons são muito amigáveis e o macarrão é mediano” em que a extração do aspecto se resume nos termos “Garçons” e “Macarrão”, enquanto a classificação do sentimento, que irá rotular o termo como positivo, neutro ou negativo, é com base nos adjetivos “Amigáveis” e “Mediano” (PENG et al., 2020). O funcionamento da extração de aspectos e classificação de sentimentos está ilustrada na Figura 3.

Figura 3 – Ilustração do funcionamento da ABSA

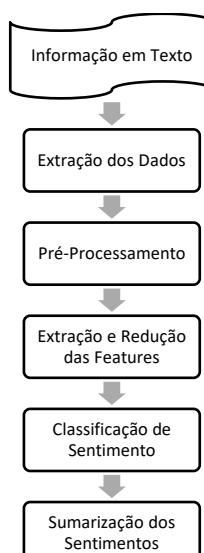


Fonte: Elaboração própria (2021)

De acordo com Pandita (2021) e conforme sumarizado na Figura 4 é possível distribuir o método de Análise de Sentimentos Baseada em Aspectos em algumas principais etapas, sendo elas:

- Extração dos Dados: A análise de sentimentos se inicia com uma coleta de dados online e posterior armazenamento em uma base local. Esta etapa é crucial pois a precisão dos dados obtidos irá influenciar todo o resultado da análise de sentimentos.
- Pré-Processamento dos Dados: Essa etapa é o primeiro passo após a mineração dos dados. Nela os dados desestruturados são preparados para serem reconhecidos durante a análise de sentimentos. Um exemplo seria a frase: “Comprei meu novo celular no site <https://www.exemplotcc.com.br/> que é ótimo, o #iphone13 é simplesmente I-N-C-R-I-V-E-L!!! 😊”. Após o pré-processamento, ela adquiriria o seguinte formato: “comprei meu novo celular no site que é ótimo, o iphone13 é simplesmente incrível!!!”
- Extração e Redução das *Features*: etapa crucial pois a identificação dos aspectos e a escolhas das *features* corretas farão com que a análise tenha uma maior acurácia. Atributos de múltiplas palavras como “*battery life*” devem ser identificados como um único aspecto, sendo tratado como “*batterylife*”.
- Classificação de Sentimento: onde as declarações são associadas como positivas, negativas ou neutras.
- Sumarização dos Sentimentos: Os resultados são sumarizados para assim possibilitar que a tomada de decisão seja feita de forma mais simples e assertiva, incluindo uma visualização de resultados objetiva.

Figura 4 – Etapas para análise de sentimentos



Fonte: Adaptado de Pandita (2021)

No ABSA é feita uma classificação de sentimento em um dado texto T para um aspecto referenciado R que dá origem a diversas tuplas, como no exemplo acima seriam as seguintes tuplas: (“são muito amigáveis”, SERVIÇO) e (“é mediano”, COMIDA), cada qual teriam classificações diferentes, sendo o serviço classificado como positivo em relação ao serviço enquanto a comida como negativa (NOH; PARK; PARK, 2019).

Na etapa de classificação, existem duas principais formas de abordar, sendo a *Supervised Learning Approach*, a qual depende de treinamento e acaba desempenhando mal ao trocar de contexto, enfrentando uma barreira para escalar o método em diferentes domínios, e a *Lexicon-Based Approach*, que evita esse tipo de dificuldade ao trocar de contexto (DING; LIU; YU, 2008). Além disso, a *Lexicon-Based Approach* tem apresentado uma performance superior em diferentes tipos de domínios (LIU, 2012).

3.3. TEXT BLOB

Text Blob é uma biblioteca utilizada em Python para processamento de linguagem natural e é utilizada para identificar a polaridade e subjetividade dos dados. A polaridade varia de -1,0 até 1,0, enquanto a subjetividade varia de 0 até 1 (LORIA, 2018).

Basicamente, como já explicitado anteriormente, existem dois tipos de algoritmos de classificação: os de aprendizagem supervisionada (*Supervised Learning Approach*) e os de aprendizagem não-supervisionada (*Lexicon-Based Approach*). O autor Kotsiantis (2007) aponta que se existem exemplos categorizados, com as corretas associações correspondentes, o modelo é chamado de supervisionado, em contraste com o não-supervisionado onde não são fornecidos exemplos categorizados.

Neste trabalho foi escolhida a abordagem não-supervisionada e para isso foi utilizado o Text Blob. O Text Blob possibilita a análise de sentimento sem exigir uma lista de exemplos categorizados, já que é um dicionário com os sentimentos léxicos (AHUJA; DUBEY, 2017). Sentimento léxico é um canal para análise de sentimentos que tem como fonte uma base de dados de unidades lexicais com suas orientações de sentimento (AKUNDIA et. al., 2018).

Embora a abordagem supervisionada proporcione análises com resultados mais satisfatórios em termos de precisão, existem duas desvantagens: um documento precisa ser previamente classificado e só é possível utilizar os exemplos categorizados que foram fornecidos durante o processo de treinamento (GUTIÉRREZ-BATISTA; VILA; MARTIN-BAUTISTA, 2021).

De forma sumarizada, Gutiérrez-Batista, Vila e Martin-Bautista levantaram alguns pontos que sustentam a escolha do Text-Blob e que se aplicam perfeitamente no contexto do presente trabalho, sendo eles:

- É uma ferramenta não supervisionada, ou seja, não é necessária uma base de dados categorizada, o que possibilita uma abordagem mais genérica que é o desejado e menos dependente do *dataset* utilizado.
- A ferramenta funciona perfeitamente com dados textuais de redes sociais.
- A ferramenta é rápida o suficiente para ser utilizada com dados contínuos (*streaming data*).
- A ferramenta permite saber o que é positivo ou negativo e o quão positivo ou negativo é o atributo em questão, além da subjetividade inerente a avaliação, a qual aponta se o texto é objetivo ou mais subjetivo.

3.4. CUSTOMER VALUE E CUSTOMER PERCEIVED VALUE

O Conceito de valor é um dos elementos básicos das teorias de marketing, como a identificação e a criação de valor para o cliente as quais tornaram-se essenciais para o futuro sucesso das companhias (GRAF; MAAS, 2008). Porém, esse assunto não é exclusivo da área de Marketing e tem sido influenciado por pesquisas de diversas áreas envolvendo assuntos como Cadeia de Valor, Comportamento do Cliente, Satisfação do Cliente e Qualidade (PAYNE; HOLT, 2001).

Os tópicos de Satisfação do Cliente e Qualidade estão mais próximos em relação à Customer Value e algumas vezes são até tratados como sinônimos na literatura (GILBERT; VELOUTSOU, 2006).

A definição de *Customer Value* é a percepção de preferência do cliente e a avaliação de quais atributos dos produtos são utilizados para facilitar com que o usuário alcance seus

objetivos e o propósito em dada situação (WOODRUFF, 1997). Porém, essa definição não é unânime e única, existem algumas variações como uma mais simplificada que assume *Customer Value* como o que o consumidor ganha (benefício, qualidade, utilidade) através da compra e o uso do produto contra o que foi pago (preço, custo, sacrifício), gerando uma emoção relacionada ao produto (BUTZ; GOODSTEIN, 1996).

Songailiene et al. (2011) destaca que, embora exista uma variedade de definições sobre o conceito, existe uma concordância generalizada na literatura que: a percepção de valor é algo individual e subjetivo, também envolve trocas entre o que é dado e o que é recebido e sempre está relacionado com competição. As diferentes definições encontradas para Customer Perceived Value estão resumidas na Tabela 1.

Tabela 1 – Diferentes definições de valor percebido pelo cliente

Autores	Campo	Jornal	Ano	Definição de Valor Percebido do Cliente
Sweeney e Soutar, 2001	Varejo	Journal of Retailing	2001	Uma combinação de quatro dimensões: emoção, social, qualidade/performance e preço/valor por dinheiro.
Song et al., 2016	Marketing/Serviços	Industrial Marketing Management	2016	Avaliação do comprador nos ambitos econômico e técnico, além de avaliar as relações de benefícios em troca de um preço pago.
Hannien e Karjaluoto, 2017	Cadeia de Suprimentos	Journal of Cleaner Production	2017	Avaliação do valor que é ganho através de uma relação de fornecedor.
Walsh et al., 2014	Varejo	Journal of Business Research	2014	A avaliação geral do usuário do produto (ou serviço) baseadas nas percepções do que é dado.
Li et al., 2015	Tomada de Decisão	Discrete Dynamics in Nature and Society	2015	O que o consumidor está disposto a pagar em troca do que é oferecido pelo produto.
Petrick 2002	Lazer	Journal of Leisure Research	2002	O que o consumidor recebe para o que ele abre mão.
Boksberger e Melsen, 2011	Marketing/Serviços	Journal of Services Marketing	2011	Uma combinação de avaliações da percepção dos consumidores em relação aos benefícios e sacrifícios com a preferência comportamental afetando a situação geral .
Songailiene et al., 2011	Serviços B2B	European Journal of Marketing	2011	Avaliação das dimensões financeira, estratégica e co-criação de valor determinadas pelo consumidor.

Fonte: Adaptado de Zanon et al. (2020)

A gestão do Valor para o Cliente tornou-se o coração das estratégias organizacionais, utilizando esse melhor entendimento dos pontos que satisfazem os clientes para maximizar o valor entregue para públicos específicos, assim ganhar vantagem competitiva e fomentar a lucratividade (EVANS, 2002). A satisfação do cliente implica em um nível maior de lealdade e comentários mais positivos, o que contribui para uma posição competitiva mais consolidada, um maior *market share* (BENEKE et al., 2013) e lucro (OKONGWU et al., 2016).

Esforços recentes para medir o valor percebido têm mostrados dificuldades em quantificar esse valor (ZANON et al., 2020). Trigos et al. (2019) demonstrou que aplicando técnicas para melhorar a performance de indicadores relacionados diretamente com os consumidores é possível melhorar a percepção de valor entregue.

3.5. MÍDIAS SOCIAIS E TWITTER

Mídias sociais são plataformas que permitem que os usuários criem, interajam, compartilhem e troquem informações e ideias em comunidades virtuais e redes (CARR; DECRETON; QIN, 2013).

Tem mudado de forma abrupta o modo como as pessoas se comunicam e colaboram (ARAL et al., 2013) com canais sendo formados, os quais permitem conexões e compartilhamento de informação através das redes sociais (AHMED et al., 2019). O uso das redes sociais expandiu-se, também, para dentro das companhias e permite uma comunicação rápida e visual entre os colaboradores (LEONARDI et al., 2013). As companhias também usam externamente para comunicação com os clientes ou o público mais genérico (GU; YE, 2014), permitindo a construção de uma relação mais próxima ao consumidor (CHAE, 2015).

De acordo com relatório intitulado "Digital 2021 October Global Statshot Report" elaborado pela organização We Are Social (2021), 61.8% (4,88 bilhões) da população utiliza a internetativamente, sendo que aproximadamente 4,55 bilhões de pessoas (57.6% da população mundial) são usuários ativos de mídias sociais, e aproximadamente 91% das pessoas utilizam a internet através de aparelhos celulares ou tablets. Somente o Twitter, atualmente, possui cerca de 436 milhões de usuários ativos por mês, sendo os Estados Unidos o país com a maior audiência, superando o valor de 77,5 milhões de usuários.

Microblogging tornou-se uma ferramenta de comunicação muito popular entre os usuários da internet, onde milhões de mensagens são publicadas diariamente em *websites* populares que fornecem esse tipo de serviço, como o Twitter (PAK; PAROUBEK, 2010). Essa expressiva quantidade de dados e a disponibilização de público acesso dessas postagens através de uma API do próprio Twitter, fazem dessa rede social uma ótima oportunidade para extração de dados e aplicação de ABSA. Além disso, alguns aspectos do Twitter tornam mais fácil o seu uso para pesquisa, como a limitação do tamanho de posts (280 caracteres), que reduz a complexidade da interpretação dos textos, além de não limitar a variedade de tópicos, nem os tipos de relações entre os usuários (POCHAMPALLY, 2011). Por fim, o Twitter já foi utilizado como fonte de dados para diversos estudos como foi citado diversos exemplos no capítulo 2.1 Sentiment Analysis.

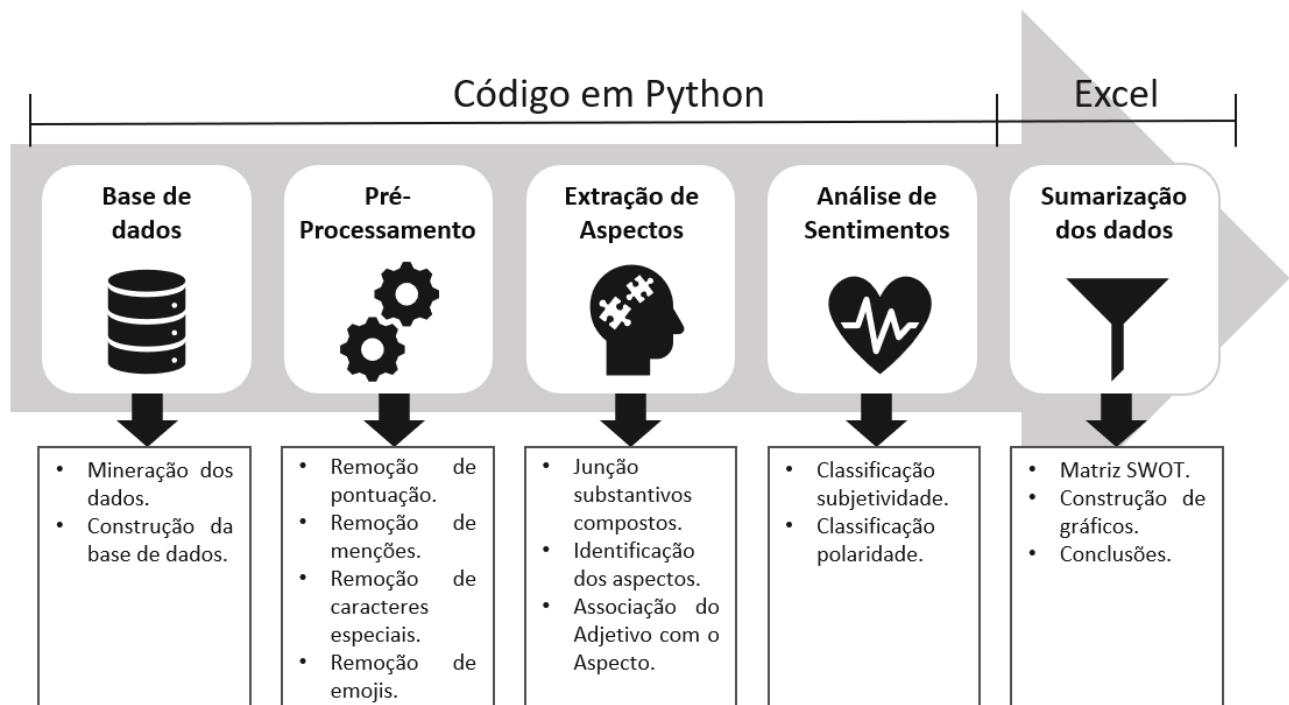
Alguns dados relevantes mostram a importância do Twitter. Smith e Anderson (2018) publicaram em sua pesquisa que aproximadamente 42% dos usuários do Twitter estão diariamente na plataforma. Ahlgren (2021) publicou uma coleção detalhada das mais atualizadas estatísticas do Twitter para 2020 a qual revela que são enviados 500 milhões de *tweets* diariamente, 83% dos líderes do mundo estão no Twitter e a plataforma serve mais de 2 bilhões de consultas de pesquisa diariamente. Esse são 3 exemplos que mostram a magnitude do Twitter.

4. DESENVOLVIMENTO DO MODELO COMPUTACIONAL

Para a criação do código foram utilizados o Google Colab e a linguagem de programação Python. Como descrito por Santos (2020) em um artigo da Alura, a ferramenta Colab é um serviço de nuvem gratuito hospedado pelo próprio Google que permite misturar código fonte (geralmente em Python) e texto rico (geralmente em *markdown*) com imagens e resultados desse código, técnica conhecida como *notebook* (“caderno”).

O código ficou dividido em 4 principais seções, as quais compõe quatro das cinco etapas de ABSA comentadas em capítulos acima e melhor desenvolvidas na Figura 5. A quinta etapa de sumarização foi realizada através do Excel, devido à maior quantidade de opções para estilizar os gráficos e deixá-los mais visíveis e compreensíveis para o presente trabalho. Além disso, possibilita a criação de gráficos e filtros dinâmicos e fácil utilização.

Figura 5 – Desenvolvimento das etapas do ABSA



Fonte: Elaboração própria (2021)

4.1. MINERAÇÃO

Para a obtenção dos dados do Twitter utilizou-se uma API disponibilizada pela própria plataforma. Para a utilização da API é necessária a criação de uma conta de desenvolvedor, a qual foi usufruída a versão acadêmica que possui algumas vantagens em relação à versão mais simples. Após a criação da conta, é possível consultar quatro informações que compõe a credencial de acesso aos *tweets*, sendo elas: *Consumer Key*, *Consumer Secret*, *Access Token* e *Access Token Secret*. Esses valores foram atribuídos a variáveis as quais são consumidas no momento de conexão com o Twitter.

Para minerar os dados, utilizamos a biblioteca Twitter Search, a qual possui funções específicas para trabalhar com a API do Twitter. No primeiro bloco de código são passados os valores da credencial, já comentado acima, para obter acesso aos dados da rede social. Em seguida, são passados os parâmetros de busca que serão utilizados para obtenção dos *tweets*. É válido ressaltar quais são esses parâmetros:

- **Set Keywords** – São estabelecidas as principais palavras-chaves de busca, no contexto do trabalho foram utilizadas iphone13, iphone 13, galaxyS21 e galaxy S21. O “or_operator” define se a condição de busca é o *tweet* possuir todas as palavras chaves ou qualquer uma delas de forma independente.
- **Set Language** – É definida a língua dos *tweets* que serão selecionados. Foi escolhida a língua inglesa pois as bibliotecas de processamento de linguagem são mais desenvolvidas e possuem mais opções em relação ao português.
- **Add Keyword** – São adicionadas mais palavras chaves à chave principal. Essa ferramenta foi utilizada duas vezes, sendo que a primeira utilização é responsável por excluir os *retweets* (replicação de um *tweet* já publicado), enquanto na outra são adicionadas ao filtro algumas das principais *features* dos celulares. Esses blocos são adicionados com a condição “e”, porém dentro dessas chaves foi estabelecida a condição “ou”. Portanto, o resultado é uma busca de 2 blocos de chaves, onde é necessário encontrar uma palavra de cada bloco simultaneamente, todavia dentro dos blocos só é necessário encontrar uma das palavras.

Após a definição dos parâmetros de busca, a conexão com o Twitter é realizada. Um *loop* corre por todas as linhas do resultado e cria um documento no formato json, com as

informações de quando o *tweet* foi publicado, o id de quem publicou e o conteúdo do texto de fato. O trecho do código descrito acima pode ser visto na Figura 6.

Figura 6 – Código mineração de dados do Twitter

```
from TwitterSearch import *

try:

    ts = TwitterSearch(
        consumer_key = consumer_key,
        consumer_secret = consumer_secret,
        access_token = access_token,
        access_token_secret = access_token_secret
    )

    tso = TwitterSearchOrder()
    tso.set_keywords(['iphone 13', 'iphone13'], or_operator = True)
    tso.set_language('en')
    tso.add_keyword(['-filter:retweets'])
    tso.add_keyword(['camera', 'battery', 'screen', 'price', 'design', 'performance', 'display'], or_operator = True)

    for tweet in ts.search_tweets_iterable(tso):

        print('created_at: ', tweet['created_at'], 'User_id: ', tweet['id_str'], 'Tweet: ', tweet['text'])

        created_at = tweet['created_at']
        user_id = tweet['id_str']
        texto = tweet['text']

        with open("tweet.json", "a+") as output:

            data = {"created_at": created_at,
                    "User_id": user_id,
                    "tweet": texto}

            output.write("{}\n".format(json.dumps(data)))

except TwitterSearchException as e:
    print(e)
```

Fonte: Elaboração própria (2021)

4.2. PRÉ-PROCESSAMENTO

O pré-processamento resume-se em moldar o conteúdo do *tweet* para que ele fique o mais próximo de uma frase escrita formalmente, sem abreviações, símbolos, ou quaisquer outros itens que podem prejudicar o processamento de linguagem natural humana. Nessa seção utilizamos a biblioteca Re, a qual auxilia o trabalho com expressões regulares no texto.

Foi criada uma função denominada Preprocessing que está dividida em pequenos blocos cada um com uma função, sendo elas:

- **Limpa links, pontos, vírgulas, ponto, vírgula e trailha** – Retira todos esses caracteres do texto. É interessante apontar que no início do desenvolvimento do trabalho, as *hashtags* eram retiradas completamente do texto. Porém, muitos dos

tweets utilizam os aspectos em *hashtags* e a exclusão delas prejudicava a extração de aspectos. Sendo assim, a opção escolhida foi de retirar apenas a cerquilha (#) e manter o conteúdo da *hashtag*. Exemplo de frase que utiliza a *hashtag* no aspecto: “*Oh my god!! The #iphone13 is amazing!!*”. Essa frase caso a *hashtag* fosse retirada: “*Oh my god!! The is amazing!!*”. A mesma frase somente retirando a tralha: “*Oh my god!! The iphone13 is amazing!!*”.

- **Limpa Menções** – No Twitter é normal realizar menções de outros perfis da rede social, os quais iniciam com um @. Nesse bloco nós retiramos a menção totalmente, já que o perfil de alguém no Twitter não é um aspecto desejado.
- **Limpa Espaços Extras** – É comum encontrar textos com espaços duplos ou até em quantidades maiores. Esse bloco garante que o texto terá apenas um espaço separando as palavras.
- **Limpa html tags** – Frequentemente encontram-se *tweets* com *links* para outros sites. Esse bloco de código retira esses *links* que poluem o conteúdo de valor para o presente trabalho.
- **Limpa emojis** – É muito comum a utilização de *emojis* em textos de redes sociais. Esse bloco de código retira esses *emojis* que são figuras já que o foco é trabalhar na análise de textos.

Com a função de pré-processamento criada ela foi utilizada na aplicação de cada *tweet* minerado. O código dessa seção pode ser consultado na Figura 7.

Figura 7 – Código função Preprocessing

```

def Preprocessing(instancia):
    #Limpa links, pontos, vírgulas, ponto, vírgula e trailha
    instancia = re.sub(r"ht\|tp\|S+", "", instancia).lower().replace('.','').replace(';', '').replace('-','').replace(':','').replace('#','')
    #Limpa menções
    instancia = re.sub(r"@[\S+]", "", instancia).lower()
    #Limpa espaços extras
    instancia = re.sub(r' +', ' ', instancia).lower()
    #Limpa html tags
    instancia = re.sub(r'<.*?>', ' ', instancia).lower()
    #Limpa emojis
    emoji_pattern = re.compile("["
        u"\U0001F600-\U0001F64F" # emoticons
        u"\U0001F300-\U0001F5FF" # symbols & pictographs
        u"\U0001F680-\U0001F6FF" # transport & map symbols
        u"\U0001F1E0-\U0001F1FF" # flags (iOS)
        u"\U00002500-\U00002BEF" # chinese char
        u"\U00002702-\U00002780"
        u"\U00002702-\U00002780"
        u"\U000024C2-\U00001F251"
        u"\U00001f926-\U00001f937"
        u"\U000010000-\U00010ffff"
        u"\u2648-\u2642"
        u"\u2600-\u2B55"
        u"\u200d"
        u"\u23cf"
        u"\u23e9"
        u"\u231a"
        u"\ufe0f" # dingbats
        u"\u3030"
    "+", re.UNICODE)
    instancia = (emoji_pattern.sub(r'', instancia))
    return(instancia)

```

Fonte: Elaboração própria (2021)

4.3. EXTRAÇÃO DE ASPECTOS

A extração de aspectos é o momento crucial para a obtenção de bons resultados, pois é nesse momento que são atribuídos os adjetivos aos aspectos, como foi ilustrado na Figura 1. Para isso utilizamos as bibliotecas Stanfordnlp, Stanza e Nltk.

Para ficar didático, as próximas explicações serão ilustradas em aplicações na seguinte frase de exemplo: The sound quality is great but the battery life is very bad.

No início ocorre a *tokenização* das palavras, onde cada palavra é tratada separadamente. Ao aplicar a função “word_tokenize” no exemplo, teríamos o seguinte resultado: ‘The’, ‘Sound’, ‘Quality’, ‘is’, ‘grat’, ‘but’, ‘the’, ‘battery’, ‘life’, ‘is’, ‘very’, ‘bad’.

A partir dessa *tokenização*, as palavras recebem uma *tag* sinalizando a classificação gramatical da palavra, como Verbo (JJ), Substantivo (NN), Determinante (DT), entre outras classificações. Aplicando no exemplo obtém-se o seguinte resultado: ('the', 'DT'), ('sound',

'NN'), ('quality', 'NN'), ('is', 'VBZ'), ('great', 'JJ'), ('but', 'CC'), ('the', 'DT'), ('battery', 'NN'), ('life', 'NN'), ('is', 'VBZ'), ('very', 'RB'), ('bad', 'JJ'), ('.', '.').

O trecho que possibilita a *tokenização* e classificação em *tags* de cada um dos *tweets* está exposto na Figura 8.

Figura 8 – Código tokenização e classificação gramatical

```
taggedList = []
for i in listaFrases:
    listaTemporaria = []
    for frase in i:
        txt_list = nltk.word_tokenize(frase)
        listaTemporaria.append(nltk.pos_tag(txt_list))
    taggedList.append(listaTemporaria)
```

Fonte: Elaboração própria (2021)

O próximo bloco de código, o qual está explícito na Figura 9, é responsável por unir os substantivos compostos em uma única palavra. De forma simplista, ele identifica a ocorrência de dois substantivos em sequência e faz a junção em um único *token*. Tendo o exemplo como base, a frase ficaria da seguinte forma: The soundquality is great but the batterylife is very bad.

Figura 9 – Junção dos substantivos compostos

```
compiladoFrases = []
for i in taggedList:
    newwordList = []
    flag = 0
    listaSimples = i[0]
    for i in range(0, len(listaSimples)-1):
        if(listaSimples[i][1]=="NN" and listaSimples[i+1][1]=="NN"):
            newwordList.append(listaSimples[i][0]+listaSimples[i+1][0])
            flag=1
        else:
            if(flag==1):
                flag=0
                continue
            newwordList.append(listaSimples[i][0])
        if(i==len(listaSimples)-2):
            newwordList.append(listaSimples[i+1][0])
    finaltxt = ' '.join(word for word in newwordList)
    compiladoFrases.append(finaltxt)
```

Fonte: Elaboração própria (2021)

Em um momento do código são retiradas as *stopwords*. As *stopwords* são palavras que não agregam nenhuma informação relevante, como preposições e conjunções. A remoção dessas palavras faz parte da lógica do código e está presente na literatura como um dos passos dentro de ABSA. O trecho do código em que ocorre essa ação está na Figura 10.

Figura 10 – Junção dos substantivos compostos

```
new_txt_list_final = []
listaSemStopWords = []
listaSemStopWordsTagged= []
for i in compiladoFrases:
    stop_words = set(stopwords.words('english'))
    new_txt_list = nltk.word_tokenize(i)
    wordsList = [w for w in new_txt_list if not w in stop_words]
    new_txt_list_final.append( new_txt_list)
    listaSemStopWords.append(wordsList)
    listaSemStopWordsTagged.append(nltk.pos_tag(wordsList))
```

Fonte: Elaboração própria (2021)

Em sequência, é utilizada uma das ferramentas da biblioteca Stanza que faz a identificação da relação entre as palavras no texto. No exemplo, teríamos o seguinte desfecho: ['the', 'soundquality', 'det'], ['soundquality', 'great', 'nsubj'], ['is', 'great', 'cop'], ['great', 0, 'root'], ['but', 'bad', 'cc'], ['the', 'batterylife', 'det'], ['batterylife', 'bad', 'nsubj'], ['is', 'bad', 'cop'], ['very', 'bad', 'advmod'], ['bad', 'great', 'conj'], ['.', 'great', 'punct']. O bloco de código correspondente ao trecho abordado acima pode ser encontrado na Figura 11, a seguir. A relação entre as palavras “*the*” e “*soundquality*” foi classificada como “*det*”, o que significa que uma delas é o sujeito da frase e a outra é seu determinante. Para conhecer o significado de cada classificação, basta acessar o manual Stanford typed dependencies manual, escrito por Marneffe e Manning (2016).

Figura 11 – Código de identificação das relações entre as palavras

```

nlp = stanza.Pipeline('en') # initialize English neural pipeline
count = 0
listaDep_node = []
for i in compiladoFrases:
    try:
        # print(i)
        doc = nlp(i)
        dep_node = []
        for dep_edge in doc.sentences[0].dependencies:
            dep_node.append([dep_edge[2].text, dep_edge[0].id, dep_edge[1]])
        newwordlist = new_txt_list_final[count]
        count = count + 1
    except:
        print(f'essa frase foi descartada {i}')
        count = count + 1
    for i in range(0, len(dep_node)):
        try:
            if (int(dep_node[i][1]) != 0):
                dep_node[i][1] = newwordlist[(int(dep_node[i][1]) - 1)]
        except:
            print(f'essa frase foi descartada {count}')
    listaDep_node.append(dep_node)
print(listaDep_node)

```

Fonte: Elaboração própria (2021)

Por fim, são feitas algumas associações entre as relações e a classificação gramatical das palavras. Sabe-se que os aspectos são substantivos e o interesse é encontrar o adjetivo que está sendo atribuído ao aspecto, para assim aplicar a análise de sentimento. Além disso, a principal relação de interesse é a `nsubj`, a qual representa um substantivo que é sujeito da frase. Sendo assim, com esse direcionamento foram feitas algumas ligações até chegarmos no último bloco onde encontram-se as relações procuradas, representadas na Figura 12. A saída obtida quando o código é aplicado no exemplo acima é a seguinte: `['soundquality', ['great']], ['batterylife', ['bad']]`.

Figura 12 – Último bloco de código na extração de aspectos

```

finalcluster = []
ListaDic = []
for frasesTagged in featureList:
    dic = {}
    for i in frasesTagged:
        dic[i[0]] = i[1]
    ListaDic.append(dic)

count = 0
for frasesClustered in fcluster:
    clusterIntermediario = []
    dic = ListaDic[count]
    for i in frasesClustered:
        if(dic[i[0]]=="NN"):
            clusterIntermediario.append(i)
            # print(clusterIntermediario)
    finalcluster.append(clusterIntermediario)
    count = count + 1

```

Fonte: Elaboração própria (2021)

4.4. ANÁLISE DE SENTIMENTO

Como descrito anteriormente, a análise de sentimento foi realizada através de um método não supervisionado, ou seja, não exige um pré-treinamento para que o código funcione satisfatoriamente. Sendo assim, a escolha foi utilizar o Text Blob, já que faz suas classificações baseadas em um método de dicionário e dispensa o *input* de dados já classificados, além de possuir um diferencial em fornecer o grau de subjetividade da frase.

Para essa seção do código, foram criadas duas funções, uma para identificar a subjetividade e outra a polaridade, as quais estão presentes na Figura 13. Com as funções definidas, elas foram aplicadas em todos os adjetivos extraídos na etapa anterior a fim de obter suas classificações. O resultado final é um *data frame* com 4 colunas, sendo elas Aspecto, Adjetivo, Subjetividade e Polaridade.

Figura 13 – Funções de classificação de Subjetividade e Polaridade

```

def getSubjectivity(text):
    return TextBlob(text).sentiment.subjectivity

def getPolarity(text):
    return TextBlob(text).sentiment.polarity

listaSub = []
listaPolaridade = []
for i in listaAdjetivos:
    try:
        listaSub.append(getSubjectivity(i))
        listaPolaridade.append(getPolarity(i))
    except:
        listaSub.append(0.0)
        listaPolaridade.append(0.0)
dfFinal['Sub'] = listaSub
dfFinal['Polaridade'] = listaPolaridade
dfFinal

```

Fonte: Elaboração própria (2021)

5. APLICAÇÃO PILOTO

O código explicado acima possui flexibilidade para ser aplicado em qualquer contexto, mas para fins práticos foi realizada uma aplicação real no contexto de dois aparelhos celulares sendo eles: Iphone 13 e Galaxy S21.

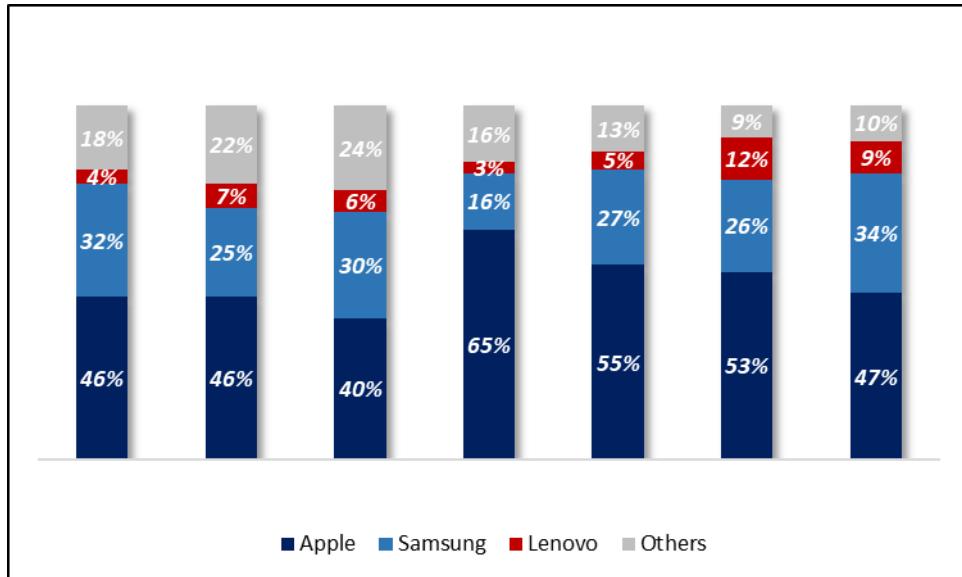
Esse contexto foi escolhido por 3 principais razões, as quais estão listadas abaixo:

- *Features* bem definidas
- Produtos utilizados por um grande volume de usuários
- Modelos mais atuais de suas respectivas empresas

A Apple, marca responsável pelo Iphone, e Samsung, marca responsável pelo Galaxy, são as líderes no segmento de *smartphones*. Um relatório publicado por Walia e Chaudhary (2021) mostra que somente as duas marcas atingiram mais de 80% do *market share* desse tipo de produto nos Estados Unidos no terceiro trimestre de 2021, como sinaliza o gráfico da Figura 14. É interessante destacar que a Apple só possui a linha Iphone na categoria de *smartphone*, enquanto a Samsung, somente o Galaxy possui diversas linhas com os mais diversos níveis de

sofisticação e preço, sendo a Galaxy S uma das principais linhas da marca Samsung. A linha Galaxy S apresenta modelos topo de linha em níveis similares ao Iphone.

Figura 14 – Market share dos smartphones por marcas



Fonte: Walia e Chaudhary (2021)

As *features* escolhidas para análise foram: Câmera, Bateria, Design, Performance, Preço, Tela e Display. Esses atributos são encontrados explicitamente nos materiais de divulgação desses aparelhos, além disso são os principais atributos encontrados em artigos que trabalham no contexto de *mobile phone* (SIVAKUMAR; UYYALA, 2021). Por isso, foram selecionadas para serem utilizadas como filtros e, assim, guiar os resultados de *tweets* para captura de frases com conteúdo e opinião sobre as *features* dos produtos. Dessa forma a utilização da análise de sentimentos baseada em aspectos foi mais bem aproveitada e mais adequada, pois uma visão geral desses aparelhos celulares não representa as visões isoladas dos atributos.

5.1. MINERAÇÃO

A API do Twitter limita as buscas em até 7 dias passados, sendo assim, para conseguir uma base de dados com um volume expressivo foram realizados diversos *rounds* de mineração com uma frequência semanal. Foram utilizados dois filtros diferentes, um deles para o Iphone 13 e outro para o Galaxy S21, como apresentado no capítulo 4.1, e cada extração gerou um arquivo CSV. Foram feitas 4 extrações, gerando um total de 8 arquivos CSV, sendo cada metade para cada aparelho móvel. Após a criação da base de dados com um volume expressivo de

tweets, esses arquivos são carregados para dentro do código e são removidos os textos repetidos. O resultado obtido do compilado da mineração sem valores duplicados foi de 11847 tweets para Iphone 13 e 6234 tweets para o GalaxyS21, os quais estão expostos nas Figuras 15 e 16, respectivamente.

Figura 15 – Resultado das minerações compilados para o Iphone 13

	created_at	User_Id	tweet	User_id
0	Mon Oct 25 23:04:18 +0000 2021	1.452773e+18	@TechKingMike The iPhone 13 Pro Max battery li...	NaN
1	Mon Oct 25 22:48:26 +0000 2021	1.452769e+18	@iamghaire @b_j_o_h_n_s_o_n If the battery get...	NaN
2	Mon Oct 25 22:19:29 +0000 2021	1.452762e+18	@RjeyTech That's a lie. \niPhone 13 Pro Max ba...	NaN
3	Mon Oct 25 22:16:56 +0000 2021	1.452761e+18	Samsung users does get paid to be so annoying?...	NaN
4	Mon Oct 25 21:58:52 +0000 2021	1.452757e+18	@Heyitsamemario @RjeyTech It's not true lol, t...	NaN
...
11842	2021-11-13 08:42:37+00:00	NaN	Sell/buy/repair Samsung smartphones best price...	1.459442e+18
11843	2021-11-13 04:07:58+00:00	NaN	Sell/buy/repair Samsung smartphones best price...	1.459372e+18
11844	2021-11-13 03:45:06+00:00	NaN	Sell/buy/repair Samsung smartphones best price...	1.459367e+18
11845	2021-11-21 21:33:24+00:00	NaN	@apple What's going on with the camera softwar...	1.462535e+18
11846	2021-11-21 21:32:17+00:00	NaN	The iphone 13 camera quality makes me want to ...	1.462534e+18
11847 rows × 4 columns				

Fonte: Elaboração própria (2021)

Figura 16 – Resultado das minerações compilados para o Galaxy S21

	created_at	User_id	tweet
0	Mon Oct 25 18:03:17 +0000 2021	1452697285310230540	@khi_daking4 @a_for_albert @realutsavdoshi My ...
1	Mon Oct 25 14:42:49 +0000 2021	1452646838922776587	Samsung Galaxy S21 Ultra Power Bank Case https...
2	Mon Oct 25 12:16:32 +0000 2021	1452610025667563525	The September Security Patch for the Samsung G...
3	Mon Oct 25 07:29:59 +0000 2021	1452537911455338497	Galaxy S21 battery replacement S21 Plus Ge...
4	Sun Oct 24 18:02:38 +0000 2021	1452334733036343306	With a sleek new design, hyperfast processing ...
...
6229	2021-11-14 03:59:16+00:00	1459732639452172288	[Dome Glass] Galaxy S21 Ultra Screen Protector...
6230	2021-11-14 03:58:33+00:00	1459732457855586304	[Dome Glass] Galaxy Note 20 Ultra Screen Prote...
6231	2021-11-14 03:58:18+00:00	1459732396622897152	[Dome FILM] [1SET 3PCS] for Samsung Galaxy Z F...
6232	2021-11-14 03:58:04+00:00	1459732336011104256	[Dome EZ Glass] Galaxy Z Fold 3 Glass Screen P...
6233	2021-11-14 03:57:50+00:00	1459732278268076032	[DOME FILM] [1SET 4PCS] for Samsung Galaxy Z F...
6234 rows × 3 columns			

Fonte: Elaboração própria

5.2. PRÉ-PROCESSAMENTO

Em sequência, a função Preprocessing é aplicada em cada uma das linhas da base de dados construída, sendo que o resultado é armazenado em uma nova coluna no *data frame*, nomeada de “preprocessed”. Nas Figuras 16 e 17 é possível visualizar a eliminação de menções e a conversão das letras para minúsculas.

Figura 16 – Resultado pré-processamento para os tweets do Iphone 13

	created_at	tweet	preprocessed
0	Mon Oct 25 23:04:18 +0000 2021	@TechKingMike The iPhone 13 Pro Max battery li...	the iphone 13 pro max battery life is un...
1	Mon Oct 25 22:48:26 +0000 2021	@jamghaire @b_j_o_h_n_s_o_n If the battery get...	if the battery gets better then the battery li...
2	Mon Oct 25 22:19:29 +0000 2021	@RjeyTech That's a lie. \niPhone 13 Pro Max ba...	that's a lie \niphone 13 pro max battery life ...
3	Mon Oct 25 22:16:56 +0000 2021	Samsung users does get paid to be so annoying?...	samsung users does get paid to be so annoying?...
4	Mon Oct 25 21:58:52 +0000 2021	@Heyitsamemario @RjeyTech It's not true lol, t...	it's not true lol, the pixel got completely de...
...
11842	2021-11-13 08:42:37+00:00	Sell/buy/repair Samsung smartphones best price...	sell/buy/repair samsung smartphones best price...
11843	2021-11-13 04:07:58+00:00	Sell/buy/repair Samsung smartphones best price...	sell/buy/repair samsung smartphones best price...
11844	2021-11-13 03:45:06+00:00	Sell/buy/repair Samsung smartphones best price...	sell/buy/repair samsung smartphones best price...
11845	2021-11-21 21:33:24+00:00	@apple What's going on with the camera softwar...	what's going on with the camera software for t...
11846	2021-11-21 21:32:17+00:00	The iphone 13 camera quality makes me want to ...	the iphone 13 camera quality makes me want to ...
11847 rows × 3 columns			

Fonte: Elaboração própria (2021)

Figura 17 – Resultado pré-processamento para os tweets do Galaxy S21

	created_at	tweet	preprocessed
0	Mon Oct 25 18:03:17 +0000 2021	@khi_daking4 @a_for_albert @realutsavdoshi My ...	my problems with the devices\ns21 horrible bat...
1	Mon Oct 25 14:42:49 +0000 2021	Samsung Galaxy S21 Ultra Power Bank Case https...	samsung galaxy s21 ultra power bank case samsu...
2	Mon Oct 25 12:16:32 +0000 2021	The September Security Patch for the Samsung G...	the september security patch for the samsung g...
3	Mon Oct 25 07:29:59 +0000 2021	Galaxy S21 battery replacement S21 Plus Ge...	galaxy s21 battery replacement s21 plus ge...
4	Sun Oct 24 18:02:38 +0000 2021	With a sleek new design, hyperfast processing ...	with a sleek new design, hyperfast processing ...
...
6229	2021-11-14 03:59:16+00:00	[Dome Glass] Galaxy S21 Ultra Screen Protector...	[dome glass] galaxy s21 ultra screen protector...
6230	2021-11-14 03:58:33+00:00	[Dome Glass] Galaxy Note 20 Ultra Screen Prote...	[dome glass] galaxy note 20 ultra screen prote...
6231	2021-11-14 03:58:18+00:00	[Dome FILM] [1SET 3PCS] for Samsung Galaxy Z F...	[dome film] [1set 3pcs] for samsung galaxy z f...
6232	2021-11-14 03:58:04+00:00	[Dome EZ Glass] Galaxy Z Fold 3 Glass Screen P...	[dome ez glass] galaxy z fold 3 glass screen p...
6233	2021-11-14 03:57:50+00:00	[DOME FILM] [1SET 4PCS] for Samsung Galaxy Z F...	[dome film] [1set 4pcs] for samsung galaxy z f...
6234 rows × 3 columns			

Fonte: Elaboração própria (2021)

5.3. EXTRAÇÃO DE ASPECTOS

Durante a extração de aspectos podemos ver alguns *outputs* relevantes. O primeiro deles representado na Figura 18, onde as classificações das relações entre as palavras são identificadas.

Figura 18 – Um exemplo aplicado das relações entre as palavras identificadas

```
print(listaDep_node[0])
[['my', 'problems', 'nmod:poss'], ['problems', 's21', 'nsubj'], ['with', 'devices', 'case'], ['the', 'devices', 'det'], ['devices', 'problems', 'nmod'], ['s21', 0, 'root'], ['horrible', 'batterylife', 'amod'],
```

Fonte: Elaboração própria (2021)

A classificação gramatical das palavras também é essencial para a extração de aspectos e está visível na Figura 19. Nota-se que na lista da figura abaixo só estão as palavras chaves da frase, pois já foram removidas as chamadas *stopwords*.

Figura 19 – Um exemplo aplicado da classificação gramatical das palavras

```
print(featureList[0])
[['problems', 'NNS'], ['devices', 'NNS'], ['horrible', 'JJ'], ['batterylife', 'NN'], ['laggy', 'NN'], ['softw...', 'NN']]
```

Fonte: Elaboração própria (2021)

Após as classificações e associações, obtém-se um cluster com os aspectos identificados e suas possíveis atribuições. Na Figura 20 nota-se que a associação Aspecto e Adjetivos não ocorre perfeitamente, algo que já é esperado devido à complexidade de interpretar a linguagem humana em redes sociais. Mais à frente ainda será realizado mais um filtro que selecionará apenas as relações que possuem um adjetivo de valor.

Figura 20 – Um exemplo aplicado do agrupamento final na etapa de extração de aspectos

```
print(finalcluster[0])
[['batterylife', ['horrible', 's21']], ['laggy', ['softw...']], ['softw...', ['laggy']]]
```

Fonte: Elaboração própria (2021)

A partir do último agrupamento, ilustrado na última figura, são criadas duas listas que são summarizadas no *data frame* presente na Figura 21. Esse *data frame* mostra uma coluna de Aspectos e outra de Adjetivos e as relações ocorrem com as palavras da mesma linha.

Figura 21 – Resultado da extração de aspectos para os tweets do Galaxy S21

	Aspectos	Adjetivos
0	batterylife	horrible
1	batterylife	s21
2	laggy	softw...
3	softw...	laggy
4	samsunggalaxy	powerbank
...
42169	domefilm	17
42170]	set
42171	samsunggalaxy	3
42172	amazonwhitestone	17
42173	innovation	17

42174 rows × 2 columns

Fonte: Elaboração própria (2021)

5.4. ANÁLISE DE SENTIMENTO

As funções de obtenção de Subjetividade e Polaridade são aplicadas em cada linha da coluna de Adjetivos. O resultado é exibido nas colunas nomeadas de Sub, representando a subjetividade, e na coluna Polaridade, assim como mostrado na Figura 22.

Figura 22 – Resultado da análise de sentimentos para os tweets do Galaxy S21

	Aspectos	Adjetivos	Sub	Polaridade
0	batterylife	horrible	1.0	-1.0
1	batterylife	s21	0.0	0.0
2	laggy	softw...	0.0	0.0
3	softw...	laggy	0.0	0.0
4	samsunggalaxy	powerbank	0.0	0.0
...
42169	domefilm	17	0.0	0.0
42170]	set	0.0	0.0
42171	samsunggalaxy	3	0.0	0.0
42172	amazonwhitestone	17	0.0	0.0
42173	innovation	17	0.0	0.0

42174 rows × 4 columns

Fonte: Elaboração própria (2021)

É notório que existem muitas relações sem valor, onde as palavras que compõe a coluna Adjetivo não são de fato adjetivos. Para filtrar os resultados é feito um filtro das linhas que não possuem nenhuma indicação de polaridade, resultado pode ser visualizado nas Figuras 23 e 24 para o Iphone 13 e Galaxy S21, respectivamente.

Figura 23 – Filtro do resultado da análise de sentimentos para os tweets do Iphone 13

	Aspectos	Adjetivos	Sub	Polaridade
6	batterylife	fine	0.5	0.416667
14	camera	better	0.5	0.500000
21	pixel	better	0.5	0.500000
31	battery	good	0.6	0.700000
35	battery	good	0.6	0.700000
...
62197	iphone	best	0.3	1.000000
62198	iphone	more	0.5	0.500000
62199	iphone	best	0.3	1.000000
62200	iphone	more	0.5	0.500000
62208	sgfonefixsgipad...	best	0.3	1.000000
5905 rows × 4 columns				

Fonte: Elaboração própria (2021)

Figura 24 – Filtro do resultado da análise de sentimentos para os tweets do Galaxy S21

	Aspectos	Adjetivos	Sub	Polaridade
0	batterylife	horrible	1.000000	-1.000000
17	batterylife	horrible	1.000000	-1.000000
24	s21	genuine	0.500000	0.400000
26	s21	genuine	0.500000	0.400000
29	design	new	0.454545	0.136364
...
42077	tempered...	first	0.333333	0.250000
42092	price	smart	0.642857	0.214286
42116	protec...	perfect	1.000000	1.000000
42118	screen	best	0.300000	1.000000
42121	glassp...	first	0.333333	0.250000
2273 rows x 4 columns				

Fonte: Elaboração própria (2021)

Após o filtro, encontram-se 5905 associações para o Iphone 13 e 2273 associações para o Galaxy S21. É visível que as relações apresentadas nas Figuras 23 e 24 é a seleção de valor das relações encontradas na etapa representada pela Figura 22. Esse *data frame* final é utilizado para criação de gráficos e análise de resultados. Todavia, os últimos filtros ainda serão aplicados durante a sumarização dos resultados.

5.5. SUMARIZAÇÃO DOS DADOS

Os *Data frames* com as colunas de Aspecto, Adjetivo, Polaridade e Subjetividade, expostos nas Figuras 23 e 24, foram convertidos para um CSV e carregados no Excel. Em seguida foram classificados nas categorias das 7 *features* definidas, pontuadas no início deste capítulo. Para a categorização foi feita uma classificação automática simples, a qual identifica dentro dos aspectos que contém no texto qualquer uma das *features* e classifica de acordo com a *feature* encontrada. Nos casos de bateria, por exemplo, funciona bem pois as variações acabam sendo: *batterylife*, *batteryquality*, *iphonebattery*, *daybattery* e entre outras opções. Nesses casos a palavra bateria encontra-se presente em todos os exemplos, porém no caso de performance as variações acabam sendo muito além da palavra performance, podendo ser utilizado os termos desempenho, velocidade e eficiência. Sendo assim, além da classificação automática foi feita em sequência uma classificação manual e detalhada, lendo os aspectos e os

classificando individualmente. Uma categoria também foi criada, além das 7 *features*, a qual abriga aspectos relacionados ao aparelho celular como um todo, sem especificar nenhum atributo. Nessa etapa muitos aspectos que não faziam sentido foram descartados e não entraram em nenhuma das classificações.

As informações da quantidade de adjetivos relacionados às categorias descritas acima estão registradas na Tabela 2.

Tabela 2 – Quantidade de adjetivos relacionados às categorias dos aspectos

	Iphone 13	Galaxy S21	Total
Bateria	316	29	345
Camera	895	82	977
Display	49	24	73
Design	96	38	134
Tela	107	13	120
Performance	71	6	77
Price	100	15	115
Aparelho	1246	234	1480
Total	2880	441	

Fonte: Elaboração própria (2021)

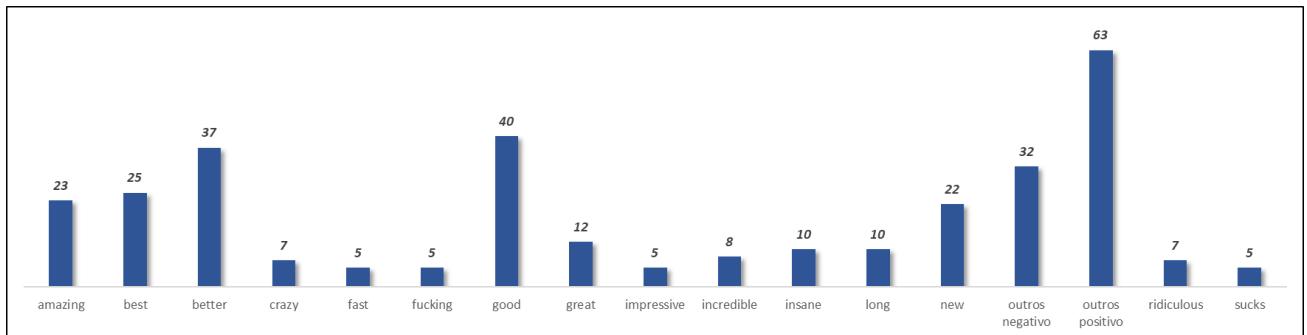
As polaridades e subjetividades médias em relação aos dois aparelhos celulares, assim como o desvio padrão desses valores, são encontrados na Tabela 3.

Tabela 3 – Média e Desvio Padrão da Subjetividade e Polaridade dos celulares

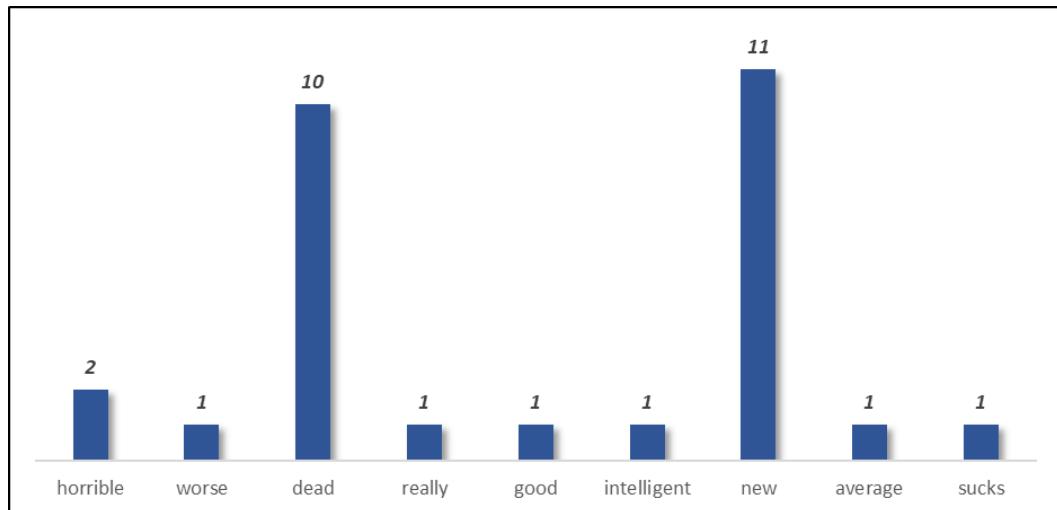
Infos	Iphone 13		Galaxy S21	
	Subjetividade	Polaridade	Subjetividade	Polaridade
Média	0.533811535	0.422016555	0.484720513	0.568431504
Desvio Padrão	0.248337356	0.397702282	0.254477253	0.381397532

Fonte: Elaboração própria (2021)

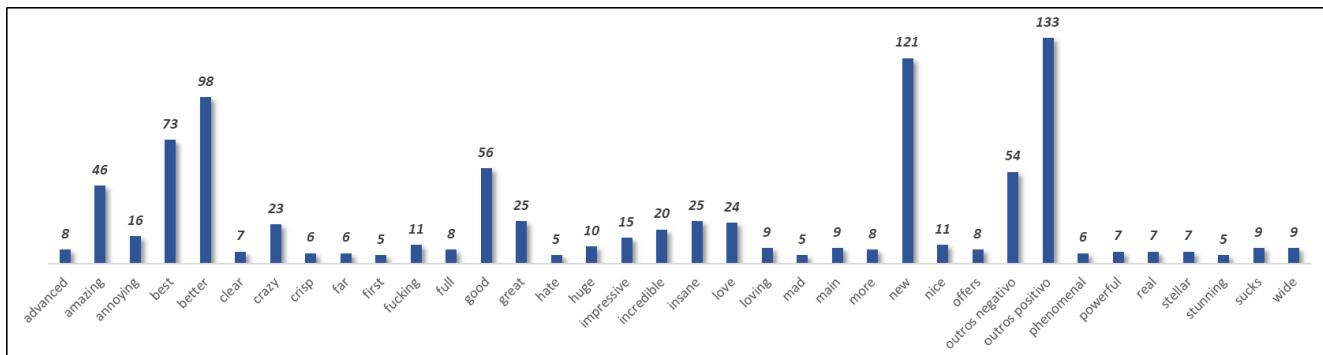
Os resultados indicando quais adjetivos e em quais quantidades foram atribuídos a cada um dos 7 aspectos e separados por celular estão expostos da Figura 25 até a Figura 40.

Figura 25 – Resultados Iphone 13 Bateria

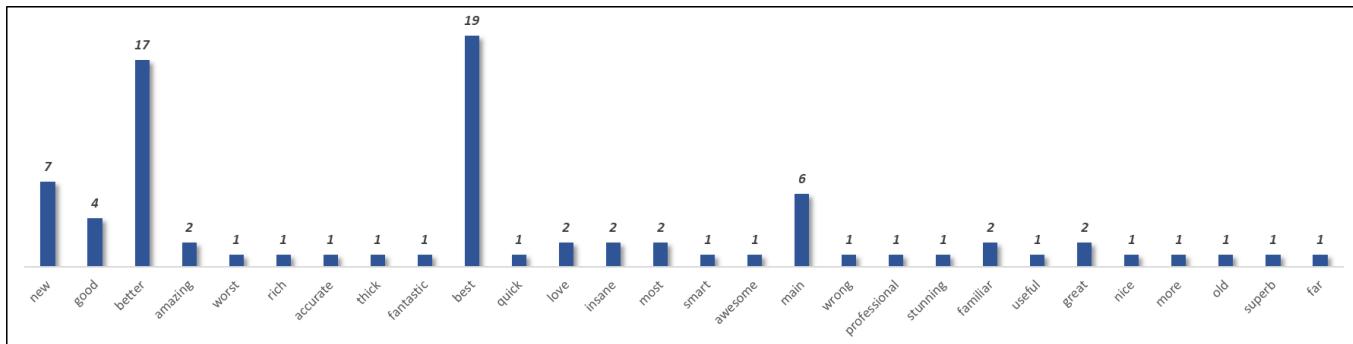
Fonte: Elaboração própria (2021)

Figura 26 – Resultados Galaxy S21 Bateria

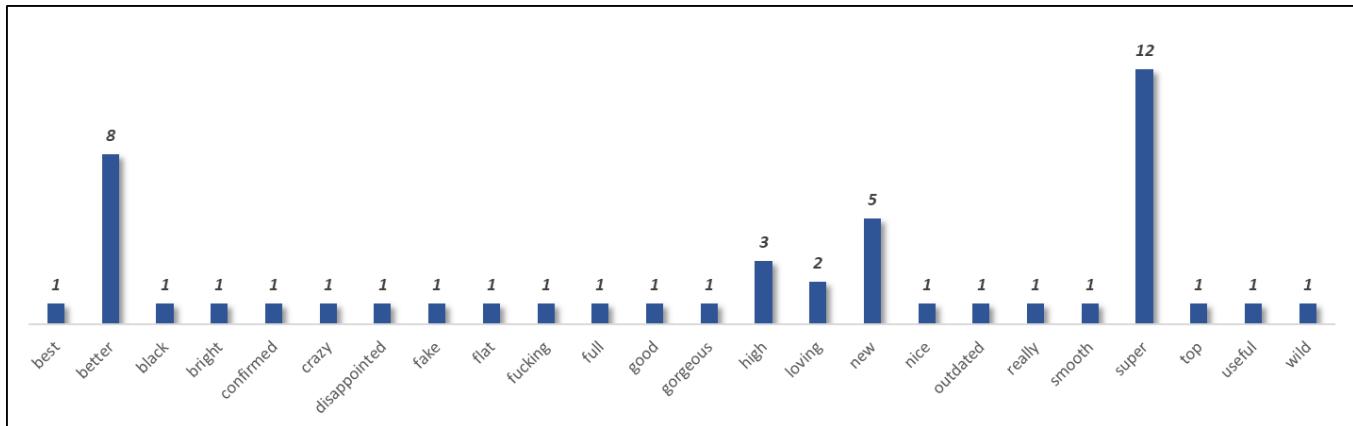
Fonte: Elaboração própria (2021)

Figura 27 – Resultados Iphone 13 Câmera

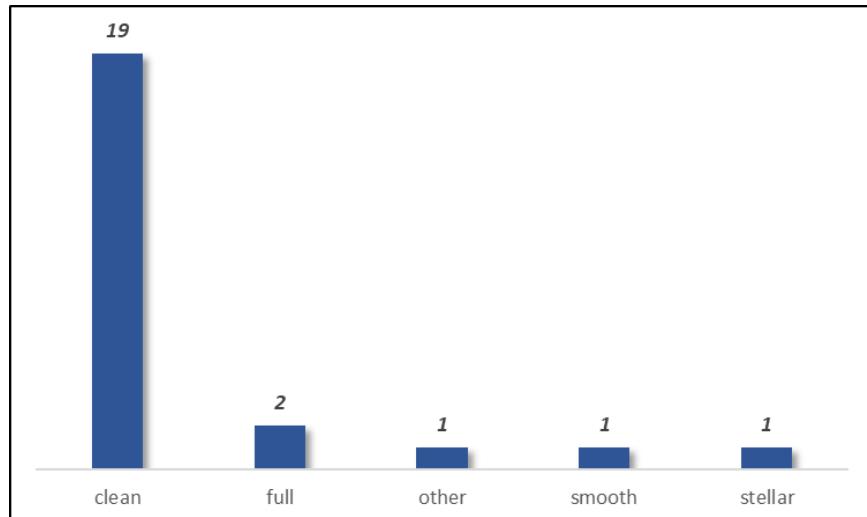
Fonte: Elaboração própria (2021)

Figura 28 – Resultados Galaxy S21 Câmera

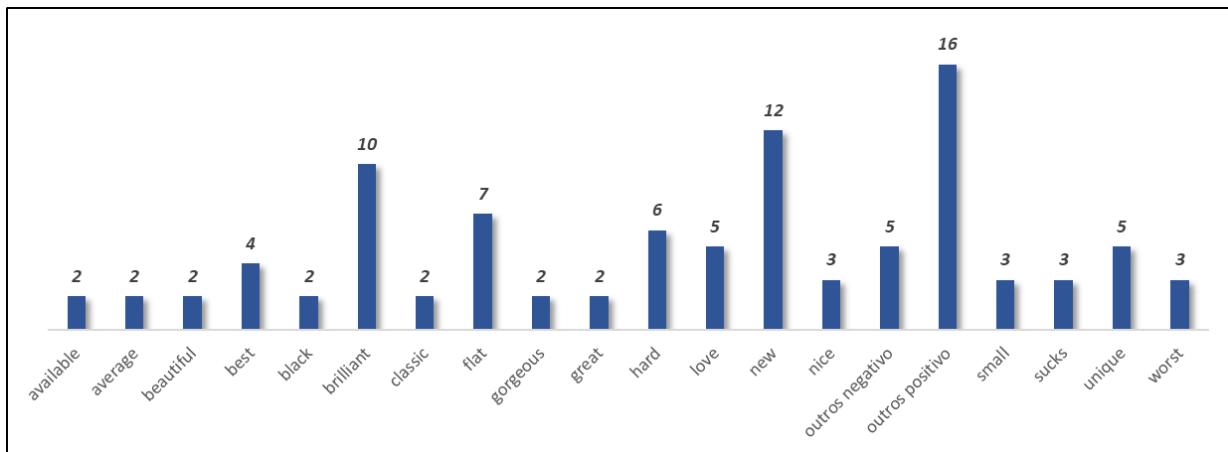
Fonte: Elaboração própria (2021)

Figura 29 – Resultados Iphone 13 Display

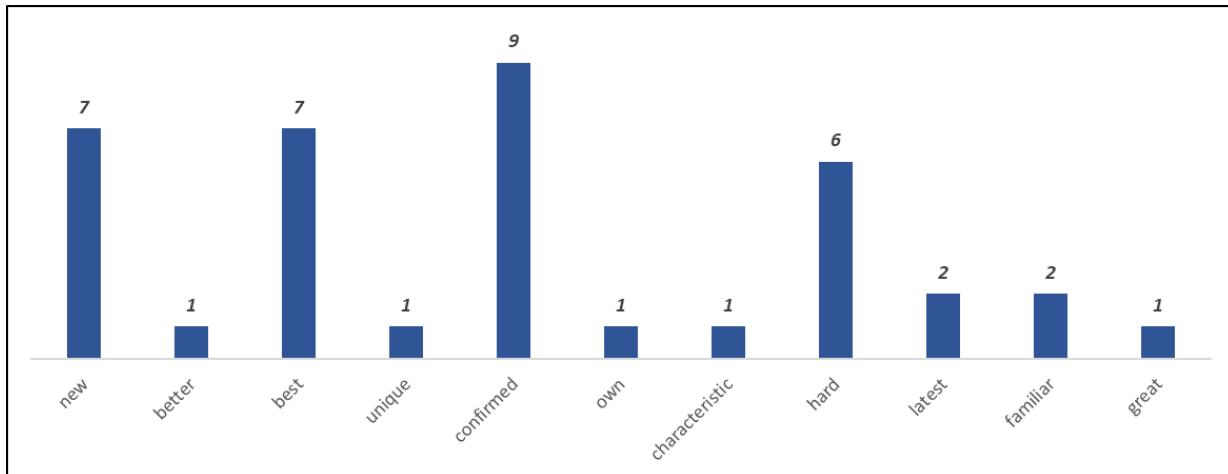
Fonte: Elaboração própria (2021)

Figura 30 – Resultados Galaxy S21 Display

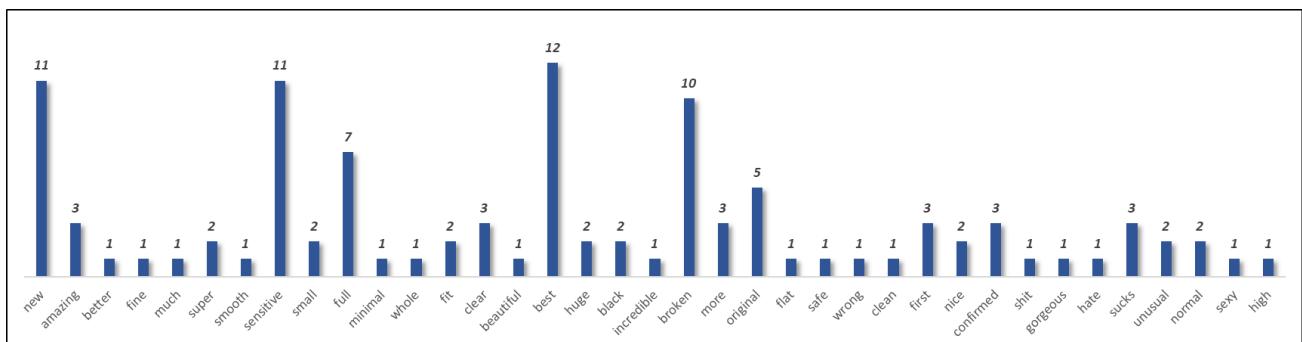
Fonte: Elaboração própria (2021)

Figura 31 – Resultados Iphone 13 Design

Fonte: Elaboração própria (2021)

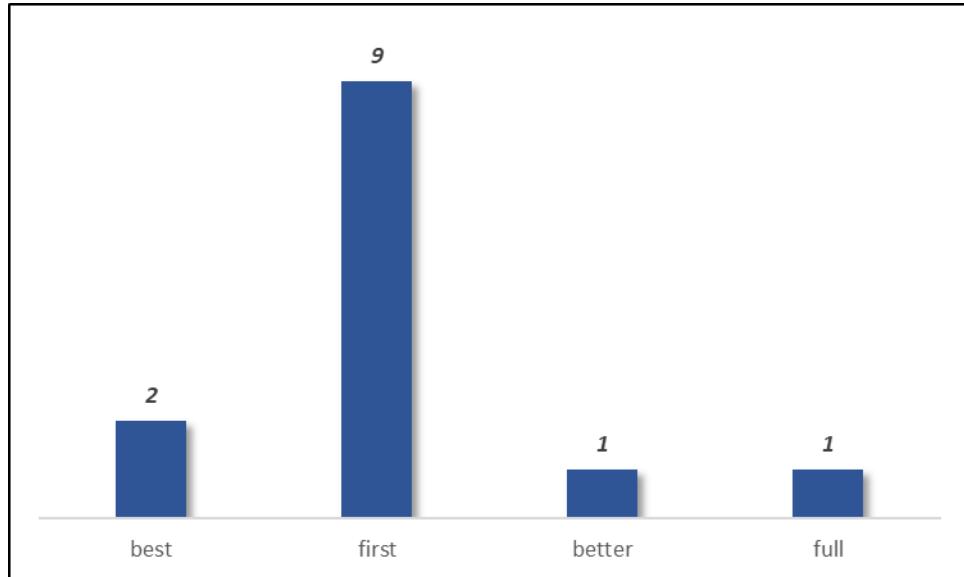
Figura 32 – Resultados Galaxy S21 Design

Fonte: Elaboração própria (2021)

Figura 33 – Resultados Iphone 13 Tela

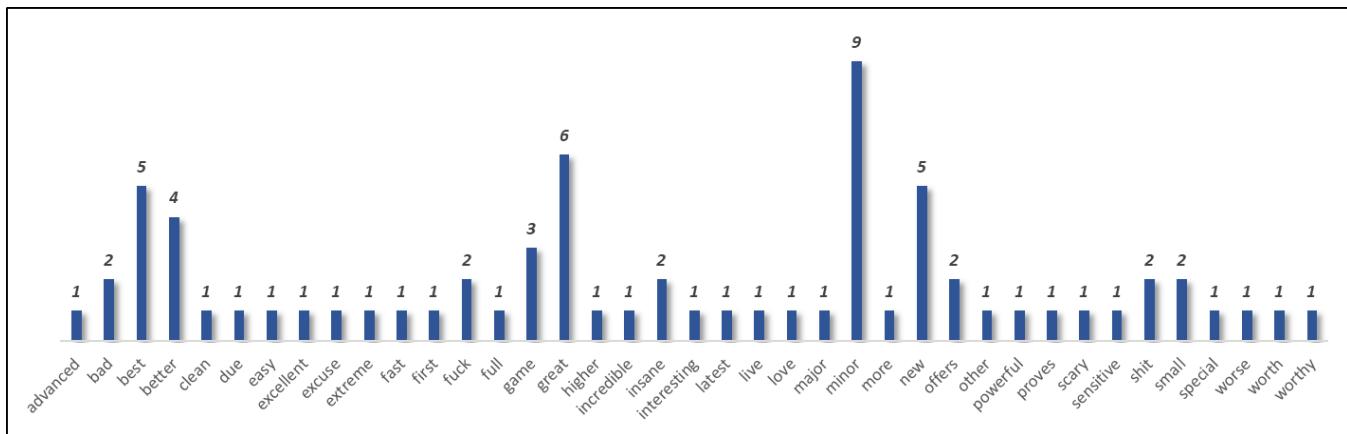
Fonte: Elaboração própria (2021)

Figura 34 – Resultados Galaxy S21 Tela



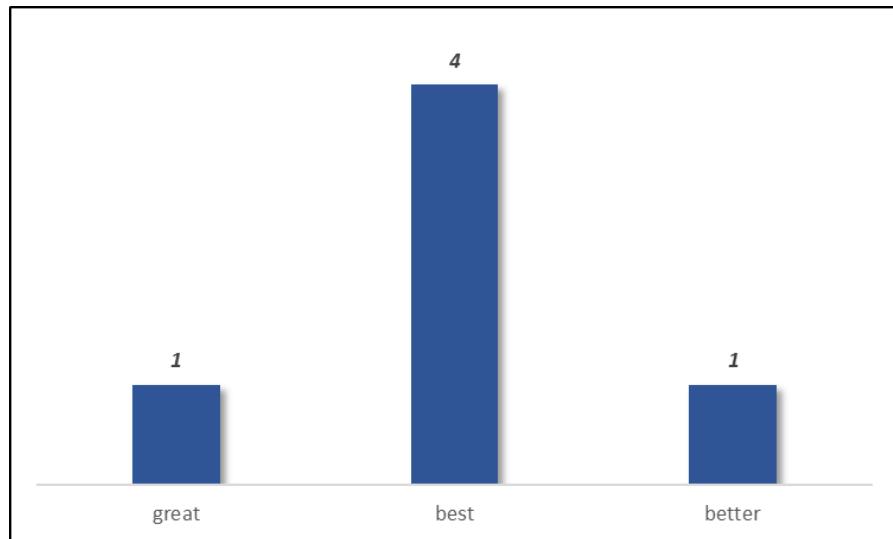
Fonte: Elaboração própria (2021)

Figura 35 – Resultados Iphone 13 Performance



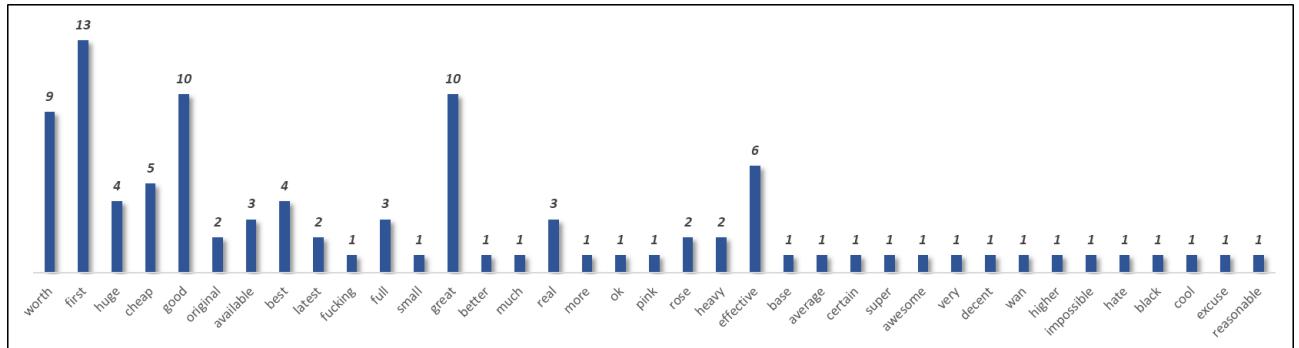
Fonte: Elaboração própria (2021)

Figura 36 – Resultados Galaxy S21 Performance



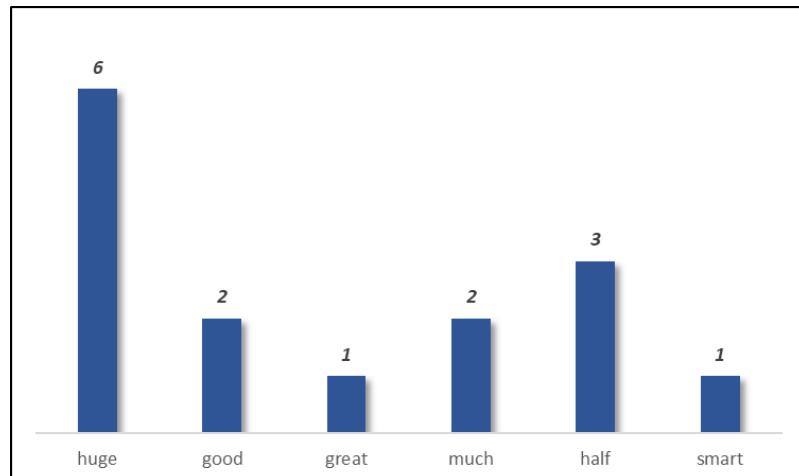
Fonte: Elaboração própria (2021)

Figura 37 – Resultados Iphone 13 Preço



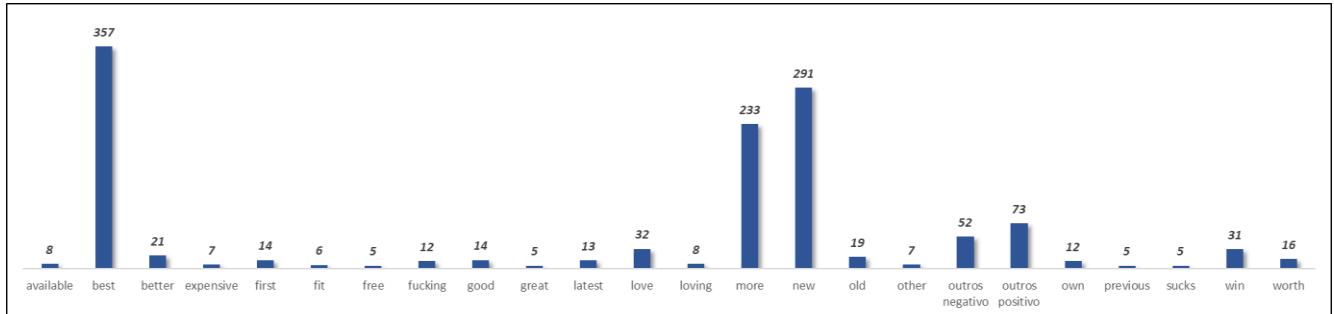
Fonte: Elaboração própria (2021)

Figura 38 – Resultados Galaxy S21 Preço



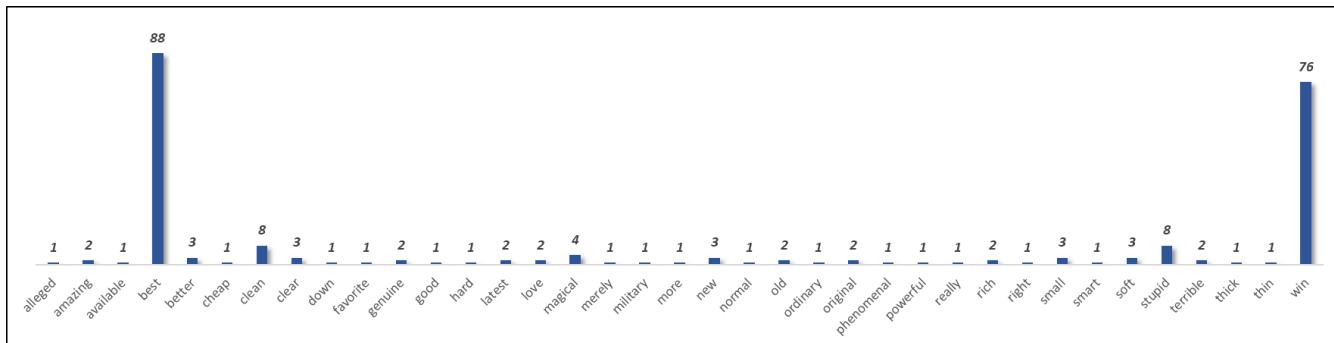
Fonte: Elaboração própria (2021)

Figura 39 – Resultados Iphone 13 Aparelho



Fonte: Elaboração própria (2021)

Figura 40 – Resultados Galaxy S21 Aparelho



Fonte: Elaboração própria (2021)

6. ANÁLISE DE RESULTADOS E DISCUSSÕES

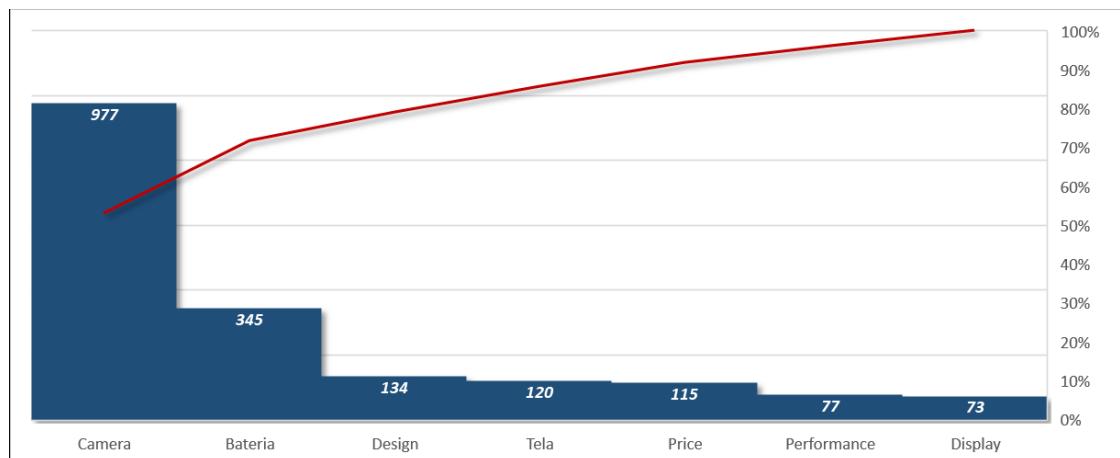
6.1. DISCUSSÕES GERAIS

É notório que o número de resultados foi muito maior para Iphone 13 em relação ao Galaxy S21. No capítulo 5 vimos que o *market share* da Apple ainda é maior perante ao da Samsung, ademais a Samsung possui uma variedade de smartphones bem maior comparado à Apple, ou seja, o valor apresentado na Figura 14 está mais pulverizado entre diferentes aparelhos para a marca responsável pelo Galaxy S21. Esses pontos sustentam o resultado obtido onde prevaleceu um volume superior de *tweets* citando o Iphone 13 comparado ao Galaxy S21.

Ainda em relação à quantidade de resultados, existe uma diferença expressiva quando comparamos o número de *tweets* minerados (18.081) com o número de combinações válidas de Aspecto-Adjetivo (8.178). Esse valor mostra que a cada dois *tweets* é encontrada algo próximo de 1 relação válida, ou seja, mais da metade dos textos não possuem nenhum conteúdo de valor. Pela característica da rede social está dentro da expectativa que existam muitos textos que não expressem nenhum sentimento relacionado ao produto, já que não estamos em um contexto em que são feitas somente avaliações do produto.

Câmera, Bateria e Design somados aproximam-se de 80% dos resultados, com um destaque notável para a Câmera, dominando mais de 50% das associações Aspecto-Adjetivo encontradas nos Tweets, ilustrado na Figura 41. Esse desfecho é muito interessante pois está em linha com a evolução dos celulares e com as estratégias que as empresas vêm seguindo.

Figura 41 – Quantidade de citações por *feature* e a representatividade em % do total acumulada.



Fonte: Elaboração própria (2021)

As redes sociais transformaram-se ao longo dos tempos e os formatos atuais têm como principal atrativo a publicação de fotos e vídeos, como podemos ver com You-Tube, Instagram e Tik-Tok, nomes que estão entre os aplicativos mais usados mundialmente. A tendência desses aplicativos está cada vez mais associada à câmera do celular, tornando-a uma das ferramentas mais utilizadas dentre as funcionalidades de um smartphone.

Em linha com as tendências, as empresas estão investindo forte no desenvolvimento das câmeras em seus aparelhos celulares e no marketing sobre essas evoluções. Tanto no site oficial do Iphone 13 quanto no do Galaxy S21, a primeira mensagem está associada ao atributo câmera, como pode ser visto nas Figuras 42 e 43, além de prevalecer comentários sobre essa *feature*.

perante as outras ao longo de todo o material de divulgação. Tal ponto ressalta o caráter preditivo do modelo computacional desenvolvido.

Figura 42 – Primeira mensagem no site oficial do Iphone 13



Fonte: www.apple.com/br/iphone-13/

Figura 43 – Primeira mensagem no site oficial do Galaxy S21



Fonte: www.samsung.com/br/smartphones/galaxy-s21-5g/

Com a Análise de Sentimento Baseadas em aspectos foi possível criar a Tabela 4, a qual apresenta a percepção de sentimento em relação às *features* de forma isolada e divididas por aparelho celular. O Galaxy S21, de forma geral, apresentou um melhor resultado tendo mais

sentimentos positivos relacionados ao aparelho celular quando comparamos com o Iphone 13, embora ambos tiveram uma avaliação positiva em todos os aspectos.

Tabela 4 – Polaridade, Subjetividade e suas médias por *feature* de ambos os celulares

	Subjetividade Iphone	Polaridade Iphone	Subjetividade Galaxy	Polaridade Galaxy	Média
Aparelho	0.451687415	0.515888869	0.435674202	0.697297872	0.52513709
Bateria	0.632851405	0.478097296	0.482758621	0.031551724	0.40631476
Camera	0.615375458	0.426388066	0.510768398	0.544477616	0.52425238
Design	0.642495265	0.294649621	0.632854864	0.403628389	0.49340703
Display	0.595599876	0.326227582	0.646875	0.346006944	0.47867735
Performance	0.523157125	0.313044479	0.408333333	0.883333333	0.53196707
Preço	0.543214286	0.419009524	0.592857143	0.344285714	0.47484167
Tela	0.581448598	0.313800623	0.302918288	0.9692607	0.54185705
Média	0.573228679	0.385888257	0.501629981	0.527480287	

Fonte: Elaboração própria (2021)

Nota-se um contraste na percepção em relação à bateria, sendo que o Iphone 13 apresentou um resultado expressivamente mais positivo em relação ao Galaxy S21, a qual possui uma classificação próxima de neutra. Porém, comparando com as especificações dos produtos em seus sites originais vemos uma divergência, onde a bateria do Iphone possui uma duração de até 17h, utilizando reprodução de vídeos como parâmetro, durabilidade inferior à do Galaxy S21 que possui uma duração de até 20h ao reproduzir vídeos. Essa incompatibilidade entre a especificação do produto e a opinião do público pode ser explicada pela evolução dos aparelhos isoladamente. O Iphone 13 mostrou uma evolução de até 2,5 horas em relação à versão anterior, enquanto o Galaxy S20 e o Galaxy S21 possuem as mesmas especificações de uma duração da bateria de até 15h utilizando internet Wi-fi e uma capacidade da bateria de 4000 mAh. Ou seja, as opiniões são subjetivas e a visão positiva ou negativa pode variar de acordo com a referência do usuário.

No preço encontra-se um outro fenômeno interessante. Os preços dos modelos mais simples de Iphone 13 e Galaxy S21 são muito próximos e até iguais no site de uma das maiores lojas de eletrônicos dos Estados Unidos. Todavia, o grau de satisfação relacionado ao preço do S21 está pouco menos de 0.1 menor quando olhamos para o celular da Apple. Esse cenário das pessoas estarem adquirindo produtos muito próximos nas especificações técnicas e pelo mesmo preço, porém terem percepções diferentes desse valor, pode ser explicada por uma entrega diferencial do Iphone, o *status*. Atualmente, o Iphone além de entregar qualidade no aparelho, ele também é responsável por entregar o *status* de possuir o aparelho, sendo que existe uma assinatura em cada novo aparelho da Apple que possibilita identificar que é o modelo mais atual. Exemplo claro é a mudança na configuração das câmeras do Iphone X para o 11 Pro, uma

diferenciação clara e visível entre os modelos. Em relação à essa assinatura, é possível encontrar na internet a venda um adesivo que não possui nenhuma função prática, não agrupa nenhum valor funcional, além de tornar o celular mais parecido visualmente com a versão mais recente. Na Figura 44 é visto a venda desse produto em uma das maiores plataformas de e-commerce dos EUA, reforçando a importância do status que um novo Iphone proporciona.

Figura 44 – Venda de lentes para simular o Iphone X como Iphone 11 Pro

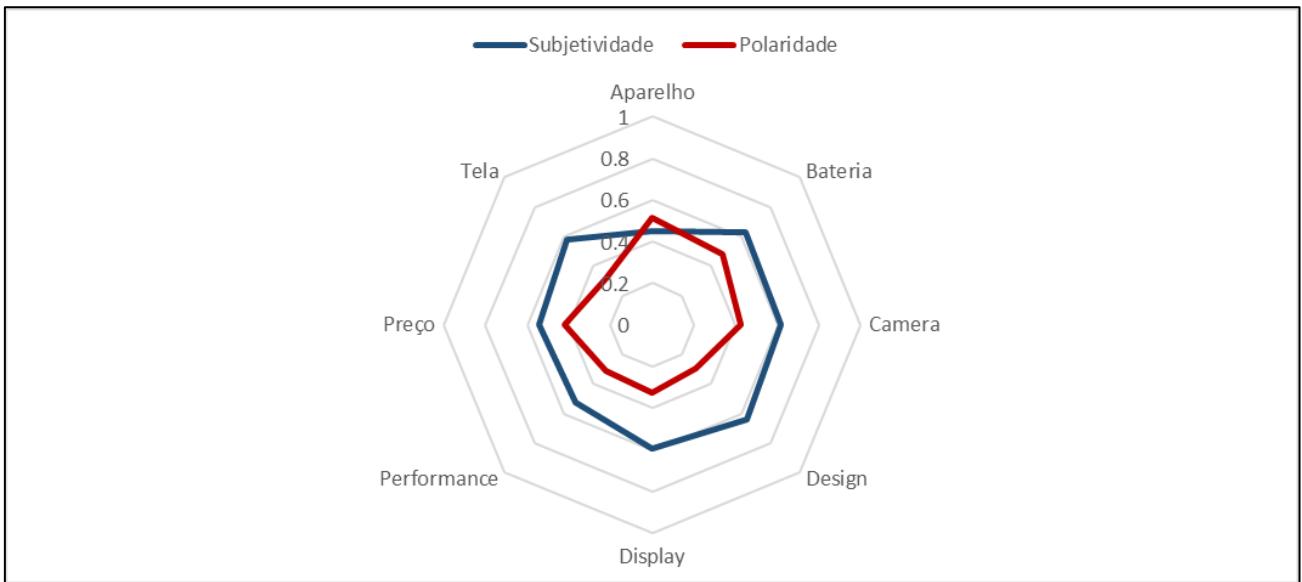


Fonte: www.ebay.com

O *Display* possui uma visão dos clientes próxima para os dois celulares, ambas com valores entre 0,30 e 0,35. A similaridade nesses valore faz total sentido com os fatos técnicos dos produtos. A Samsung, dona da marca Galaxy, foi a principal fornecedora de telas OLED de 120 Hz para a Apple na produção dos novos iPhone 13, ou seja, o Galaxy S21 e o Iphone 13 possuem o mesmo fornecedor de *Display*.

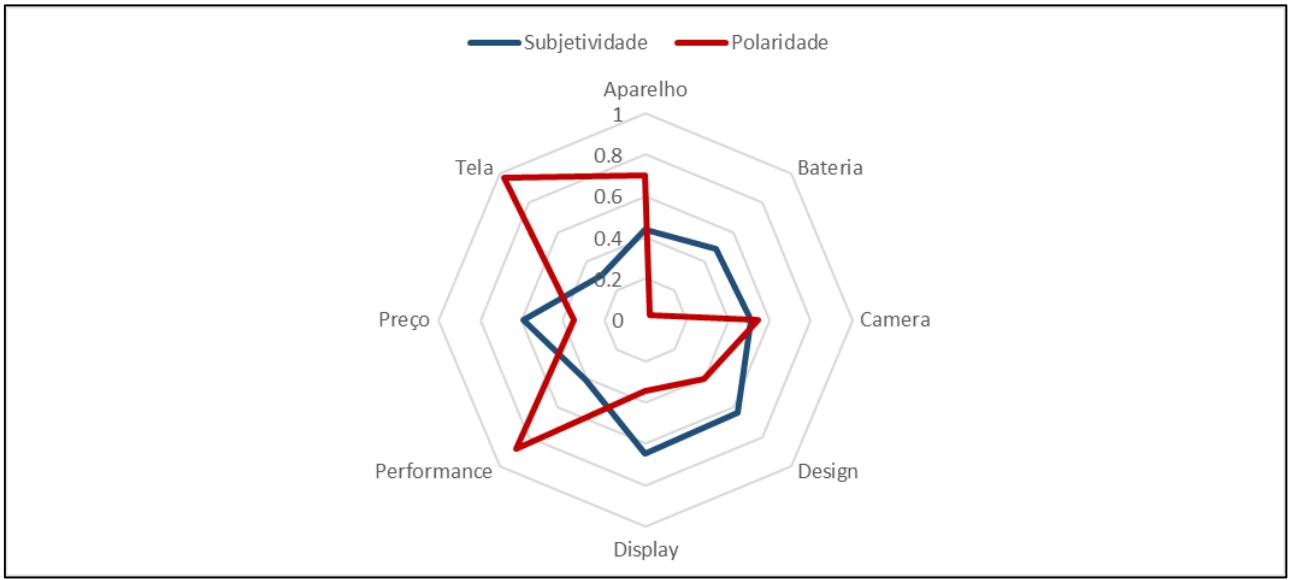
As *features* Tela, Câmera, Design e Performance possuem todas uma vantagem para o Galaxy S21, inclusive na categoria Aparelho que é uma visão mais geral. Esses resultados revelam que os usuários do Galaxy S21 possuem uma visão mais positiva em relação aos usuários do Iphone 13. O fato do número de resultados para o aparelho da Samsung ter sido expressivamente menor comparado ao do Iphone pode estar impactando nos resultados e trazendo valores menos representativos e fidedignos à opinião em massa sobre os *smartphones* em questão. Comparando as Figuras 45 e 46 é possível visualizar como o maior volume de dados colabora para uma suavização nos resultados, evitando discrepâncias e resultados muito polarizados.

Figura 45 – Polaridade e Subjetividade para as *features* do Iphone 13 (Gráfico Radar)



Fonte: Elaboração própria (2021)

Figura 46 – Polaridade e Subjetividade para as *features* do Galaxy S21 (Gráfico Radar)



Fonte: Elaboração própria (2021)

6.2. ANÁLISE DA CONFIABILIDADE

Para obter uma noção geral da confiabilidade dos resultados foi realizada uma análise manual de 10 frases com opiniões selecionadas durante a mineração no Twitter, representadas na Figura 47. As palavras destacadas em verdes nessa figura formam os pares Aspecto-Adjetivo que foram identificados e classificados corretamente pelo código. O par em amarelo foi

devidamente identificado, porém o adjetivo “prettiest” não foi classificado adequadamente e acabou sendo descartado no filtro de palavras sem polaridade. Por fim, as palavras em vermelho não foram identificadas como par Aspecto-Adjetivo. Na Tabela 5 está o resultado fornecido pelo código, sendo que apenas o par em vermelho está incorreto, sinalizando uma associação inexistente. Ou seja, nesse estudo em cima de 10 frases mineradas do Twitter, houve 4 erros entre 12 acertos, totalizando uma precisão de 75%.

Figura 47 – Frases selecionadas da mineração para teste

```
iPhone 13 Pro has incredible battery, lasting around 24 hours with 6 hours of screen on time. Just too good!',  
Iphone 13 battery is so good bruh',  
The iPhone 13 camera is wonderful',  
The iphone 13 battery life is really good yo.',  
The #iphone 13 pro has officially the worst design in terms of comfort in hand. Extremely inconvenient and causes joint pain #iPhone13Pro',  
@BiteSizeRi No, what you need is the samsung galaxy s21 ultra.. that muthafucka got best camera not cap!',  
@KellieADennis74 I got my new Galaxy S21 a few weeks ago. Its the prettiest phone I have ever had.',  
@EzTech231 The Samsung Galaxy S21 Ultra is my favorite phone. The Screens great and performance is great',  
@ashdogx I have the Galaxy S21 Ultra and this phone is pretty darn big. The Camera is great and the battery life is fantastic'  
Yeah, Galaxy S21 Ultras battery life still sucks.
```

Fonte: Elaboração própria (2021)

Tabela 5 – Resultado após rodar o código para as frases de teste

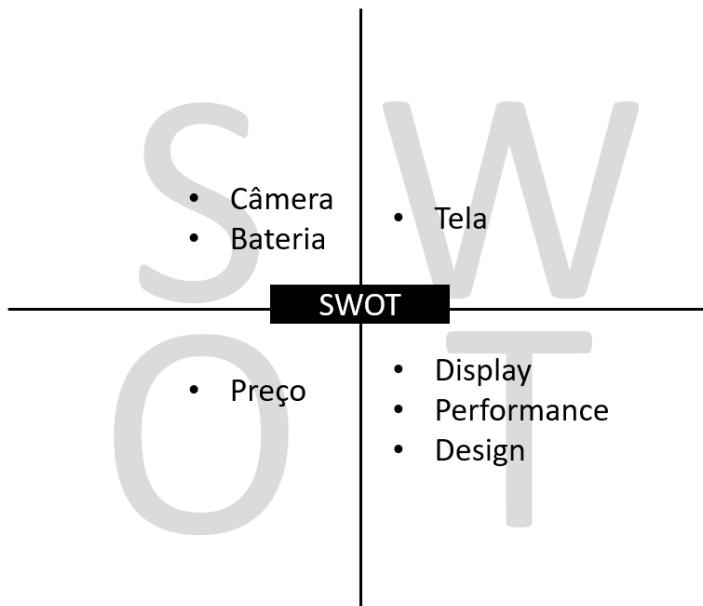
Adjetivos	Sub	Subjetividade	Polaridade
battery	incredible	0.9	0.9
battery	good	0.6	0.7
camera	wonderful	1	1
batterylife	good	0.6	0.7
design	worst	1	-1
camera	best	0.3	1
s21	new	0.454545	0.136364
phone	favorite	1	0.5
phone	great	0.75	0.8
performance	great	0.75	0.8
camera	great	0.75	0.8
batterylife	fantastic	0.9	0.4
batterylife	sucks	0.3	-0.3

Fonte: Elaboração própria (2021)

6.3. SWOT

Como o Iphone 13 apresentou um volume maior de resultados, uma matriz SWOT foi montada com base em suas avaliações, representada na Figura 48.

Figura 48 – Matriz SWOT



Fonte: Elaboração própria (2021)

6.3.1. STRENGTHS

Câmera e Bateria, como visto anteriormente, são as *features* mais citadas pelos usuários e são os principais pontos de atenção das empresas. Esses atributos estão em constante evolução e fomentam sentimentos positivos. São pontos que estão bem desenvolvidos e podem ser considerados pontos fortes da atual geração de celulares.

6.3.2. WEAKNESS

A Tela apresenta uma polaridade abaixo da média das *features* e a palavra mais recorrente com classificação negativa foi “*broken*”. Podemos ver que a resistência dos smartphones ainda é baixa e quedas frequentemente danificam a tela do aparelho. A resistência ainda é uma fraqueza desses celulares e já podemos ver uma preocupação com esse fator nos materiais de divulgação do Iphone e Galaxy, onde resistência está entre os principais tópicos abordados.

6.3.3. OPPORTUNITY

O preço está acima da média das *features* e mostra que o público usuário desse celular está satisfeito com o preço. Existe uma visão de valor em relação aos smartphones consideravelmente maior que o preço de custo. O fato de o preço ser bem aceito tem como um dos principais motivos a questão do status comentada anteriormente, de forma que o preço elevado do produto corrobora para a imagem do produto. Esse fator cria uma oportunidade para aumento nas margens de lucro, já que reduzir o preço pode minimizar o glamour do celular, de forma que reduções nos custos de produção não impliquem numa diminuição no preço, mas sim em um aumento de margem. O próprio Galaxy S21 se beneficia desse fenômeno e apresenta valores semelhantes ao do Iphone.

6.3.4. THREATS

O *design* é a *feature* com a menor polaridade. Desde 2007, ano em que a Apple revolucionou o mercado de celulares com o lançamento do primeiro Iphone, não vimos nenhuma mudança muito expressiva no modelo dos celulares. O *Design* desse aparelho tornou-se comum dentre os modelos de *smartphones* e acaba gerando sentimentos positivos próximos ao neutro. Sendo assim, um *design* inovador pode ameaçar os modelos atuais. Embora a revolução tenha sido iniciada com o Iphone, os celulares da linha Galaxy S apresentam um *design* semelhantes. A linha Galaxy Z já está trazendo um design diferenciado o qual permite que o celular seja dobrado.

Display e Performance apresentaram polaridade abaixo da média das *features*. Como são aspectos mais técnicos e talvez pouco perceptíveis para um usuário comum acabam não sendo o principal foco. Porém, são tecnologias que evoluem rapidamente e que a qualquer momento podem surgir inovações que tornem ultrapassadas as tecnologias atuais.

7. CONCLUSÃO

Este trabalho de conclusão de curso se propôs a desenvolver um modelo computacional que possibilitasse a mineração de dados de uma rede social, sendo Twitter a escolhida, de forma que fosse possível identificar qual a percepção de valor que os usuários possuem em relação à um determinado produto. Foi escolhido o contexto de smartphones, mais especificamente os modelos Iphone 13 e Galaxy S21 para realizar uma aplicação piloto.

Concluiu-se que a percepção desses usuários em relação aos aparelhos selecionados é positiva em todos os aspectos. O ponto principal é a intensidade do sentimento positivo em relação à cada *feature*. Os aspectos câmera e bateria demonstraram-se como os tópicos mais citados pelos clientes, mostrando a importância que essas ferramentas têm para os usuários e apresentaram uma polaridade acima da média das *features* em geral. Foi verificado, também, que a relevância desses dois fatores está alinhada com as estratégias das empresas, mostrando o potencial preditivo do trabalho desenvolvido.

Existiram alguns contrastes quando foram comparadas as opiniões presentes nos *tweets* com as especificações técnicas dos produtos nos aspectos de bateria e preço. Esses pontos permitiram associar que o sentimento dos usuários não está correlacionado apenas à qualidade do produto, mas também outros fatores como experiências sociais e evolução do produto em relação suas próprias versões anteriores.

O *Design* foi o aspecto com o menor grau de positividade dentre as *features*. A trajetória dos modelos de *smartphones* está em linha com esse resultado, já que desde 2007 os modelos dos celulares apresentaram nenhuma disruptão no visual comparados ao primeiro Iphone. Como já são mais de 10 anos sem mudanças expressivas no *Design*, é natural que o sentimento em relação a esse aspecto aproxime-se do neutro.

Quando comparados os resultados de Iphone 13 e Galaxy S21, é visível como os resultados de iphone estão mais sólidos e representativos, consequência do volume superior de *tweets* encontrados citando o celular da Apple perante o celular da Samsung. Inclusive, o resultado da mineração revela a popularidade superior do Iphone 13 sobre o Galaxy S21 em países com a língua inglesa, sobretudo os Estados Unidos.

Por fim, verificou-se que o número de *tweets* minerados é expressivamente superior ao número de associações Aspectos-Adjetivos. Essa diferença nos volumes entre as 2 etapas é reflexo da característica do Twitter, onde o conteúdo não possui um tema, propósito e formato definido, provocando um número elevado de textos sem informações significativas para o presente trabalho. A aplicação desse código em outros ambientes, como sites da Amazon, Reclame aqui, ou qualquer outra plataforma que possui campos específicos para avaliações dos produtos, pode apresentar resultados valiosos. A restrição da API do Twitter em *tweets* de até 7 dias passados juntamente com o período habitual para o desenvolvimento de um trabalho de conclusão de curso limitou a relevância da base de dados criada. Essa limitação impactou principalmente as conclusões sobre o Galaxy S21 devido ao menor volume encontrado, sendo que uma base com maior volume sobre esse aparelho permitiria conclusões mais embasadas e assertivas. Sendo assim, recomenda-se a realização de uma mineração por um período maior de forma a construir uma base de dados mais extensa e representativa.

8. REFERÊNCIAS

- AHLGREN, MATT. **Mais de 50 estatísticas e fatos do Twitter.** Disponível em <https://www.websiterating.com/pt/research/twitter-statistics/>.
- AKUNDI, A. et al. **Text Mining to Understand the Influence of Social Media Applications on Smartphone Supply Chain**, 2018.
- BING, LIU. **Sentiment Analysis and Opinion Mining**, Morgan & Claypool Publishers, 2012.
- CAROSIA, A.; COELHO, G.; SILVA, A. **Investment strategies applied to the Brazilian stock market: A methodology based on Sentiment Analysis with deep learning**, 2021.
- CHAE, B. **Insights from hashtag# supplychain and Twitter Analytics: Considering Twitter and Twitter data for supply chain practice and research**, 2015.
- DAY, G.; STAELIN, R. **The Path to Customer Centricity**, 2006.
- DESARBO, W.; JEDIDI, K; SINHA, I. **Customer Value Analysis in a heterogeneous market**, 2001.
- EVANS, G. **Measuring and managing customer value**, 2002.
- FERREIRA, E. **Análise de Sentimento em Redes Sociais Utilizando Influência das Palavras**, 2010.
- GRAF, A.; MAAS, P. **Customer value from a customer perspective: a comprehensive review**, 2008.
- HUANG, S.; POTTER, A.; EYERS, D. **Social Media in Operations and Supply Chain Management: State-of-the-Art and Research Direction**, 2019.
- KIRITCHENKO, S. et al. **Detecting aspects and sentiment in customer reviews**, 2014.
- KOTSIANTIS, S. **Supervised Machine Learning: A Review of Classification Techniques**, 2007.
- LORIA, STEVEN. **Textblob Documentation**. Release 0.15, v. 2, p. 269, 2018.
- MARNEFFE, M; MANNING, C. **Stanford typed dependencies manual**, 2016.

NOH, Y.; PARK, S.; PARK, SEONG-BAE. **Aspect-Based Sentiment Analysis Using Aspect Map**, 2019.

PANDITA, H.; GONDHI, N. **A literature survey of sentiment analysis based on E-commerce reviews**, 2021.

PENG, H. et al. **Knowing What, How and Why: A Near Complete Solution for Aspect-Based Sentiment Analysis**, 2020.

SMITH, A.; ANDERSON, M. **Social Media Use in 2018**, 2018.

STINE, R. **Annual Review of Statistics and Its Application Sentiment Analysis**, 2018.

WALIA, CHAUDHARY. **Mobile Devices Monitor – Q3 2021**. Disponível em https://report.counterpointresearch.com/posts/report_view/Monitor/2564.

WE ARE SOCIAL. **Digital 2021 October Global Statshot Report**. Disponível em <https://wearesocial.com/uk/blog/2021/10/social-media-users-pass-the-4-5-billion-mark/>.

Acesso em: 08 nov. 2021.