

UNIVERSIDADE DE SÃO PAULO  
ESCOLA DE ENGENHARIA DE SÃO CARLOS

JOÃO VITOR DE MEDEIROS

Dados de redes sociais para o desenvolvimento de novos produtos: um estudo  
de caso com *smartwatches*

Monografia apresentada ao Curso de  
Engenharia de Produção, da Escola de  
Engenharia de São Carlos da Universidade de  
São Paulo, como parte dos requisitos para  
obtenção do título de Engenheiro de Produção.

Orientador: Prof. Dr. Janaína Hornos Mascarenhas da Costa

Co-orientador: Dra. Carina Campese

São Carlos

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).

M488d      Medeiros, João Vitor  
              Dados de redes sociais para o desenvolvimento de  
              novos produtos: um estudo de caso com smartwatches /  
              João Vitor Medeiros; orientadora Janaina Hornos  
              Mascarenhas da Costa; coorientadora Carina Campese.  
              São Carlos, 2019.

              Monografia (Graduação em Engenharia de Produção  
              Mecânica) -- Escola de Engenharia de São Carlos da  
              Universidade de São Paulo, 2019.

              1. desenvolvimento de novos produtos. 2. análise  
              de mídias sociais. 3. extração de dados de redes  
              sociais. 4. desenvolvimento de produtos. I. Título.

Eduardo Graziosi Silva - CRB - 8/8907

## FOLHA DE APROVAÇÃO

<b>Candidato:</b> João Vitor de Medeiros
<b>Título do TCC:</b> Uso de dados de mídias sociais no Desenvolvimento de produtos: um estudo de caso com <i>Smartwatches</i>
<b>Data de defesa:</b> 14/10/2019

Comissão Julgadora	Resultado
Professor Doutor Janaina Mascarenhas Hornos da Costa (orientador)	Aprovado
Instituição: EESC - SEP	
Professor Doutor Humberto Filipe de Andrade Januário Bettini	Aprovado
Instituição: EESC - SEP	
Pesquisador Carina Campese	Aprovado
Instituição: EESC - SEP	

Presidente da Banca: **Professor Doutor Janaina Mascarenhas Hornos da Costa**



## AGRADECIMENTOS

A Profa. Dra. Janaína Costa e Dra. Carina Campese pelo imenso auxílio neste trabalho, apoio durante a graduação e projetos de Iniciação científica.

A minha família que proporcionou suporte e apoio incondicional para que pudesse realizar minha graduação em Engenharia de Produção.

Ao time Enactus USP-São Carlos onde tive a oportunidade de trabalhar com pessoas incríveis e trazer impacto social para a comunidade de São Carlos.

A todos os atuais e ex-moradores da República Tira Gosto, os quais me ofereceram grande apoio profissional e pessoal. Levarei estes ensinamentos por toda a minha vida, assim como todas as experiências trocadas.

A Giulia Di Giacomo e Caio Arbulu que apoiaram, incentivaram e colaboraram com este projeto.

A toda comunidade do Github, Stackoverflow e Pushishift.io por disponibilizarem informações e conteúdos gratuitos, e por serem sempre solícitos com novos programadores.



## RESUMO

MEDEIROS, J. V. **Dados de redes sociais para o desenvolvimento de novos produtos:** um estudo de caso com *smartwatches*. 2019. Monografia (Trabalho de Conclusão de Curso) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2019.

O âmbito da inovação colaborativa e envolvimento do usuário no desenvolvimento de novos produtos vêm sendo amplamente estudado e explorado nos últimos anos, e seu potencial de gerar resultados positivos para o desenvolvimento de produtos é reconhecido, como no aumento da aceitação de mercado dos produtos. O surgimento de novas tecnologias de informação e comunicação provocou uma mudança na relação empresa-consumidor e abriu uma nova gama de possibilidades, uma grande quantidade de dados públicos passou a ser gerada, principalmente com o crescimento da popularidade de mídias sociais. Porém a utilização de mídias sociais para o desenvolvimento de novos produtos ainda é escassa e pouco explorada. Portanto, esta pesquisa tem como objetivo explorar a utilização da análise de mídias sociais para o desenvolvimento de novos produtos, e comparar o uso de diferentes redes sociais como contribuição para equipes de desenvolvimento utilizando um estudo de caso com a categoria de produtos *smartwatches* e as redes sociais Twitter e Reddit. Com este estudo conclui-se que o Twitter não é uma plataforma recomendada para utilização de dados como forma de *input* para a fase de ideação do processo de desenvolvimento, mas pode ser utilizado na fase de teste/suporte do produto. Por outro lado, a plataforma Reddit se mostrou com alto potencial para ser utilizada na fase inicial do desenvolvimento de produtos, provendo *insights* para a equipe de desenvolvimento.

Palavras-chave: Desenvolvimento de novos produtos. Análise de mídias sociais. Extração de dados de redes sociais.





## ABSTRACT

MEDEIROS, J. V. **Social media for new product development:** a study case with smartwatches. 2019. Monografia (Trabalho de Conclusão de Curso) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2019.

The field of collaborative innovation and user involvement in the new product development process is being largely studied and explored in the past few years, and its potential to generate positive results is already recognized, like the increase of products market fit. The emerging of new information and communication technologies caused a change in the producer-consumer relationship and opened new possibilities. A large amount of public data is generated each day, mainly due to social media popularity growth and availability. However, the use of social media data in new product development still small and poorly explored. Therefore, through a study case with smartwatches, the objective of this research is to explore the potential use of social media analytics in new product development and compare how different social networks can provide insights development teams. It's concluded that Twitter is not a recommended platform to be used as a data source to generate inputs in the ideation phase of the product development process. In the other side, the social media Reddit shows a great potential to be used in the initial phase of the process and generate insights to the development team.

Keywords: New product development. Social media analytics. Social network data extraction.



## SUMÁRIO

1. INTRODUÇÃO.....	13
2. REVISÃO BIBLIOGRÁFICA .....	15
2.1 Envolvimento do Usuário no processo de Desenvolvimento de Novos Produtos .....	15
2.2 Mídias Sociais e redes sociais .....	17
2.2.1 Definições .....	17
2.2.2 Análise de Mídias Sociais.....	19
2.3 Twitter e Reddit como fonte de dados para pesquisa.....	21
3. METODOLOGIA.....	23
3.1 Entendimento Inicial da Pesquisa (F1).....	24
3.2 Fundamentação teórica (F2).....	24
3.3 Estudo de caso (F3) .....	26
3.3.1 Preparação e pré-processamento (M3) .....	26
3.3.2 Mineração dos dados (M4) .....	32
3.3.3 Análise dos dados (F4) .....	35
4. RESULTADOS .....	36
4.1 Extração de dados e pré-processamento.....	36
4.1.1 Twitter .....	36
4.1.2 Reddit.....	36
4.2 Análise de share of voice .....	37
4.3 Análise de Frequência .....	40
4.4 Análise de sentimentos.....	43
5. CONCLUSÃO.....	47
6. BIBLIOGRAFIA .....	49
APÊNDICE A – Análises de sentimentos da plataforma Twitter .....	51
APÊNDICE B - Análises de sentimentos da plataforma Reddit .....	52



# 1. INTRODUÇÃO

Com o aumento da disponibilidade e popularidade da internet, uma quantidade significativa de dados vem sendo gerada através de mídias sociais. Muitas ferramentas de pesquisa para mídias sociais foram desenvolvidas, e muitas empresas desenvolveram programas de relacionamento com cliente através de mídias sociais. Porém a exploração de mídias sociais para pesquisas de mercado voltadas para a inovação no desenvolvimento de produtos ainda é pouco utilizada. Existe uma preocupação quanto à confiabilidade dos dados, e em encontrar uma forma de gerar *insights* relevantes (JEAANMARIE CARR, LUCIEN DECRETON, WEI QIN, 2013).

A área de estudo da gestão da inovação vem recebendo cada vez mais ênfase, principalmente quando falamos dos modelos de inovação aberta (*open innovation*) e inovação colaborativa (*collaborative innovation*). Estes conceitos enfatizam a importância de recolher e fazer um uso eficiente do conhecimento adquirido de fontes externas à organização, como fornecedores, outras organizações, universidades, comunidades, e principalmente dos clientes da empresa.

Os benefícios e incentivos para a inovação aberta e a inclusão dos clientes no desenvolvimento de produtos são apresentados por diversos autores. Connor (1999) argumenta que a informação advinda do mercado e dos clientes é importante para diversos tipos de inovação. O'Connor, (1998) salienta que os usuários geram *inputs* importantes para inovações incrementais, e Eisenhart (2019) mostra que o envolvimento de clientes na inovação aberta pode melhorar aspectos como a efetividade do conceito do produto, e a aceitação de mercado. Já Martini et. al (2014) afirmam que utilizar clientes no processo de Desenvolvimento de Novos Produtos, *New Product Development (NPD)*, torna possível adquirir ideias incrementais e também radicais concomitantemente.

A interação e o envolvimento de clientes no NPD podem ocorrer em diferentes *touchpoints* (pontos de contato). Berendsen et al. (2015) define *touchpoints* como os diferentes pontos de interação entre clientes e uma empresa, onde o cliente pode possuir diversos papéis nessas interações, dependendo por exemplo, em qual fase do processo está sendo envolvido (NAMBISAN, 2002).

Esta pesquisa tem como objetivo primário comparar os dados que podem ser obtidos de duas redes sociais distintas relacionados ao produto *smartwatch*, e que poderiam ser

utilizados como um recurso para geração de *insights* para a fase de geração de ideias do processo de Desenvolvimento de Novos Produtos. Assim, o objetivo secundário desta pesquisa é realizar a extração e análise de dados das redes sociais Twitter e Reddit.

## 2. REVISÃO BIBLIOGRÁFICA

### 2.1 Envolvimento do Usuário no processo de Desenvolvimento de Novos Produtos

O surgimento de novas tecnologias de informação e comunicação iniciou uma transformação radical na relação produtor-consumidor em muitas indústrias, com implicações importantes para o NPD (*New Product Development*). Novas tecnologias, como a internet, permitem o surgimento de vastas comunidades de interesse se formem ao redor de produtos e serviços específicos (NAMBISAN, 2002, p.01). Nambisan (2002) também afirma que estas chamadas Comunidades Virtuais de Clientes, ou *Virtual Customer Communities* (VCC) podem facilitar o desenvolvimento de modelos de inovação que envolvem vários papéis do consumidor no NPD.

O desenvolvimento de novos produtos é uma tarefa desafiadora, equipes de NPD estão cada vez mais procurando por recursos externos para superar a curva de aprendizado relacionada a novas tecnologias e novos mercados (NAMBISAN, 2002). O potencial de clientes como um recurso externo para o desenvolvimento de novos produtos é reconhecido há muito tempo na teoria e na prática. Por exemplo, o envolvimento do cliente se mostrou capaz de melhorar a efetividade do conceito de um produto, ou seja, a aceitação do mercado (*Market fit*) (EISENHARDT, 1995).

Na literatura de gestão estratégica e gestão da qualidade, pesquisadores identificaram cinco diferentes papéis dos clientes na criação de valor: recurso, coprodutor, comprador, usuário e produtor (NAMBISAN, 2002 *apud* FINCH, 1999; GERSUNY & ROSENGR; KAULIO, 1998; LENGNICK-HALL). Os primeiros dois papéis estão focados na fase de *input* da atividade organizacional, por outro lado os outros três papéis estão localizados nas fases de *output* do sistema. Dentro do contexto de NPD, apenas os primeiros três papéis são relevantes (usuário como recurso, coprodutor e usuário). O primeiro papel, refere-se ao cliente como fonte para a inovação, o segundo, como participante do design e desenvolvimento do produto, e o terceiro como o envolvimento do cliente no teste e suporte do produto. Os demais papéis são menos relevantes para o panorama de desenvolvimento de novos produtos, ao passo que enfatizam o cliente mais como objetos do que agentes da criação de valor. Os principais papéis do cliente no NPD, suas fases de envolvimento, assim como os principais problemas e desafios desse envolvimento são apresentados a seguir na Figura 1.

Figura 1 – Papeis do cliente no NPD.

Papel do cliente	Fase do NPD	Problemas e desafios
Cliente como um recurso	Ideação	<ul style="list-style-type: none"> <li>• Adequação do cliente como fonte de inovação</li> <li>• Seleção de cliente inovador</li> <li>• Necessidade de incentivos variados para o cliente</li> <li>• Infraestrutura para capturar o conhecimento do cliente</li> <li>• Papéis diferentes de clientes atuais e potenciais</li> </ul>
Cliente como cocriador	Design e desenvolvimento	<ul style="list-style-type: none"> <li>• Envolvimento em uma vasto espectro de atividades de design e desenvolvimento</li> <li>• Natureza do contexto de NPD: industrial/bens de consume</li> <li>• Aproximação com os times internos do NPD</li> <li>• Gerenciar as incertezas do projeto</li> <li>• Aumentar o conhecimento do cliente quanto ao produto ou tecnologia</li> </ul>
Cliente como usuário	Teste de produto Suporte do produto	<ul style="list-style-type: none"> <li>• Limitação de tempo</li> <li>• Garantir diversidade de clientes</li> <li>• Atividade recorrente</li> <li>• Infraestrutura para suportar as relações com os clientes</li> </ul>

Fonte: Adaptado de Nambisan, 2002

O papel do usuário mais comentado na literatura é do consumidor como fornecedor de informações (LEGNICK-HALL, 1996). O cliente como fonte de ideias para novos produtos é o tipo de envolvimento utilizado nesta pesquisa, uma vez que este tema foi bem explorado na literatura (e.g. LEONARD, 1995; URBAN; HIPPEL, 1989).

A utilidade do cliente como um recurso varia principalmente com a maturidade da tecnologia e o alinhamento da linha de produtos com a base de clientes atual (CHRISTENSEN, 1995; LEONARD, 2015). Quando ambas as dimensões estão em um patamar alto, os clientes são um ótimo recurso para inovação, por outro lado, quando ambas as dimensões são baixas (como em tecnologias em desenvolvimento, e mercados emergentes), o valor de clientes como um recurso é limitado. Mesmo nestes contextos, os clientes podem ter um papel passivo importante, empresas aplicam uma variedade de pesquisas estruturadas (pesquisas de mercado, grupos de foco) para acessar o conhecimento dos clientes. Clientes raramente fornecem ideias para novos produtos sem que sejam solicitadas por organizações, além disso, estas pesquisas estruturadas limitam a riqueza e frequência das contribuições dos clientes. Questões logísticas e econômicas forçam firmas a envolver uma minoria de clientes, frequentemente não representativos da diversa população de clientes (NAMBISAN, 2002 *apud* WAYLAND; COLE, 1997).

As empresas encontram três principais desafios para envolver usuários como fontes de novas ideias de produtos, o primeiro é relacionado à seleção de clientes inovadores, e o



estabelecimento de laços com eles. Principalmente nos casos de bens de consumo, e produtos industriais, existe uma dificuldade em evolver o usuário de uma maneira efetiva em termos de custo. O segundo desafio consiste em criar o incentivo correto para que os clientes contribuam com novas ideias. E o terceiro desafio está relacionado à dificuldade em capturar o conhecimento do cliente. Nambisan (2002) afirma que é mais efetivo entender clientes e seus requisitos no seu ambiente natural, do que em ambientes artificiais.

## **2.2 Mídias Sociais e redes sociais**

### **2.2.1 Definições**

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

Para uma definição formal do conceito de mídia social, é necessário delimitar uma linha entre outros conceitos relacionados: Web 2.0 e conteúdo gerado pelo usuário. O termo Web 2.0 foi primeiramente utilizado em 2004 para descrever a nova forma como desenvolvedores e usuários passaram a utilizar a World Wide Web, que se resume em uma plataforma onde o conteúdo não é mais somente criado e publicado por usuários individuais, mas criado e modificado por todos os usuários que participam de maneira colaborativa (KAPLAN; HAENLEIN, 2009). Páginas individuais e de sentido único de criação e publicação ficaram para trás, e foram substituídas por blogs, wikis e projetos colaborativos. Apesar da Web 2.0 não representar uma evolução técnica/tecnológica específica, um conjunto de funcionalidades básicas foi necessário para seu funcionamento, como o Adobe Flash e o RSS (*Really Simple Syndication*). Assim, a Web 2.0 pode ser considerada a plataforma utilizada para a evolução das mídias sociais (KAPLAN; HAENLEIN, 2010, p.03)

Em 1979, Tom Truscott e Jim Ellis, da Universidade de Duke, fundaram o Usenet, uma rede mundial de discussões que permitia aos usuários o compartilhamento de mensagens públicas. No entanto, a era das mídias sociais teve seu início 20 anos antes, quando Bruce e Susan Abelson fundaram o “Open Diary”, uma rede social que reunia escritores de diários online em uma única comunidade. O termo “weblog” foi utilizado pela primeira vez nesta mesma época, posteriormente reduzido apenas para “blog”.

O aumento da velocidade da internet e a popularidade do conceito proporcionaram o rápido crescimento destas plataformas, levando à criação de sites de redes sociais como

MySpace (2003) e Facebook (2004). Foi nesta época que o termo “mídia social” foi cunhado. Uma das inclusões mais recentes neste conceito foram os chamados “Mundos virtuais”, ambientes virtuais simulados e habitados por avatares tridimensionais, o seu representante mais conhecido é o Second Life (KAPLAN; HAENLEIN, 2009).

Enquanto a Web 2.0 é considerada a fundação ideológica e tecnológica, o Conteúdo gerado pelo usuário, ou *User Generated Content* (UGC), pode ser visto como a soma de todas as formas pelas quais as pessoas utilizam as mídias sociais. O termo, que se popularizou em 2005, e é geralmente utilizado para descrever conteúdos públicos que foram criados por usuários finais. Segundo a Organização para Cooperação e Desenvolvimento Econômico (OECD), existem três requisitos que caracterizam o UGC (SCIENCE; INFORMATION; POLICY, 2007):

- i) Conteúdo público e disponível na Internet
- ii) Que reflete uma “certa quantidade de esforço criativo”
- iii) Que é “criado fora de rotinas e práticas profissionais”

O primeiro requisito exclui mensagens privadas trocadas entre usuários, como e-mails e mensagens de texto. O segundo requisito exclui cópias de conteúdos já existentes, e o terceiro exclui todos os conteúdos gerados dentro de um contexto profissional ou voltado para o mercado (KAPLAN; HAENLEIN, 2010). Enquanto o UGC existe desde antes da Web 2.0, a combinação de incentivos tecnológicos (como o aumento da velocidade da internet), econômicos (como o aumento da disponibilidade de ferramentas gratuitas para criação de UGC), e sociais (como o surgimento de novas gerações que são “digitais nativas”), tornou este conceito muito diferente do que era observado em 1980.

Com uma melhor clarificação dos conceitos de Web 2.0 e UGC, é possível dar uma descrição mais detalhada sobre o significado de mídias sociais. Mídia social é um grupo de aplicativos da internet que formam as fundações tecnológicas e ideológicas da Web 2.0, e possibilita a criação e troca de conteúdo gerado pelo usuário (KAPLAN; HAENLEIN, 2010). Dentro desta definição geral, existem diversos tipos de Mídias Sociais que podem ser distinguidas entre si, como blogs, projetos colaborativos (Wikipedia), redes sociais (Facebook, Twitter), comunidades de conteúdo (Youtube), mundos virtuais sociais (Second Life), e mundos virtuais de jogos (World of Warcraft).

Para fins desta pesquisa, trataremos apenas com a exploração de conteúdos de Redes Sociais. Redes sociais são plataformas que permitem que usuários criem perfis pessoais, convidem colegas a acessar seus perfis, e enviar e-mails ou mensagens privadas uns aos outros, estes perfis podem conter todo tipo informação, como fotos, vídeos, áudios, documentos e blogs (KAPLAN; HAENLEIN, 2010).

De acordo com o WeAreSocial (2019), 57% (4,38 bilhões) da população utilizam a internet ativamente, sendo que aproximadamente 3,5 bilhões de pessoas (45% da população mundial) são usuários ativos de mídias sociais, e 3,26 bilhões utilizam mídias sociais através de aparelhos celulares ou tablets. Nos Estados Unidos, especificamente, cerca de 95% da população (312,3 milhões) utilizam a internet, e 70% (230 milhões) são usuários ativos de mídias sociais, com uma utilização média diária de 2,07 horas, onde 86% destes usuários utilizam mídias sociais todos os dias.

A expansão de mídias sociais causou uma mudança de paradigma em como as pessoas se comunicam, colaboram, criam, e consomem informação. Além disso, apesar da desigualdade social persistir, a curadoria e difusão de informação pública não é mais tão facilmente controlada por um pequeno número de instituições (ZENG et al., 2010).

Além do uso pessoal no dia a dia, principal responsável pela sua adoção em massa, as mídias sociais estão sendo utilizadas cada vez mais como canais de comunicação em contextos políticos e contextos de negócios. Por exemplo, empresas começaram a utilizar plataformas de mídias sociais internas e externas para inúmeros propósitos. Enquanto o uso de mídias sociais internas pode melhorar a comunicação e colaboração entre funcionários, gestão do conhecimento, e inovação de produtos/serviços; empresas também passaram a formar redes em mídias sociais com parceiros de negócios, e também a engajar em mídias sociais públicas com propósitos de marketing, relações públicas, relação com o cliente, gestão de reputação, e recrutamento (STIEGLITZ et al., 2014).

### **2.2.2 Análise de Mídias Sociais**

Pesquisas acadêmicas de várias disciplinas recentemente passaram a dar mais atenção às mídias sociais. Esse interesse recente na chamada “*Big Social Data*” é motivado principalmente pelo acesso facilitado a grandes bases de dados de mídias sociais populares, como Twitter, Facebook, e LinkedIn, assim como de outras plataformas colaborativas como *wikis*, e blogs (HALPERIN et al., 2012; STIEGLITZ et al., 2014).

Do ponto de vista acadêmico, as mídias sociais podem ser vistas como um “laboratório vivo”, que possibilita que pesquisadores coletem uma grande quantidade de dados gerados no mundo real. Existe um grande interesse em analisar a *Big Social Data* das mídias sociais não só no meio acadêmico, mas também para propósitos práticos. Por exemplo, analisando mídias sociais, empresas podem encontrar oportunidades para veicular anúncios, estabelecer relacionamento com o cliente (CRM), entre outros. Em particular, o principal interesse por trás de atividades corporativas em mídias sociais, é como utilizar a plataforma como um canal adicional de marketing. Além disso, empresas B2B (*Business to business*) começaram a utilizar mídias sociais como uma forma de encontrar potenciais clientes.

Do ponto de vista político, instituições também passaram a demonstrar interesse em monitorar a opinião pública sobre políticos e posições políticas, detectando e monitorando tópicos políticos, e gerenciando sua própria reputação em mídias sociais (STIEGLITZ et al., 2014).

Funcionários públicos também poderiam utilizar mídias sociais para identificar informações em casos de desastres naturais (BRUNS; LIANG, 2014; JEAN; ED, 2012). Baseado em dados de mídias sociais, organizações de saúde podem estabelecer um sistema de alerta para surtos de doenças que pode ajudar a prover medidas de resposta mais rapidamente. Indivíduos e consumidores também buscam fazer uso de informações e opiniões de diversas fontes de forma a tomarem decisões mais embasadas (STIEGLITZ et al., 2014).

No âmbito de desenvolvimento de produtos, existem estudos de casos em diversos mercados onde o comportamento e publicações dos usuários foram utilizados por equipes de desenvolvimento. J. Carr et. al (2013) realizou um estudo relacionado à “frescor do café”, já Sandor et. al (2013) fez uma pesquisa sobre novas tendências do mercado de cosméticos,. As maiores dificuldades apontadas nestes estudos foram a confiabilidade dos dados, e como utilizar estes dados para gerar *insights* relevantes.

O interesse em *Big Social Data* e análise de mídias sociais possui dois principais motivadores. O primeiro, é o avanço tecnológico que permite o monitoramento contínuo, automatizado, e a análise dos conteúdos e interações de mídias sociais. O segundo motivo é a mudança da participação pública, que levou ao aumento da complexidade do ambiente de comunicação, como por exemplo o aumento da heterogeneidade dos usuários que estão se comunicando, maior difusão da informação com respeito a escala e velocidade, e também o rápido desenvolvimento e adoção de dispositivos móveis (STIEGLITZ et al., 2014).

Estudos e pesquisas recentes revelam a necessidade de coletar, monitorar, analisar, sumarizar e visualizar continuamente informações de interações sociais e conteúdos gerados pelos usuários em vários contextos, desde negócios e administração pública até política e decisões do consumidor (STIEGLITZ; BROCKMANN; DANG-XUAN, 2012; ZENG et al., 2010).

Tais tarefas são consideradas difíceis devido ao grande número de plataformas de mídias sociais, assim como a grande quantidade, dinamicidade, e complexidade dos dados advindos de mídias sociais. De forma mais específica, as mídias sociais geram uma grande quantidade de dados e meta-dados, que ainda estão sendo explorados pela literatura de mineração de textos e dados (ZENG et al., 2010). Como exemplos destes dados, temos opiniões subjetivas, emoções, avaliações, que podem se manifestar na forma de imagens, vídeos, *tags*, e outros tipos de dados, como “*likes*”, comentários, *retweets*, menções, entre outros.

Nos últimos anos, uma nova área de pesquisa interdisciplinar chamada *Social Media Analytics* (SMA), ou Análise de Mídias Sociais, vem se estabelecendo para tratar justamente destas dificuldades. Seu principal objetivo é desenvolver e avaliar métodos científicos, assim como modelos técnicos e ferramentas de software para rastrear, modelar, analisar e minerar grandes quantidades de dados de mídias sociais para diversos propósitos. No contexto de negócios, a SMA pode ser considerada uma subdivisão da área de *Business Inteligente* (BI), que trata de metodologias, processos, arquiteturas e tecnologias que transformam dados puros de mídias sociais em informações úteis e significativas para o propósito do negócio.

### **2.3 Twitter e Reddit como fonte de dados para pesquisa**

Dentre as redes sociais existentes, o Twitter é uma das plataformas de *microblogging* mais famosa, e vem provendo um impulso para a utilização de *microblogging* como uma ferramenta de pesquisa com diversos objetivos (PLATANIA; SPADONI, 2018).

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, porém não limita a variedade de tópicos, nem os tipos de relações entre os usuários (POCHAMPALLY, 2011). Diversos autores conduziram pesquisas envolvendo análises baseadas no Twitter. Foram publicados estudos com foco em problemas de pesquisas que poderiam ocorrer ao utilizar o Twitter como fonte de informação (STIEGLITZ et al., 2014); análises de *supply chain* e o papel do Twitter nas relações entre diferentes atores (CHAE,

2015); investigações sobre o papel que o twitter desempenha em estratégias de negócio e marketing (BARRY et al., 2016).

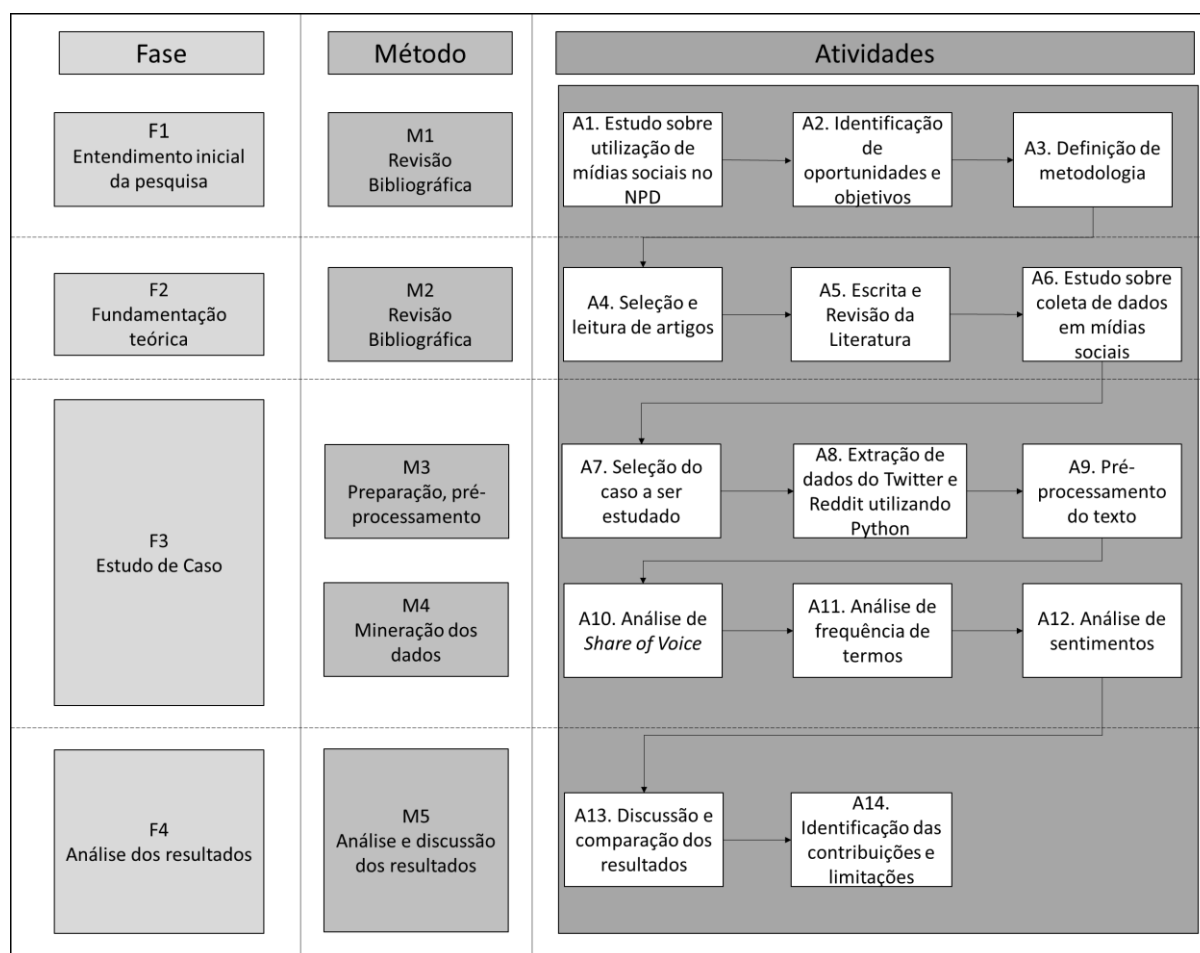
Por outro lado, o Reddit, é uma das redes sociais mais populares de fóruns de discussões. O Reddit é constituído por *subreddits*, que são subfóruns direcionados para a discussão de determinado assunto. Diferente do Twitter, as submissões do Reddit são limitadas a 40.000 caracteres, tornando as publicações mais ricas em informações.

Assim, o uso do Reddit como fonte de dados para o desenvolvimento de produtos, é mais utilizado quando se busca informações mais completas e profundas. Assim, o conteúdo gerado pelo Reddit são *leading indicators*, que podem ser utilizados para identificar novas tendências de uso e necessidades dos clientes, em contraste com o Twitter que possui um carácter mais pessoal e pontual (SANDOR et al., 2018).

### 3. METODOLOGIA

A metodologia adotada para esta pesquisa é dividida em quatro principais fases, apresentadas na Figura 2. A primeira fase é denominada “entendimento inicial da pesquisa”, buscando compreender os tópicos e panorama geral das áreas envolvidas na pesquisa. A segunda fase, denominada “aprofundamento teórico”, trata-se do aprofundamento nos temas estudados na primeira fase, com o intuito de embasar e direcionar a pesquisa. A terceira fase constitui um estudo de caso, que compreende a coleta e análise de dados do Twitter e Reddit relacionados à categoria de produtos *smartwatch*. Buscando demonstrar e testar a utilização de mídias sociais para a geração de ideias para o desenvolvimento de produtos, assim como comparar os dados obtidos das diferentes redes sociais. As atividades realizadas em cada uma das fases estão descritas a seguir.

Figura 2 – Metodologia de pesquisa com fases, métodos e atividades



Fonte: elaborado pelo autor

### **3.1 Entendimento Inicial da Pesquisa (F1)**

Esta fase teve como objetivo a compreensão geral do tema a ser estudado. Foram levantados artigos relacionados à utilização de redes sociais em desenvolvimento de novos produtos. Tal pesquisa, além de proporcionar um embasamento no tema, permitiu a identificação de lacunas e oportunidades para esta pesquisa, também possibilitando a definição de metodologia e objetivos (A1, A2 e A3).

### **3.2 Fundamentação teórica (F2)**

A segunda fase consistiu na busca, seleção e leitura de artigos científicos extraídos da base de dados *Science Direct*. Foram buscados artigos com as seguintes palavras-chave: “*social media analytics*”, “*product development*”, “*consumer research*”, “*market research*”.

A busca foi realizada apenas em inglês, e os artigos foram selecionados conforme a relevância destes quanto a citações, e com base em uma leitura prévia de seus resumos. Ao todo, a pesquisa retornou aproximadamente 14.000 resultados. Através de uma avaliação inicial de títulos e número de citações, 373 artigos foram escolhidos para leitura dos resumos, dos quais, 22 foram selecionados para leitura completa. Com a leitura destes artigos, foram encontrados outros 20 artigos referenciados que também foram separados para estudo. Ao todo foram estudados 42 artigos relacionados aos temas (A4), o Quadro 1 apresentado a seguir, apresenta os artigos selecionados para leitura.



Quadro 1 – Artigos selecionados para leitura na fase de Fundamentação teórica (F2)

Título	Autor	Ano
Social media, new digital technologies and their potential application in sensory and consumer research	Olsen, Nina Veflen; Christensen, Kasper	2015
Social media's emerging importance in market research	Patino, Anthony; Pitta, Dennis A.; Quinones, Ralph	2012
Using Twitter data for food-related consumer research: A case study on "what people say when tweeting about different eating situations"	Vidal, Leticia; Ares, Gastón; Machín, Leandro; Jaeger, Sara R.	2015
Opportunities for Innovation in Social Media Analytics	Moe, Wendy W.; Schweidel, David A.	2017
Social media analytics for new product development: Case of a pizza	Rathore, Ashish Kumar; Ilavarasan, P. Vigneswara	2017
Social media content and product co-creation: an emerging paradigm	Rathore, Ashish K.; Ilavarasan, P. Vigneswara; Dwivedi, Yogesh K.	2016
Opportunities for Innovation in Social Media Analytics	Moe, Wendy W.; Schweidel, David A.	2017
Influence of the adoption and use of social media tools on absorptive capacity in new product development	Peltola, Tero; Mäkinen, Saku J.	2014
Social media infrastructure: supporting communication practices from behind the scenes	Palmer, Laura A	2013
Consumers' motivation to interact in virtual food communities – The importance of self-presentation and learning	Jacobsen, Lina Fogt; Tudoran, Ana Alina; Lähteenmäki, Liisa	2017
Social media, new digital technologies and their potential application in sensory and consumer research	Olsen, Nina Veflen; Christensen, Kasper	2015
Social media in product development	Sandor, Logan; Midura, Justis; Abedin, Sabreena; Ingber, George; Pederson, Maxwell; Sander, Tim; Wolf, Liam; Bailey, Reid; Flower, Abigail	2018
Social media in product development	Jeanmarie, Carr; Lucien, Decreton; Wei, Qin; Beatriz, Rojas	2013
Use of Social Media Applications for Supporting New Product Development Processes in Multinational Corporations	Bashir, Naheed; Papamichail, K. Nadia; Malik, Khaleel	2017
New Insights from Practice: Exploring Online Channel Management Strategies and the Use of Social Media as a Market Research Tool	Klaus, Philipp 'Phil'	2014
Social Media and New Product Development	Kenly, Amy	2013
Social media within sustainable product development: an exploratory multiple case study on the perception of social media usability in the new product development process	Berendsen, Gerard; Middel, Rick; Pieters, Ineke; Angard, Fabian; Hillerström, Fredrik	2015
Social media analytics – Challenges in topic discovery, data collection, and data preparation	Stieglitz, Stefan; Mirbabaie, Milad; Ross, Björn; Neuberger, Christoph	2018
How people share information about food: Insights from tweets regarding two Italian Regions	Platanía, M.; Spadoni, R.	2018
Social media in the product development process of the automotive industry: A new approach	Klein, Andreas; Spiegel, Götz	2013
Use of emoticon and emoji in tweets for food-related emotional expression	Vidal, Leticia; Ares, Gastón; Jaeger, Sara R.	2016
Mineração de Textos - technical report	Morais, Edison Andrade Martins; Ambrósio, Ana Paula L.	2007
Tools and methods for capturing Twitter data during natural disasters	Bruns, Axel; Liang, Yuxian Eugene	2014
Entifying your brand among Twitter-using millennials	Sashittal, Hemant C.; Hodis, Monica; Sriramachandramurthy, Rajendran	2015
Wellspring of Knowledge	Leonard, Dorothy A	2015
Historicizing new media: A content analysis of twitter	Humphreys, Lee; Gill, Phillipa; Krishnamurthy, Balachander; Newbury, Elizabeth	2013
The fairyland of Second Life: Virtual social worlds and how to use them	Kaplan, Andreas M.; Haenlein, Michael	2009
Usage of Social Media for Political Communication	Stieglitz, Stefan; Brockmann, Tobias; Dang-Xuan, L	2012
Working within a black box: Transparency in the collection and production of big twitter data	Driscoll, Kevin; Walker, Shawn	2014
Local and Global Responses to Disaster	Jean, E; Ed, Peter	2012
Campaign and corporate goals in conflict: Exploring company-issue congruence through a content analysis of Coca-Cola's twitter feed	Gaither, Barbara Miller; Austin, Lucinda	2016
Food policy in cyberspace: A webometric analysis of national food clusters in South Korea	Kim, Jang Hyun; Park, Han Woo	2014
Designing Virtual Customer Environments for New Product Development	Nambisan, Satish	2002
Customer Contributions to Quality : A Different View of the Customer-Oriented Firm	Lengnick-Hall, Cynthia A.	1996
Users of the world, unite! The challenges and opportunities of Social Media	Kaplan, Andreas M.; Haenlein, Michael	2010
The Innovator's Dilemma; Pages 1-14	Christensen, By Clayton M	1995
Socialmedia analytics	Stieglitz, Stefan; Dang-Xuan, Linh; Bruns, Axel; Neuberger, Christoph	2014
Social media analytics and intelligence	Zeng, Daniel; Chen, Hsinchun; Lusch, Robert; Li, Shu Hsing	2010
User context as a source of topic retrieval in Twitter	Pochampally, Ravali	2011
How to plan and perform a qualitative study using content analysis	Bengtsson, Mariette	2016
A hands-on guide to doing content analysis	Erlingsson, Christen; Brysiewicz, Petra	2017

Fonte: elaborado pelo autor

A revisão bibliográfica desta pesquisa foi fundamentada nestes artigos estudados, e a escolha das redes sociais Twitter e Reddit também foi feita com base nas pesquisas realizadas dos artigos revisados (A5).

Também foi realizada uma pesquisa preliminar em fóruns de programação, como o “Github” e “stackoverflow” para entender métodos de extração de dados. Esta pesquisa mostrou a facilidade na obtenção de dados das redes sociais escolhidas (A6).

### **3.3 Estudo de caso (F3)**

A fase de estudo de caso foi baseada no framework de análise de mídias sociais descrito por STIEGLITZ et al., 2018, o qual é subdividido nas seguintes etapas:

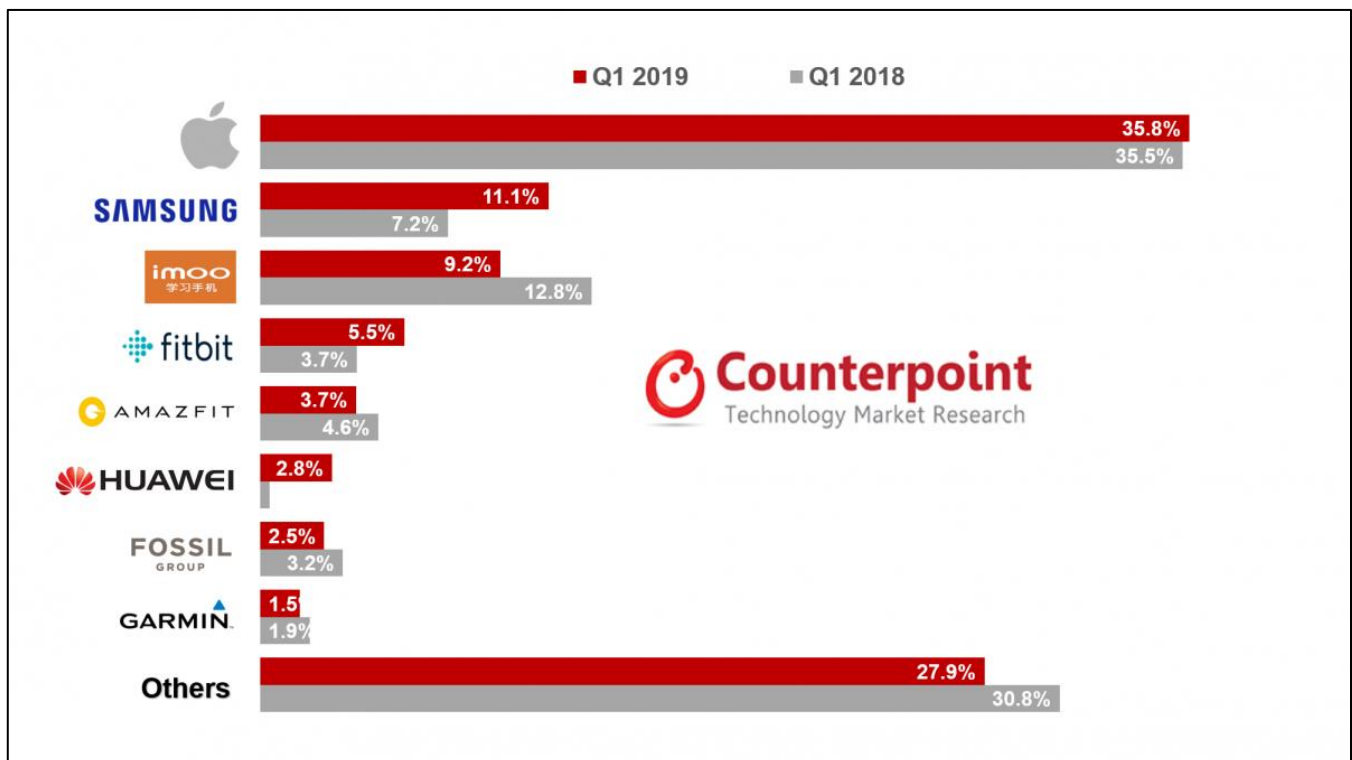
- *Descoberta (Discovery)*: esta etapa é utilizada para a identificação de quais tópicos serão extraídos. Mesmo em casos onde o tópico é bem conhecido, esta etapa pode auxiliar na identificação de palavras-chaves e *hashtags* relacionados ao tópico, que serão utilizadas na extração de dados (A7).
- *Extração ou rastreamento (Tracking)*: esta etapa envolve decisões quanto à fonte de dados (Twitter, Facebook, Reddit), abordagem, método, e saída de dados (A8).
- *Preparação (Preparation)*: esta etapa consiste na alocação e preparação dos dados extraídos para futura análise, como limpeza de textos, e remoção de dados duplicados (A9).
- *Análise (Analysis)*: fase de análise de dados, onde, dependendo do propósito, existem diversos métodos disponíveis, como Análise de sentimentos, análise de frequência, e análise de rede. Nesta etapa foram realizadas análises de *share of voice* (A10), análises de frequência (A11), e análise de sentimentos utilizando o método TextBlob (A12).

#### **3.3.1 Preparação e pré-processamento (M3)**

Tendo em vista o objetivo da pesquisa, de exemplificar a utilização de análise de mídias sociais para geração de ideias no desenvolvimento de novos produtos, e comparar dados de diferentes fontes, o tópico escolhido para ser explorado foi a categoria de produtos *Smartwatch*. Como apresentado no tópico F2, as mídias sociais escolhidas para serem comparadas foram o Twitter e Reddit.

Assim, com o intuito de melhor direcionar a fase de extração de dados, foi realizada uma pesquisa preliminar com o intuito de identificar os principais modelos de *Smartwatches* no mercado, maiores mercados consumidores, e períodos com maior volume de vendas (A7). A Figura 3 mostra os dados de *market-share* do mercado mundial de *Smartwatches*, no qual a fabricante Apple é o principal *player*, seguido pela Samsung, Imoo, Fitbit e AmazFit, respectivamente.

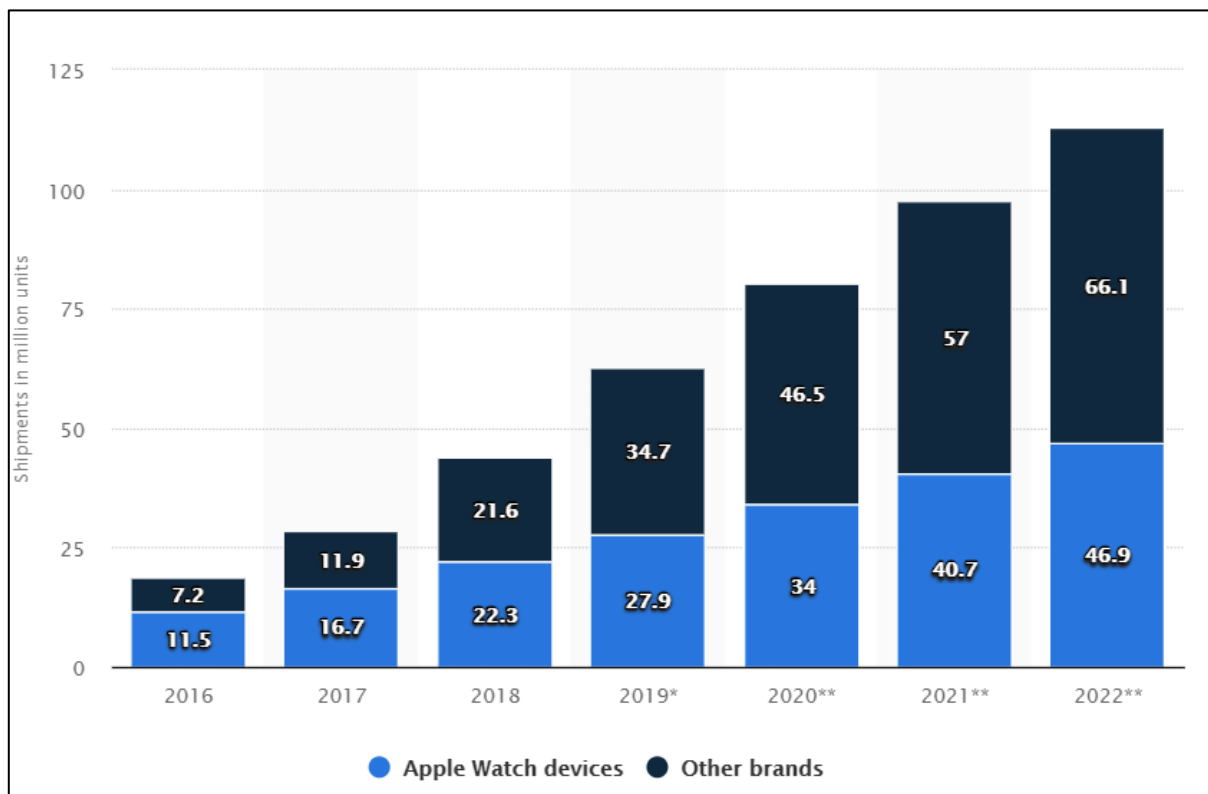
Figura 3 - *market-share* número de smartwatches vendidos por fabricante.



Fonte: Sinha, 2019.

Também é possível observar que o número de vendas de *smartwatches* é crescente desde sua inserção no mercado (2014), como mostrado na Figura 4. Este crescimento nas vendas leva ao aumento do número de usuários desta categoria de produtos. Assim, foram escolhidos períodos mais recentes para a extração de dados, onde existe uma maior quantidade de usuários, e existe um maior volume de conteúdos públicos relacionados a modelos mais recentes e, portanto, dados mais relevantes para uma pesquisa de mercado.

Figura 4 - Número de unidades de smartwatches vendidas globalmente por ano.



Fonte: Liu, 2019.

Em outra pesquisa prévia, foi utilizado o termo de pesquisa “Smartwatch” para identificar outros possíveis termos que usuários utilizam para se referir à esta categoria de produtos. Como resultado desta pesquisa preliminar, foi observado que o termo “Apple Watch” era constantemente utilizado para se referir aos *Smartwatches* fabricados pela empresa Apple, e o termo “Amazfit” era predominante para se referir a relógios inteligentes produzidos pela empresa Xiaomi. Por outro lado, para se referir aos produtos de outros fabricantes, os termos mais utilizados se tratavam de uma combinação do nome do fabricante e a palavra Smartwatch (Ex: Samsung Smartwatch, Fitbit Smartwatch). Devido ao grande volume de dados do Twitter, a palavra-chave “*smartwatch*” foi escolhida como termo de pesquisa.

### 3.3.1.1 Twitter

Com base nos estudos anteriores, a extração de dados do Twitter foi realizada utilizando o pacote ‘GetOldTweets3’ desenvolvido em Python para a coleta de *tweets* (A8). Assim, foram coletadas publicações no idioma inglês, utilizando o termo de pesquisa “*smartwatch*” e

contidos no período de 19 meses, entre 01-01-2018 e 31-07-2019. O código utilizado para esta atividade é mostrado na Figura 5.

Figura 5 - código para coleta de *tweets* utilizando Python

```
python GetOldTweets3 --querysearch "smartwatch" --since "2018-01-01" --until 2019-07-31 --lang en
```

Fonte: Elaborado pelo autor

O comando “--querysearch” define o termo de pesquisa utilizado para a busca. Os comandos “--since” e “--until” estabelecem o início e fim do período a ser pesquisado no formato ano, mês e dia. Já o comando “--lang” define a linguagem dos *tweets* a serem pesquisados, utilizando uma codificação de dois caracteres baseado na ISO 639-2.

A execução destes comandos faz a coleta de todos os *tweets* dentro dos parâmetros especificados. Os *tweets* são então automaticamente exportados para um arquivo “smartwatchquery.csv”. Neste arquivo são armazenados dados de cada *tweet*, como data e horário de publicação, texto, nome do usuário, número de respostas ao *tweet*, número de *retweets* (compartilhamentos), número de vezes que foi marcado como favorito, geolocalização do usuário (quando disponível) e *hashtags* contidas no *tweet*.

O pré-processamento de texto dos *tweets* a serem utilizados nas atividades posteriores foi realizado utilizando códigos construídos em Python (A9). Primeiramente foram removidas URLs contidas no *tweets*, e em seguida foram removidos *tweets* duplicados, mantendo apenas a primeira publicação. Foi necessário executar o processo exatamente nesta sequência para evitar que *tweets* repetidos, porém com URLs diferentes, não fossem identificados pelo algoritmo. Uma função também foi utilizada para transformar todas as palavras em minúsculas, de forma a evitar que palavras com diferentes capitalizações fossem identificadas como distintas na análise de frequência de termos. O código utilizado para a fase de pré-processamento é apresentado na Figura 6.

Figura 6 - Código em Python utilizado para pré-processamento do texto.

```
import pandas as pd
import re

df = pd.read_csv(r'C:\Users\jvmsk\Anaconda3\Lib\site-packages\GetOldTweets3-master\bin\Pre-clean\Smartwatch.csv')
tweets = df.text

def remove_url(txt):
    return " ".join(re.sub("([^\0-9A-Za-z \t])|(\w+:\/\/\S+)", "", str(txt)).split())

all_tweets_no_urls = [remove_url(text) for text in tweets]
all_tweets_no_urls = [tweet.lower() for tweet in all_tweets_no_urls] #Lower case words
df.text = all_tweets_no_urls

clean = df.drop_duplicates(subset=['text'], keep='last')
```

Fonte: elaborada pelo autor

### 3.3.1.2 Reddit

Os dados do Reddit foram extraídos no mesmo período que os do Twitter (entre 01-01-2018 e 31-07-2019). A rede social Reddit é dividida em *subreddits*, ou subfóruns, referentes a assuntos específicos. Devido à esta especificidade da rede social, e métodos de extração, a busca de dados foi feita extraindo *posts* publicados em *subreddits* das principais marcas e modelos de smartwatch presentes nesta rede social. Para tal, foram considerados apenas *subreddits* que possuísem mais de 4.000 participantes. Ao todo, foram extraídos dados de cinco diferentes subfóruns: Amazfit (empresa Xiaomi), Fitbit (empresa Fitbit), Applewatch (empresa Apple), GearS3 e GalaxyWatch (Samsung).

O método utilizado para tal atividade, foi o uso das conexões API (*Application Programming Interface*) fornecidas pela plataforma Pushshift, que disponibiliza pontos de conexões entre usuários e dados do Reddit. APIs são padrões estabelecidos por um software que permitem o uso de suas funcionalidades através de programação.

Assim, um código em Python foi construído para realizar a extração de dados de cada *subreddit* e salvá-los em um arquivo de formato .csv, um exemplo do código utilizado é apresentado a seguir na Figura 7.

Figura 7 – Código em Python utilizado para a extração de publicações de *subreddits*.

```
from psaw import PushshiftAPI
import datetime as dt
import pandas as pd

api = PushshiftAPI()
start_epoch=int(dt.datetime(2018,1, 1).timestamp())
end_epoch=int(dt.datetime(2019,7,31).timestamp())

gen =(api.search_submissions(subreddit='GearS3',after=start_epoch,before=end_epoch))
df = pd.DataFrame([c.d_ for c in gen])
df1= df[['subreddit', 'selftext', 'title','created','created_utc']]
epoch_t = df1['created']
real_t = pd.to_datetime (epoch_t,unit='s')
df1['created'] = real_t
df1 ['text'] = df1['title'] +" "+ df1['selftext']
df1.to_csv(r'C:\Users\47994\Documents\TCC\Reddit Data\Subreddit\GearS3.csv',index= None, header=True,sep=';')
```

Fonte: elaborada pelo autor

Similar à extração realizada no Twitter, os parâmetros “start\_epoch” e “end\_epoch” determinam o período de extração. Em seguida, é determinado o *subreddit* do qual os dados serão extraídos. Na extração são obtidos diversos parâmetros de cada submissão, para a pesquisa em questão, foram coletados, o nome do *subreddit* do qual foi extraído, o texto da submissão, título, e data de postagem.

Assim, ao final do processo de extração foram obtidos um total de cinco arquivos, referentes a cada *subreddit* explorado. Os arquivos dos *subreddits* GearS3 e GalaxyWatch foram agregados, já que fazem referência à smartwatches produzidos pela mesma empresa (Samsung). Também foi criado um arquivo agregado com todos os dados extraídos da plataforma, para ser utilizado nas análises de visão geral.

A atividade de pré-processamento dos dados extraídos do Reddit foi realizada de forma similar ao Twitter, com algumas poucas alterações. Os posts extraídos desta rede social são constituídos por título e corpo da submissão, assim, estes dois dados foram concatenados em uma única coluna, que foi chamada de “text”, este ajuste foi feito no próprio código de extração mostrado acima. Outras diferenças são; a quase ausência de posts duplicados, e URLs não incluídas nos campos de título ou corpo do texto, não necessitando, portanto, da atividade de remoção de URLs e duplicatas. Assim, estes dados passaram apenas pelo tratamento de transformação das palavras para minúsculas.

### 3.3.2 Mineração dos dados (M4)

#### 3.3.2.1 Twitter

Utilizando todos os *tweets* extraídos foi realizada uma análise para identificar a quantidade de menções feitas às principais marcas ou modelos de Smartwatch. Os tweets foram analisados em busca dos seguintes termos: samsung, apple, fitibit e amazfit.

A análise de menções foi realizada através do software Excel, utilizando as ferramentas de identificação de texto. Foram criadas colunas para cada um dos fabricantes e caso o *tweet* da linha correspondente possuísse algum dos termos que se referisse à marca, o valor “1” era atribuído à célula, caso contrário, era retornado o valor “0”. Assim foi possível identificar quais *tweets* faziam menção à cada marca de smartwatch. Assim, os resultados foram agrupados em períodos mensais, com a soma de referências a cada marca, e também a porcentagem relativa de menções:

$$n^{\circ} \text{ de menções à marca} \div n^{\circ} \text{ de tweets com menções}$$

Através da identificação das menções dos *tweets*, foram criados outros quatro diferentes arquivos .csv, contendo apenas *tweets* referentes à cada marca, estes arquivos foram utilizados em análises posteriores.

Posteriormente, foi realizada uma análise de frequência dos principais termos referentes a *smartwatches* dentro do período analisado, utilizando uma rotina em Python. Para isto, o texto foi tratado previamente, excluindo palavras que não agregam valor, chamadas de *Stop Words*, e palavras referentes às marcas ou fabricantes (Samsung, Apple, AppleWatch, Fitbit etc.). Além disso, outras palavras não relevantes foram removidas manualmente no arquivo final gerado. O algoritmo utilizado para esta análise é apresentado na Figura 8. Desta forma, foram criadas nuvens de palavras mostrando os 50 termos mais frequentes contidos nos *tweets* durante todo o período.



Figura 8 – Código em Python utilizada para analisar a frequência de termos condidos nos *tweets*.

```
from collections import Counter
from itertools import chain
import pandas as pd
from nltk.corpus import stopwords

df= pd.read_csv(r'C:\Users\jvmsk\Downloads\TCC\Extração-Twitter\Smartwatch\Clean Files\Apple\Apple-all.csv', index_col=0)
tweets = df.text

stops = set(stopwords.words('english'))

col = ['smartwatch', 'rt', 'applewatch', 'apple', 'watch', 'smart', 'amazfit', 'samsung', 'galaxy', 'fitbit', 'reviews', '2019']

words = chain.from_iterable(str(tweet).split() for tweet in tweets)

count = Counter(word for word in words if word not in stops if word not in col)

result = pd.DataFrame(count.most_common(),
                      columns=['words', 'count'])
```

Fonte: elaborada pelo autor

Em seguida, a análise foi realizada para cada conjunto de *tweets* referentes às diferentes marcas de *smartwatch* analisadas, com o intuito de mostrar diferentes termos e aspectos que se sobressaem em cada caso.

Por fim, foi realizada uma análise polar de sentimentos do *tweets* utilizando uma rotina em Python e através da biblioteca TextBlob (A12), mostrado na Figura 9. O método de análise de sentimentos utilizado por essa biblioteca é baseado na classificação de palavras, e atribuição de sentimentos a ele. Portanto, todas as palavras contidas no texto a ser analisado são classificadas dentre uma das classes gramaticais (substantivo, adjetivo, advérbio, preposição, conjunção ou artigo). Em seguida é atribuído um valor entre -1 e 1 para cada palavra, sendo -1 uma palavra de significado muito negativo, e 1 de significado muito positivo. O método também é capaz de identificar intensificadores de adjetivos, como no termo “*very bad*”, onde “*very*” intensifica o valor negativo da palavra “*bad*”. Para gerar o resultado da análise, é calculada uma média das notas atribuídas às palavras para identificar o sentimento do *tweet* como um todo, que é classificado como positivo (valores maiores que 0), neutro (valores iguais a 0), ou negativo (valores menores que 0).

Assim como na análise anterior, os resultados foram agrupados em períodos mensais para analisar a evolução de sentimentos dos *tweets* extraídos no período. A análise de sentimentos também foi realizada para cada uma das diferentes marcas de *smartwatch*, para avaliar se havia uma discrepância entre sentimentos de menções a cada um deles.

Figura 9 – Código em Python utilizado para a análise de sentimentos

```
import pandas as pd
from textblob import TextBlob

df= pd.read_csv(r'C:\Users\jvmsk\Downloads\TCC\Extração-Twitter\Smartwatch\Resultados\Sentiment Analysis\smartwatch-all.csv', index_col=0)
tweets=df.text

def get_tweet_sentiment (tweet) :
    """
    Utility function to classify sentiment of passed tweet
    using textblob's sentiment method
    """
    # create TextBlob object of passed tweet text
    analysis = TextBlob(tweet)
    # set sentiment
    if analysis.sentiment.polarity > 0:
        return 'positive'
    elif analysis.sentiment.polarity == 0:
        return 'neutral'
    else:
        return 'negative'

results = []

for tweet in tweets:
    parsed_tweet = {}
    parsed_tweet['text'] = str(tweet)
    parsed_tweet['sentiment'] = get_tweet_sentiment(str(tweet))
    results.append(parsed_tweet)

sentiment_data = pd.DataFrame (data=results,columns=['text','sentiment'])
```

Fonte: elaborada pelo autor

### 3.3.2.2 Reddit

Sendo o intuito desta pesquisa, mostrar e comparar as informações obtidas de diferentes redes sociais referentes ao mesmo tópico, as mesmas análises de *share of voice*, frequência, e sentimentos foram realizadas para os dados extraídos do Reddit, tanto de forma a dar uma visão geral (utilizando todos os dados extraídos), quanto para dados de cada uma das marcas (utilizando os dados segmentados).

Para a análise de *share of voice* (A10), foi tomada a premissa de que, posts contidos em *subreddits* destinados à certo modelo ou marca, fazem menção a eles. Assim o número de menções a cada marca no período analisado, é igual ao número de posts publicados em *subredits* daquela marca/modelo no período.

As demais análises, de frequência, e sentimentos, tanto para todo o conjunto de dados, assim como para cada marca em específico, foi realizada de maneira análoga às feitas para o Twitter, utilizando os mesmos códigos, salvo mudanças em nomes de variáveis (A11 e A12).

### 3.3.3 Análise dos dados (F4)

Os dados coletados foram analisados e sintetizados com o objetivo de mostrar algumas das informações que podem ser obtidas através da extração de dados de redes sociais e também com o intuito de realizar uma comparação entre as informações obtidas de redes sociais distintas (A13).

Para uma visão preliminar dos dados, as unidades de análise (*tweets* e submissões do Reddit) foram classificados de acordo com a presença de alguns termos com a ajuda da ferramenta Excel, principalmente com o objetivo de identificar a presença de conteúdos de marketing, e pedidos de ajuda com algum problema técnico ou uso de alguma funcionalidade do produto. Para identificar a presença de conteúdos de marketing, os termos “*sales, offer, promotion, e coupon*” foram utilizados. Já para a identificação de conteúdos relacionados à problemas, ou ajuda, foram utilizados os termos “*help, fix e problem*”.

As análises de *Share of voice* e análise de sentimentos foram comparadas através do software Excel, onde também foi possível ter uma visão quantitativa dados.

Por fim, os resultados da pesquisa foram compreendidos e analisados, de forma a apresentar suas contribuições e também pontuando as limitações do trabalho (A14).

## 4. RESULTADOS

### 4.1 Extração de dados e pré-processamento

#### 4.1.1 Twitter

Dentro do intervalo de 19 meses, a coleta de dados do Twitter gerou 545.644 linhas no arquivo final, onde cada linha corresponde a um *tweet* extraído.

Após a remoção dos *tweets* duplicados na atividade de pré-processamento do texto (A9), o número de *tweets* foi reduzido para 298.173, os quais foram classificados em grupos de menções às diferentes marcas de *smartwatch* e posteriormente utilizados para as análises. Ao todo, 73.375 *tweets* fazem menção a alguma marca analisada, uma visão mais detalhada dos grupos é apresentada no tópico 4.2.

A classificação para uma impressão inicial dos dados mostrou uma grande presença de *tweets* relacionados a marketing, e uma pequena porção relacionada a problemas ou ajuda, como pode ser observado na Tabela 1 a seguir.

Tabela 1. Número de *posts* contidos na segmentação inicial dos dados do Twitter.

Classificação	Número <i>tweets</i>	% total <i>tweets</i>
Marketing	11.935	4%
Ajuda/problema	4.528	2%
<b>Total</b>	16.463	6%

Fonte: elaborada pelo autor

#### 4.1.2 Reddit

A coleta de dados do Reddit dentro do período analisado gerou um total de 58.824 linhas, cada uma delas correspondendo a um post na plataforma. Destaca-se que esta segunda plataforma possui um volume de dados consideravelmente menor que o Twitter.

Como mencionado anteriormente, o conteúdo do Reddit é livre de duplicatas, portanto as submissões apenas foram tratadas para que pudessem ser melhor utilizadas nas análises posteriores. Devido ao método de extração (baseado em subfóruns), todas as submissões extraídas fazem menção à alguma das marcas analisadas.

A classificação para uma impressão inicial do conteúdo mostrou uma que uma pequena porção dos dados eram relacionados a marketing, por outro lado, existe uma grande fração das

submissões relacionadas a problemas ou ajuda com o produto, como pode ser observado na Tabela 2 a seguir.

Tabela 2. Número de *posts* contidos na segmentação inicial dos dados do Reddit.

Classificação	Número submissões	% total submissões
Marketing	480	1%
Ajuda/problema	4.205	7%
<b>Total</b>	16.463	6%

Fonte: elaborada pelo autor

## 4.2 Análise de share of voice

Essa primeira análise foi realizada com o intuito de avaliar a presença de cada marca nos *tweets* e posts no Reddit relacionados a *smartwatch*. A análise não reflete necessariamente o número de vendas ou presença digital da marca, porém é uma boa forma de avaliar sua popularidade entre usuários das plataformas. As Tabelas 3 e 4 mostram o total de menções no Twitter e Reddit, respectivamente, a cada uma das marcas analisadas nesta pesquisa.

Tabela 3. Número de menções a cada marca analisada no Twitter

Marca	Número de Menções
Apple	23.962
Fitbit	17.935
Samsung	22.502
Amazfit	8.976
<b>Total</b>	73.375

Fonte: elaborada pelo autor

Tabela 4. Número de posts de *subreddits* relacionados a cada marca analisada

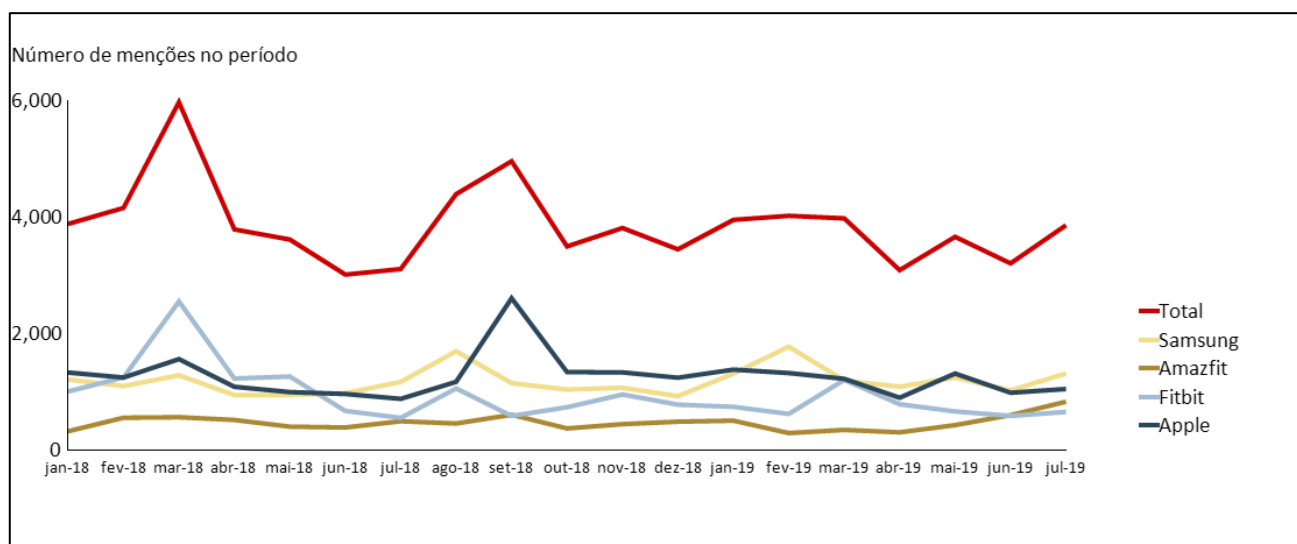
Marca	Número de Menções
Apple	37.434
Samsung	11.101
Fitbit	8.170
Amazfit	2.119
<b>Total</b>	58.824

Fonte: elaborada pelo autor

Para avaliar a evolução do número de menções à *smartwatches* e a cada marca, os dados foram segmentados em períodos mensais, e plotados nos Gráficos 1 e 2, para as plataformas Twitter e Reddit respectivamente. Os dados também foram sintetizados de forma a mostrar a relação do número de menções a cada marca em relação ao total de publicações com menções a alguma das marcas em análise em cada uma das plataformas, como mostrado pelos Gráficos 3 e 4. No caso do Twitter, a análise não inclui *tweets* sem menções a alguma marca avaliada.

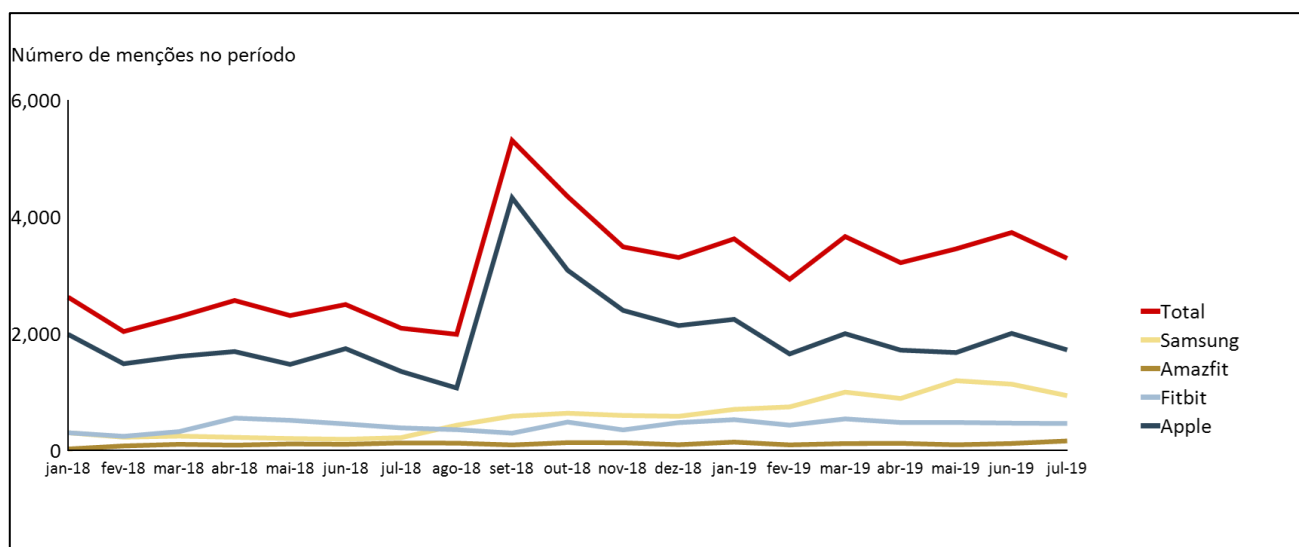
Assim, podemos observar que o número de menções às diferentes marcas oscila em picos, no caso da marca Apple, o pico de menções em setembro de 2018, em ambas plataformas, coincide com o lançamento de um novo modelo de *smartwatch*, o Apple Watch S4.

Gráfico 1 – Número de menções no Twitter às marcas selecionadas durante o período de análise



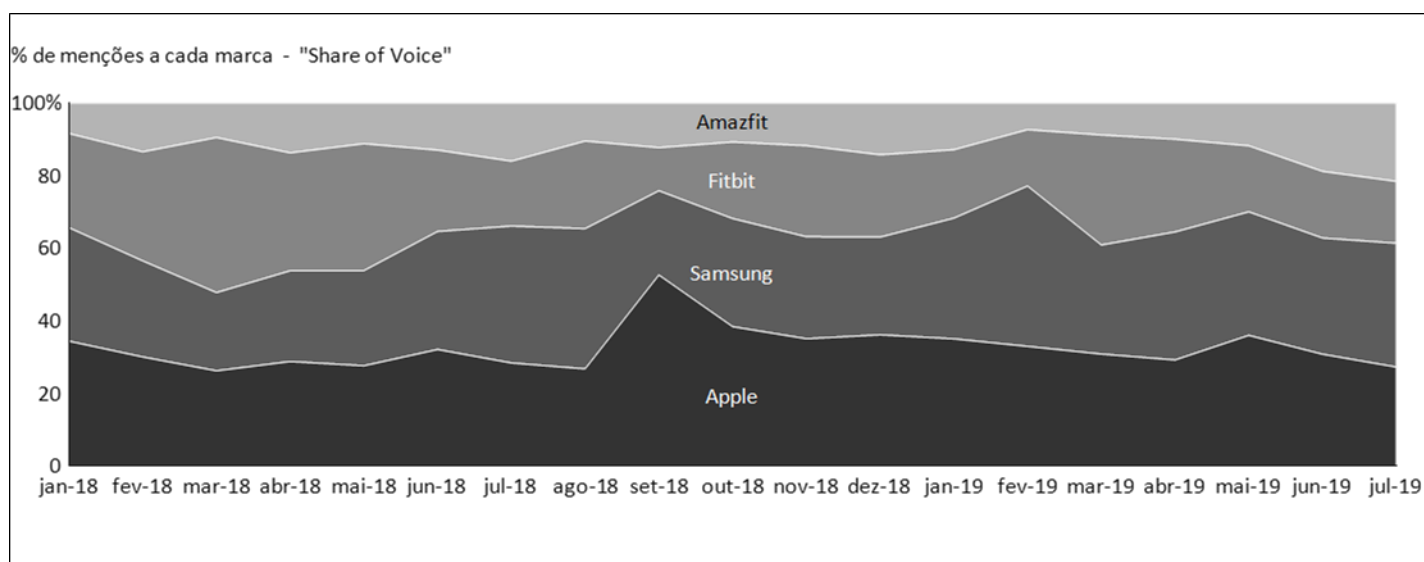
Fonte: elaborado pelo autor

Gráfico 2 – Número de menções no Reddit às marcas selecionadas durante o período de análise



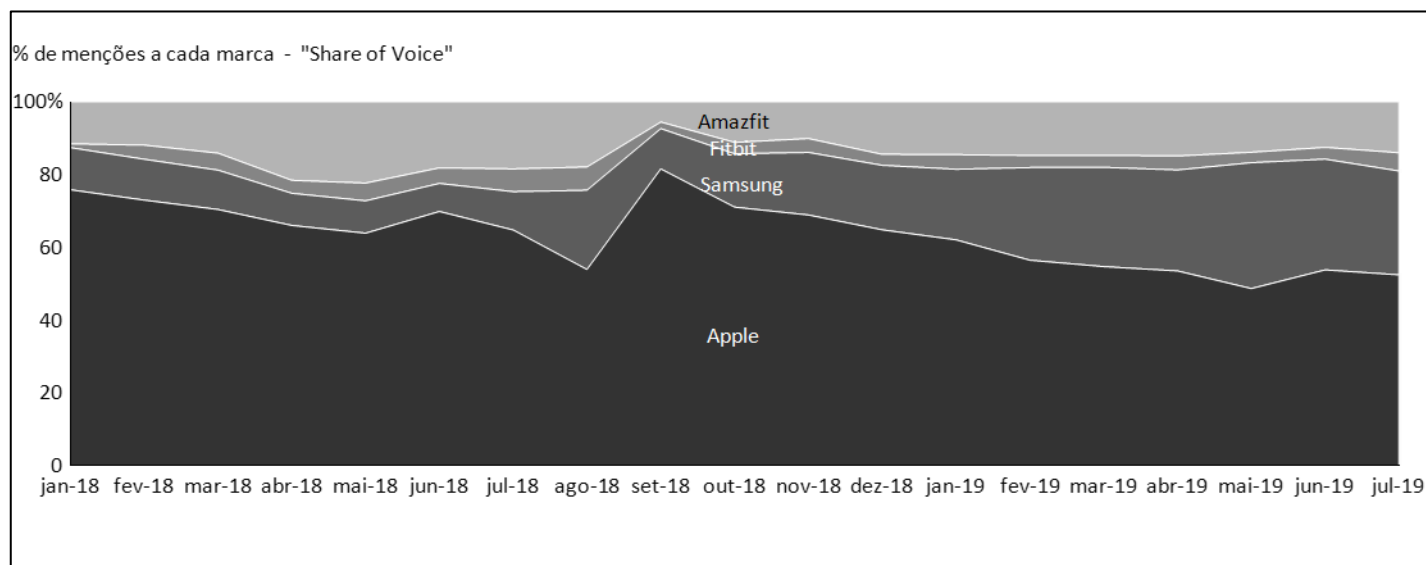
Fonte: elaborado pelo autor

Gráfico 3 – Participação nas menções do Twitter das marcas analisadas na pesquisa



Fonte: elaborado pelo autor

Gráfico 4 – Participação de menções no Reddit das marcas analisadas na pesquisa



Fonte: elaborado pelo autor

Com estes dados é possível observar que, apesar da Apple possuir uma grande vantagem em número de vendas em relação a outras marcas, a diferença entre o número de menções das outras principais marcas no Twitter não é tão discrepante. Já no Reddit, podemos ver que existe uma comunidade da AppleWatch mais consolidada, enquanto as outras marcas, possuem comunidades consideravelmente menores. A marca Samsung é a segunda marca com mais menções em ambas a plataformas, mostrando consistência em sua presença entre as diferentes redes sociais. Por outro lado, a marca Fitbit, que possui um número considerável de menções no Twitter, possui pouca relevância na comunidade dedicada à marca no Reddit.

### 4.3 Análise de Frequência

A análise de frequência foi realizada com o intuito de ter uma visão dos termos mais frequentes relacionados à *smartwatch*. Além de apresentar uma visão inicial da voz dos usuários quanto ao tema, os termos mais frequentes podem ser utilizados como base para a elaboração de métodos de pesquisa de mercado tradicional.

A nuvem de palavras apresentada na Figura 10 foi gerada para identificar os principais termos presentes em todos os *tweets* extraídos, já a Figura 11 apresenta as nuvens de palavras dos principais termos presentes nos *tweets* relacionados à cada marca analisada. Cada nuvem foi criada utilizando os 50 principais termos de cada grupo de *tweets*.

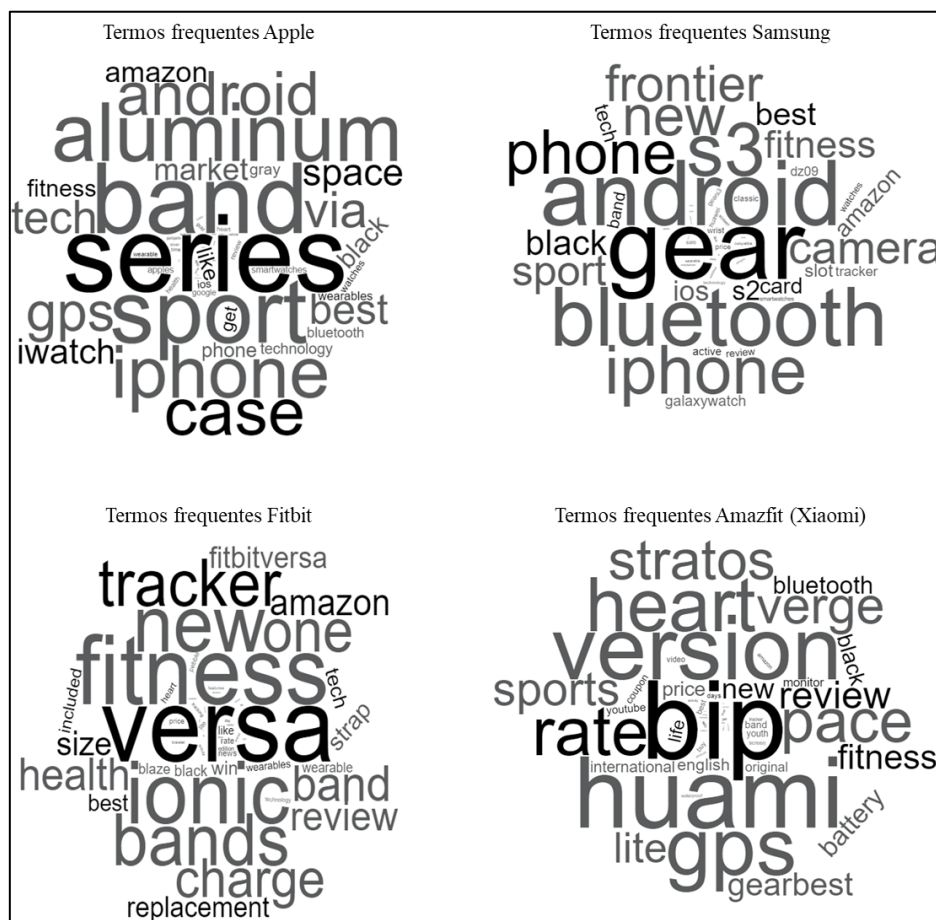


Figura 10. Principais termos recorrentes nos *tweets* extraídos com menções à categoria de produto Smartwatch



Fonte: elaborada pelo autor

Figura 11 – Principais termos presentes em *tweets* relacionados às marcas analisadas na pesquisa.



Fonte: elaborada pelo autor

Analisando tais nuvens de palavras, nota-se que os termos encontrados no Twitter, devido à própria natureza de utilização da rede social, variam desde adjetivos, funções do produto e componentes de *smartwatches*, até menções a modelos específicos e situações de uso do produto. Termos relacionados a características do produto, como *bluetooth*, *gps*, *camera*, *tracker*, e também ao seu uso, principalmente relacionado a atividades físicas; *heart rate*, *sports*, *fitness*, se destacam entre os termos frequentes do Twitter. Uma análise mais profunda pode ser realizada para a categorização destes termos, porém, tal análise não está dentro do escopo deste projeto.

A mesma análise de frequência foi realizada para os dados extraídos do Reddit, a nuvem de palavras apresentada na Figura 12 foi gerada com os termos mais frequentes dentre todos os dados extraídos da plataforma. Já a Figura 13 apresenta os termos frequentes dos dados de cada marca analisada.

Figura 12. Termos mais frequentes nas publicações de *subreddits* relacionados a *smartwatches*



Fonte: elaborada pelo autor

Figura 13. Principais termos contidos nos dados extraídos de *subreddits* relacionados a cada marca analisada.



Fonte: elaborada pelo autor

Os termos mais frequentes em submissões do Reddit também são variados. Além de menções a modelos específicos de *smartwatches*, nota-se também a presença de termos relacionados ao seu uso com esportes. Por outro lado, tratando-se de características ou funções do produto, percebe-se a presença de diferentes termos, como *charge*, *battery*, *app*, *screen* e *notifications*. Através da identificação dos termos mais frequentes, é possível realizar uma análise de conteúdo mais profunda em submissões que contenham estes termos, para um melhor entendimento do que foi abordado em relação ao termo.

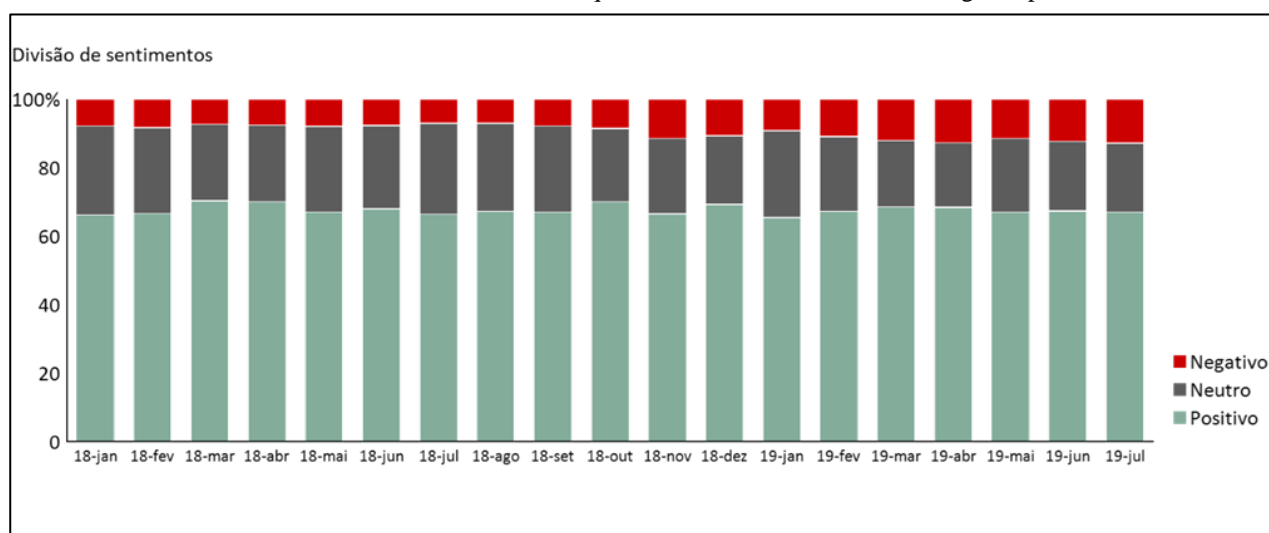
#### 4.4 Análise de sentimentos

A análise de sentimentos dos dados usando o método TextBlob foi realizada com o intuito de avaliar a polaridade de sentimentos expressadas pelos usuários ao falarem da categoria de produto, ou das marcas avaliadas ao longo do período analisado. A análise de

sentimentos se mostra uma ferramenta útil quando se considera o grande número de usuários de redes sociais, o alcance das publicações e impacto das redes sociais na imagem da marca. Esta análise também pode trazer *insights* de comparação de sentimento dos usuários em relação às diferentes marcas. Para que esta análise seja útil para empresas, também é necessário entender como esta proporção de sentimentos positivos e negativos se distribui para a categoria como um todo, para utilizá-la como base de comparação.

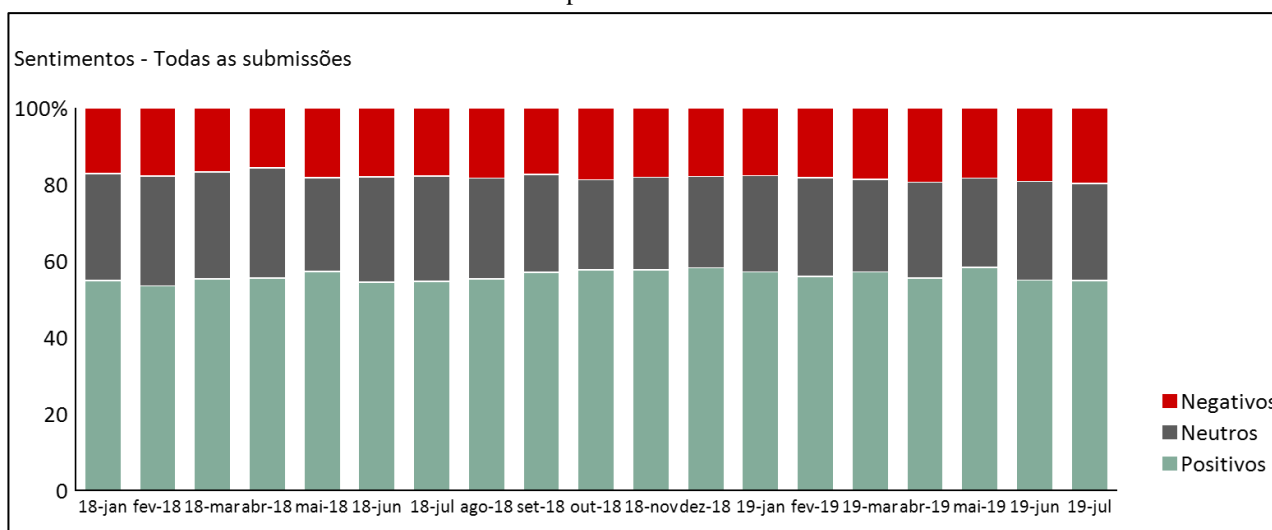
Assim, o Gráfico 3 mostra o resultado da análise de sentimentos realizado para todos os *tweets* extraídos, e o Gráfico 4 mostra o resultado desta análise para os dados extraídos do Reddit, buscando apresentar uma visão geral do sentimento em relação à categoria de produtos *smartwatch*.

Gráfico 3 – Divisão de sentimentos de *tweets* que mencionam *smartwatch* ao longo do período analisado



Fonte: elaborado pelo autor

Gráfico 4 – Divisão de sentimentos de publicações de *subreddits* relacionados à *smartwatches* ao longo do período analisado



Fonte: elaborado pelo autor

Com estes gráficos, é possível observar que, apesar de se notar um aumento na proporção de sentimentos negativos ao longo do período em ambas as plataformas, existe uma predominância de sentimentos positivos em relação ao tema. Por outro lado, podemos notar que a proporção de sentimentos negativos na plataforma Reddit, é consideravelmente maior que no Twitter, uma análise mais profunda no conteúdo de cada plataforma pode ajudar no entendimento desta diferença. Em seguida, foram plotados dois gráficos com os resultados da mesma análise de sentimentos, porém aplicada a cada uma das marcas analisadas em ambas as plataformas (Apêndices 1 e 2). Tais resultados estão resumidos nas tabelas 5 e 6, para as plataformas Twitter e Reddit respectivamente.

Tabela 5. Resumo da análise de sentimentos do Twitter

<b>Empresa</b>	<b>Positivos</b>	<b>Neutros</b>	<b>Negativos</b>
Apple	62%	26%	11%
Samsung	64%	25%	10%
Fitbit	65%	27%	8%
Amazfit	64%	29%	7%
Todos Tweets	68%	23%	9%

Fonte: elaborada pelo autor

Tabela 6. Resumo da análise de sentimentos do Reddit

<b>Empresa</b>	<b>Positivos</b>	<b>Neutros</b>	<b>Negativos</b>
Apple	56%	26%	17%
Samsung	57%	19%	23%
Fitbit	54%	27%	19%
Amazfit	62%	17%	21%
Todas as publicações	56%	25%	18%

Fonte: elaborada pelo autor

Ao comparar os resultados das análises de sentimento obtidos das diferentes plataformas, podemos concluir que:

- O Reddit possui uma maior quantidade de publicações com sentimentos negativos, e menor com sentimentos positivos. Uma análise de conteúdo pode ser realizada para entender os principais tópicos relacionados a sentimentos negativos e positivos.
- A marca Fitbit, que possuía o melhor desempenho de sentimentos positivos no Twitter, possui a menor proporção de publicações com sentimentos positivos no Reddit. Já a

marca Amazfit, que possuía um dos piores desempenhos no Twitter, possui a maior proporção de sentimentos positivos no Reddit, mostrando a disparidade dos resultados de uma mesma marca entre as diferentes plataformas.

## 5. CONCLUSÃO

Uma primeira análise dos resultados, principalmente em relação à frequência de termos e análise do conteúdo levou a algumas conclusões:

- Muitas empresas e pessoas autônomas utilizam o Twitter como uma mídia de marketing, compartilhando avaliações, fotos e promoções dos produtos. Estes *tweets*, apesar de comporem o conjunto de dados analisados, não são vozes do cliente, e não contribuem para os objetivos da análise desta pesquisa. Tais *tweets* apresentam apenas nomes de *smartwatches*, informações de promoções e links para páginas de compra.
- Existe uma grande quantidade de *tweets* repetidos, muitas vezes submetidos pelo mesmo usuário, onde apenas o link final do conteúdo é diferente. Optou-se por remover tais duplicatas, já que muitas vezes correspondem ao mesmo usuário ou são apenas mensagens de marketing.
- Diferentemente do Twitter, não existem posts com o objetivo de marketing e promoções no Reddit. Cada *subreddit* possui regras e moderadores, que vetam este tipo de publicações, melhorando a qualidade dos dados extraídos para os objetivos destas análises.
- As publicações do Reddit tratam, em sua maioria, de dúvidas e discussões sobre modelos, funcionalidades, problemas com o produto, e novidades. Também em contraste com o Twitter, poucos posts são comentários pontuais, além de, devido ao maior limite de caracteres, possuírem mais conteúdo.

Ao avaliarmos as demais análises obtidas das diferentes plataformas:

- Observando a análise de sentimentos realizada para cada uma das marcas apresentada, é possível observar que existe uma maior heterogeneidade de sentimentos entre as marcas na plataforma Reddit, portanto, tal plataforma se mostrou uma melhor escolha para analisar diferenças entre as marcas neste aspecto.
- A análise de frequência obtida através do Twitter mostrou uma baixa presença de termos relacionados às funcionalidades ou características do produto, termos que estão mais presentes no Reddit, indicando que esta segunda plataforma é melhor que o Twitter para a avaliação de requisitos dos usuários.
- O Twitter se mostrou uma boa plataforma para se avaliar efeitos e percepções dos usuários quanto aos produtos, no entanto, não apresentou grandes oportunidades para de gerar inputs que pudessem ser utilizados no processo de desenvolvimento.

Com base nos pontos apresentados acima, conclui-se que o Twitter não é uma plataforma indicada para ser utilizada como fonte de dados para o desenvolvimento de produtos, porém pode ser útil para a avaliação de lançamentos de novos produtos e a percepção dos usuários quanto a eles. Portanto, retomando a Figura 1, o Twitter não é recomendado como uma fonte de dados do usuário como fonte durante a fase de geração de ideias, porém pode ser utilizada como uma forma de se obter a voz do usuário na fase de teste/suporte do produto.

Por outro lado, a plataforma Reddit se mostrou uma plataforma útil para a fase de geração de ideias, se mostrando uma fonte mais confiável, com maior variedade, e qualidade de dados. As análises realizadas nesta pesquisa mostraram que através de um aprofundamento na análise de conteúdo, mais informações poderiam ter sido obtidas, como por exemplo, através da classificação dos termos frequentes, e análise de conteúdo em submissões com termos muito frequentes.

As principais limitações da pesquisa se concentram nos métodos de extração, pré-processamento e análise utilizados no estudo de caso. A escolha da palavra-chave para extração de dados do Twitter pode ser otimizada para cobrir uma maior quantidade de *tweets* relacionados à *smartwatches*, testes com diferentes palavras-chave podem melhorar a fase de extração. O mesmo se aplica ao caso da plataforma Reddit, onde diferentes *subreddits* podem ser utilizados como fonte de dados, englobando um maior número de submissões. Já durante a fase de pré-processamento, não foi realizada a identificação de todos os *tweets* relacionados à marketing e spams, de modo que os dados analisados não estavam completamente livres deste tipo de conteúdo não relevante para as análises. Por fim, a análise de sentimentos foi realizada utilizando apenas um método (TextBlob), o qual não necessariamente trás o resultado mais preciso para este tipo de conteúdo, outros métodos podem ser testados e validados para uma melhor confiabilidade dos resultados desta análise.

Como recomendações para pesquisas futuras, existem oportunidades no aprofundamento da análise do conteúdo de dados extraídos de redes sociais, como a classificação dos termos obtidos através da análise de frequência e mineração dos dados. Também existe a oportunidade de automatização da extração e análise de dados das plataformas, tornando uma rotina e permitindo o acompanhamento contínuo dos dados.



## 6. BIBLIOGRAFIA

- BARRY, A. E. et al. Alcohol Marketing on Twitter and Instagram : Evidence of Directly Advertising to Youth / Adolescents. v. 51, n. November 2015, p. 487–492, 2016.
- BERENDSEN, G. et al. Social media within sustainable product development: an exploratory multiple case study on the perception of social media usability in the new product development process. **International Journal of Technology Intelligence and Planning**, v. 10, n. 3/4, p. 273, 2015.
- BRUNS, A.; LIANG, Y. E. First Monday , Volume 17 , Number 4 - 2 April 2012. **First Monday**, v. 17, n. 4, p. 1–17, 2014.
- CHAE, B. K. Int . J . Production Economics Insights from hashtag # supplychain and Twitter Analytics : Considering Twitter and Twitter data for supply chain practice and research. **Intern. Journal of Production Economics**, v. 165, p. 247–259, 2015.
- CHRISTENSEN, B. C. M. 1 The Innovator's Dilemma. **Boston: Harvard Business School Press.**, p. 1–14, 1995.
- CONNOR, T. O. M.; CONNOR, T. CUSTOMER-LED AND MARKET-ORIENTED : A MATTER OF BALANCE. v. 1163, n. February, p. 1157–1163, 1999.
- EISENHARDT, K. M. Product Development : Past Research , Present Findings , and Future Directions Author ( s ): Shona L . Brown and Kathleen M . Eisenhardt Source : The Academy of Management Review , Vol . 20 , No . 2 ( Apr . , 1995 ), pp . 343-378 Published by : Academy of . v. 20, n. 2, p. 343–378, 1995.
- HALPERIN, R. et al. First Monday. v. 17, n. 4, 2012.
- JEAANMARIE CARR, LUCIEN DECRETON, WEI QIN, B. R. Social media in product development (J. Carr). **International Flavors and Fragrances, Inc., 521 W 57th St #9, New York, NY 10019, United States**, v. 40, p. 88–93, 2013.
- JEAN, E.; ED, P. Local and Global Responses to Disaster : # eqnz and the Christchurch Earthquake Associate Professor Axel Bruns ARC Centre of Excellence for Creative Industries & Innovation Queensland University of Technology , Brisbane , Australia ARC Centre of Excellenc. **Disaster and Emergency Management Conference, Conference Proceedings**, p. 86–103, 2012.
- KAPLAN, A. M.; HAENLEIN, M. The fairyland of Second Life: Virtual social worlds and how to use them. **Business Horizons**, v. 52, n. 6, p. 563–572, 2009.
- KAPLAN, A. M.; HAENLEIN, M. Users of the world, unite! The challenges and opportunities of Social Media. **Business Horizons**, 2010.
- LENGNICK-HALL, C. A. Customer Contributions to Quality : A Different View of the Customer-Oriented Firm Author ( s ): Cynthia A . Lengnick-Hall Source : The Academy of Management Review , Vol . 21 , No . 3 ( Jul . , 1996 ), pp . 791-824 Published by : Academy of Management Sta. v. 21, n. 3, p. 791–824, 1996.
- LEONARD, D. A. Wellspring of Knowledge. n. January 1998, 2015.
- MARTINI, A.; MASSA, S.; TESTA, S. Customer co-creation projects and social media : The case of Barilla of Italy. **Business Horizons**, v. 57, n. 3, p. 425–434, 2014.
- NAMBISAN, S. Designing Virtual Customer Environments for New Product Development :

Toward a Theory Author ( s ): Satish Nambisan Source : The Academy of Management Review , Vol . 27 , No . 3 ( Jul ., 2002 ), pp . 392-413 Published by : Academy of Management Stable URL. **Academy of Management**, v. 27, n. 3, p. 392–413, 2002.

O’CONNOR, G. C. **Market Learning and Radical Innovation: A cross case comparison of Eight Radical Innovation Projects**, 1998.

PLATANIA, M.; SPADONI, R. How people share information about food: Insights from tweets regarding two Italian Regions. **International Journal on Food System Dynamics**, v. 9, n. 2, p. 149–165, 2018.

POCHAMPALLY, R. User context as a source of topic retrieval in Twitter. n. Enir, p. 1–3, 2011.

SANDOR, L. et al. Social media in product development. **2018 Systems and Information Engineering Design Symposium, SIEDS 2018**, p. 88–93, 2018.

SCIENCE, D. F. O. R.; INFORMATION, C. F. O. R.; POLICY, C. Participative Web and User-Created Content. n. 2006, 2007.

STIEGLITZ, S. et al. Socialmedia analytics. **Business and Information Systems Engineering**, v. 6, n. 2, p. 89–96, 2014.

STIEGLITZ, S. et al. Social media analytics – Challenges in topic discovery, data collection, and data preparation. **International Journal of Information Management**, v. 39, n. October 2017, p. 156–168, 2018.

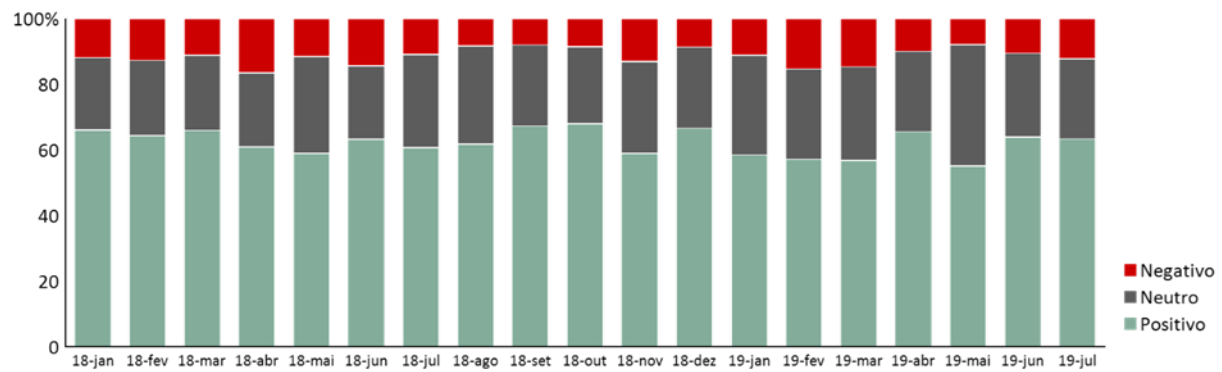
STIEGLITZ, S.; BROCKMANN, T.; DANG-XUAN, L. Usage of Social Media for Political Communication. **16th Pacific Asia Conference on Information Systems**, n. February, p. 22, 2012.

URBAN, G.; HIPPEL, E. VON. Lead User Analyses For The Development Of New Industrial Products. **Journal of Chemical Information and Modeling**, v. 53, n. 5, p. 160, 1989.

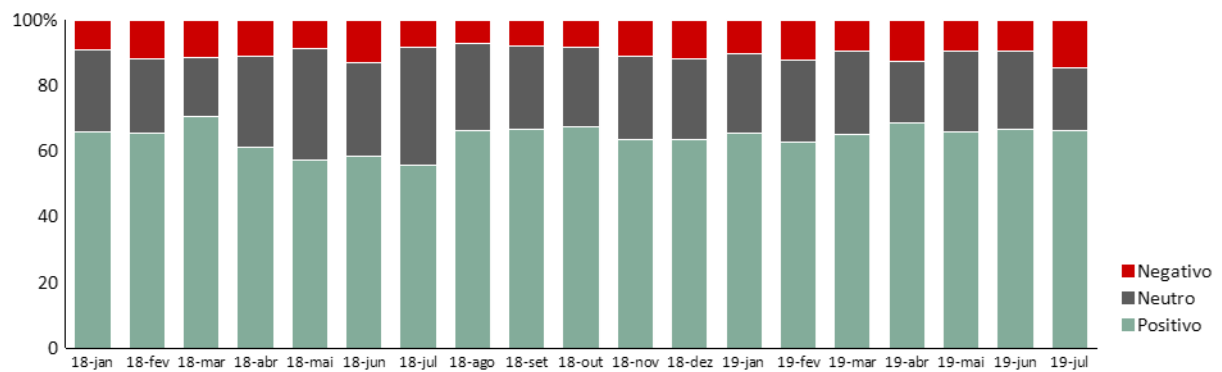
ZENG, D. et al. Social media analytics and intelligence. **IEEE Intelligent Systems**, v. 25, n. 6, p. 13–16, 2010.

## APÊNDICE A – Análises de sentimentos da plataforma Twitter

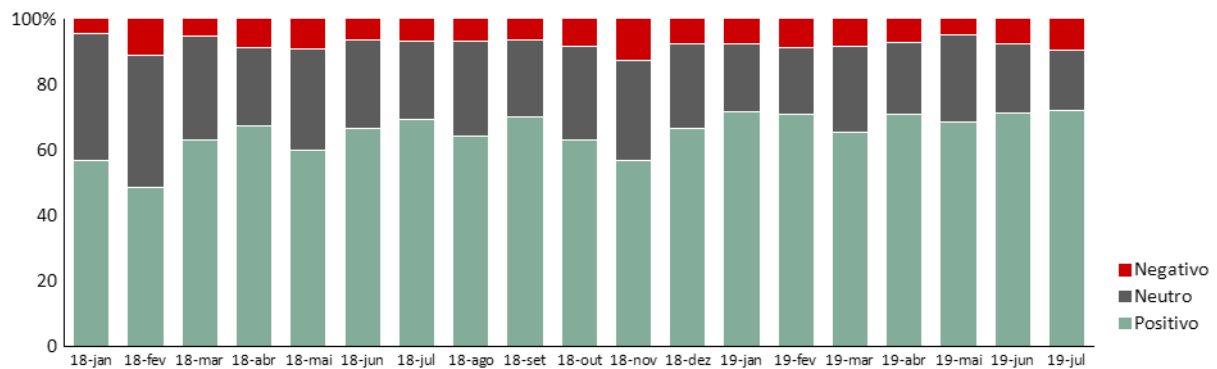
Apple Tweets



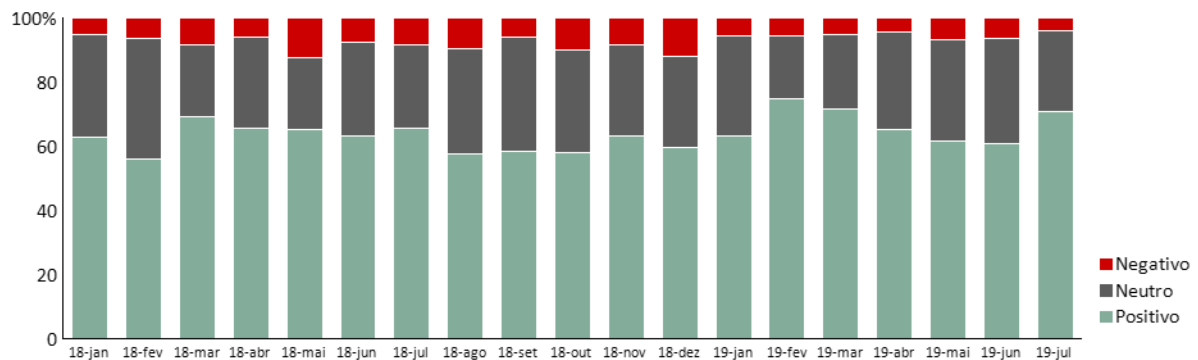
Samsung tweets



Fitbit tweets

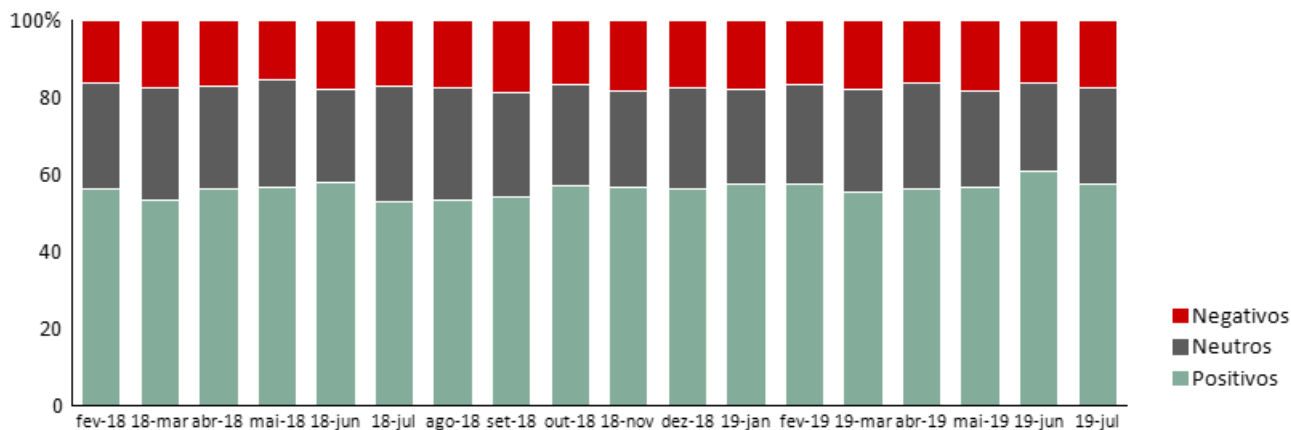


Amazfit tweets

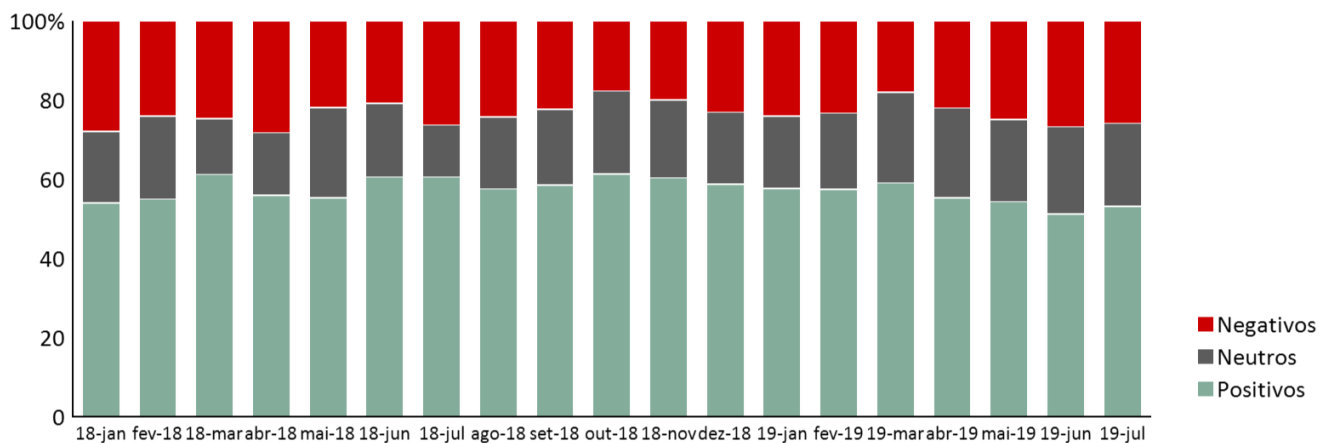


## APÊNDICE B - Análises de sentimentos da plataforma Reddit

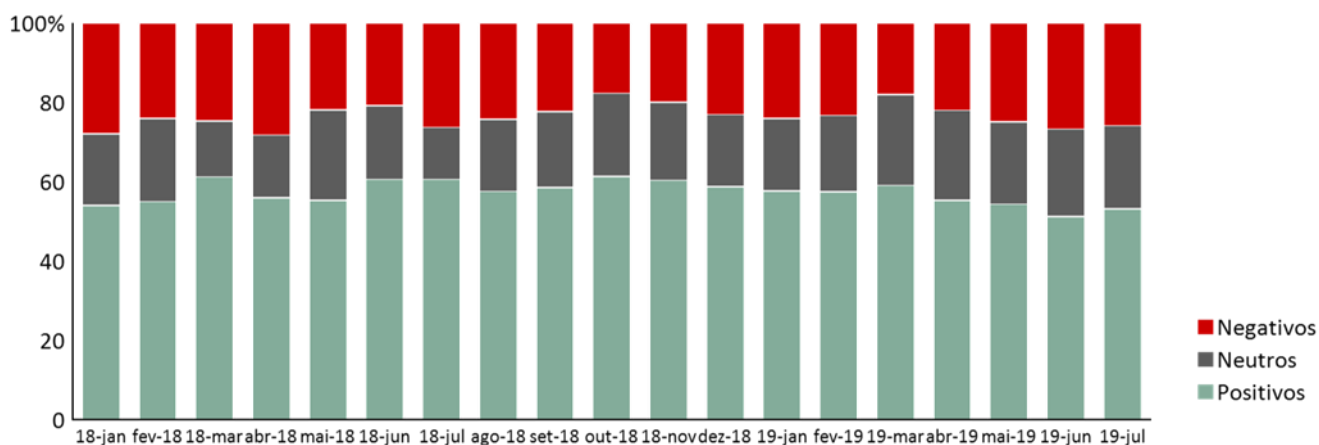
Sentimentos - Applewatch



Sentimentos - Galaxy Watch + Gear S3



Sentimentos - Galaxy Watch + Gear S3



Sentimientos - Galaxy Watch + Gear S3

