



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE ARTES E COMUNICAÇÃO
PROGRAMA DE PÓS GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

CHRIS TAÍNE MUNIZ DA SILVA

**UMA REFLEXÃO DOS ASPECTOS HUMANOS NA ANÁLISE DE SENTIMENTO SOB
A ÓTICA DA TEORIA DAS MÁQUINAS MORAIS**

Recife

2021

CHRIS TAÍNE MUNIZ DA SILVA

**UMA REFLEXÃO DOS ASPECTOS HUMANOS NA ANÁLISE DE SENTIMENTO SOB
A ÓTICA DA TEORIA DAS MÁQUINAS MORAIS**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Ciência da Informação.

Área de Concentração: Informação, Memória e Tecnologia.

Orientador: Prof^o Dr. Célio Andrade de Santana Júnior

Recife
2021

Catálogo na fonte
Bibliotecária Jéssica Pereira de Oliveira – CRB-4/2223

S586r Silva, Chris Taíne Muniz da
Uma reflexão dos aspectos humanos na análise de sentimento sob a
ótica da teoria das máquinas morais / Chris Taíne Muniz da Silva. – Recife,
2021.
107f.: il.

Sob orientação de Célio Andrade de Santana Júnior.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. Centro
de Artes e Comunicação. Programa de Pós-Graduação em Ciência da
Informação, 2021.

Inclui referências.

1. Análise de sentimento. 2. Máquinas morais. 3. Classificação.
I. Santana Júnior, Célio Andrade de (Orientação). II. Título.

020 CDD (22. ed.) UFPE (CAC 2022-27)

CHRIS TAÍNE MUNIZ DA SILVA

**UMA REFLEXÃO DOS ASPECTOS HUMANOS NA ANÁLISE DE SENTIMENTO SOB
A ÓTICA DA TEORIA DAS MÁQUINAS MORAIS**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Ciência da Informação.

Aprovada em: 31/05/2021

BANCA EXAMINADORA

Prof. Dr. Célio Andrade de Santana Júnior (Orientador)
Universidade Federal de Pernambuco

Prof. Fábio Mascarenhas e Silva (Examinador Interno)
Universidade Federal de Pernambuco

Prof. Dr. Diego Andres Salcedo (Examinador Externo)
Universidade Federal de Pernambuco

Profa. Maria Amália Oliveira de Arruda Câmara (Examinadora Externa)
Universidade de Pernambuco

A Deus de onde provém toda sabedoria, tudo que vem dele volta para ele.

AGRADECIMENTOS

A Deus, que tem sido minha força e fortaleza, socorro sempre presente na hora da angústia. A meus pais, meus maiores incentivadores e patrocinadores. Ao meu orientador que aceitou embarcar nessa jornada junto comigo.

"A defesa da liberdade individual repousa, em última análise, no reconhecimento das limitações do nosso próprio conhecimento". (HAYEK, 1983 apud BARBIERI, 2014, p. 1).

RESUMO

A mudança nos meios de comunicação e as novas práticas de relacionamento em redes sociais impulsionaram o surgimento da análise de sentimento, uma técnica que aborda a informação como principal agente de fomento demonstrando a necessidade de identificação e consolidação de opiniões para aprimoração. A motivação desta pesquisa é a nova perspectiva informacional que influencia em meios como economia, política, processos e tomada de decisão baseado nos dados disponíveis, observados o trajeto histórico e comunicacional, além dos agentes relacionados tanto individual quanto coletivamente. O objetivo é discorrer e fazer uma análise crítica sobre as questões sociais e de uso da informação ligadas a análise de sentimento, partindo de uma visão infocomunicacional, identificando seus elementos e elaborando um ciclo de vida para posterior uso e investigação. Para a metodologia será efetuado um conjunto de levantamento e observação sobre a execução no mundo real que evidencie as manifestações da análise de sentimento e quais os seus impactos. Como método, utiliza uma abordagem qualitativa pois exerce interpretação dos fatos. Com meios de revisão bibliográfica, recolhendo informações e detalhando o assunto, abordando visões e aplicações. Ao final desta pesquisa serão respondidas questões ambíguas para a ciência da informação sobre a realidade dos aspectos sociais e das ferramentas utilizadas para uma melhor abordagem da análise de sentimento, fomentando a reflexão para que auxilie na construção de uma crítica sobre como os aspectos humanos, informacionais e tecnológicos cercam a análise de sentimento. Conclui-se que apesar do comportamento receoso da sociedade perante as plataformas e elementos de tecnologia da análise de sentimento, sua utilização e desenvolvimento podem causar transformações estruturais.

Palavras-chave: análise de sentimento; máquinas morais; classificação.

ABSTRACT

The changes in communication and the new relationship practices in social medias drove the emergence of sentiment analysis, a technique that addresses information as the main fostering agent, demonstrating the need to identify and consolidate opinions for improvement. The motivation for this research is the new informational perspective that influences means such as economics, politics, processes and decision-making based on the available data, observing the historical and communicational path, in addition to the related agents both individually and collectively. The objective is to discuss and make a critical analysis on social issues and the use of information related to sentiment analysis, starting from an infocommunicational view, identifying its elements and elaborating a life cycle for later use and investigation. For the methodology, a set of surveys and observations on the execution in the real world will be carried out, showing the manifestations of the sentiment analysis and what are its impacts. It uses a qualitative approach because it interprets the facts. With means of bibliographic review, gathering information and detailing the subject, addressing visions and applications. It is expected that at the end of this research, questions that are still pending for information science about the reality of social aspects and the tools used for a better approach to sentiment analysis will be answered, encouraging reflection to assist in the construction of a critique on how the human, informational and technological aspects surrounding sentiment analysis.

Keywords: sentiment analysis; moral machines; classification.

LISTA DE FIGURAS

Figura 1 –	Ciclo de Vida de Análise de Sentimento de Lima.....	42
Figura 2 –	Ciclo de Vida de Análise de Sentimento da Devopedia.....	46
Figura 3 –	Ciclo de Vida de Análise de Sentimento de Assis.....	49
Figura 4 –	Ciclo de Vida de Análise de Sentimento de Silva, Lima e Barros.	52
Figura 5 –	Ciclo de Vida de Análise de Rodrigues <i>et al.</i>	55
Figura 6 –	Ciclo de Vida de Análise de Sentimento de Godsay.....	58
Figura 7 –	Ciclo de Vida de Análise de Sentimento de Kdnuggets.....	61
Figura 8 –	Processo geral para análise de sentimento.....	63
Figura 9 –	Processo Unificado de Análise de Sentimento.....	77
Figura 10 –	Processo Unificado de Análise de Sentimento.....	78
Figura 11 –	Detalhamento da Atividade Identificar Objetivos da AS.....	79
Figura 12 –	Detalhamento da Atividade Delimitação do Escopo.....	80
Figura 13 –	Detalhamento da Atividade Planejar Execução da AS.....	81
Figura 14 –	Detalhamento da Atividade Elaborar Modelo da AS.....	82
Figura 15 –	Detalhamento da Atividade Testar Modelo de AS.....	83
Figura 16 –	Detalhamento da Atividade Analisar Modelo de AS.....	84
Figura 17 –	Revisar e Atualizar Base de Dados.....	85
Figura 18 –	Aquisição dos Dados.....	86
Figura 19 –	Pré-Processamento.....	88
Figura 20 –	Análise de Sentimento.....	91
Figura 21 –	Auditar Modelo.....	92
Figura 22 –	Validar AS (Usuário).....	93
Figura 23 –	Validar AS (Organização Contratada).....	94
Figura 24 –	Elaborar Relatório Final.....	95
Figura 25 –	Auditar Modelo.....	97

SUMÁRIO

1	INTRODUÇÃO.....	12
1.1	Motivação.....	14
1.2	Objetivos.....	17
2	REFERENCIAL TEÓRICO.....	19
2.1	Aspectos Sociais da Informação.....	19
2.2	Análise de Sentimento.....	21
2.2.1	Histórico e conceitos.....	21
2.2.2	Utilizações da Análise de Sentimento.....	27
2.2.3	Facetas da Análise de Sentimento.....	29
2.3	Máquinas Morais.....	31
2.4	Ciclos de Vida de Análise de Sentimento.....	40
3	METODOLOGIA.....	65
4	CICLO UNIFICADO DE ANÁLISE DE SENTIMENTO.....	68
4.1	Regulamentar Análise de Sentimento.....	78
4.2	Identificar objetivos da Análise de Sentimento.....	79
4.3	Delimitar o escopo.....	79
4.4	Planejar Execução da AS.....	81
4.5	Elaborar Modelo de AS.....	81
4.6	Testar Modelo de AS.....	82
4.7	Analisar Modelo de AS.....	83
4.8	Criar Conteúdo.....	84
4.9	Revisar e Atualizar Bases de Dados.....	84
4.10	Aquisição dos Dados.....	85
4.11	Pré-Processamento.....	86
4.12	Análise de Sentimento.....	89
4.13	Auditar dados processados.....	91
4.14	Utilizar os Dados Ajustados.....	92
4.15	Validar AS (Usuário).....	92
4.16	Validar AS (Organização Contratada).....	93

4.17	Atualizar Modelo da AS.....	94
4.18	Elaborar Relatório Final.....	94
4.19	Tomada de Decisão.....	96
4.20	Auditar Uso da AS.....	96
4.21	Ajustar modo de Uso.....	97
4.22	Conclusão.....	97
5	CONSIDERAÇÕES FINAIS.....	98
5.1	Contribuições.....	99
5.2	Impactos Teóricos e Práticos.....	100
	REFERÊNCIAS.....	101

1 INTRODUÇÃO

A transformação nos meios de comunicação propiciou o surgimento de uma nova cultura de troca de informações e de possibilidades de expressão. As interações que costumavam acontecer em contexto de proximidade, de modo tangível e fortemente influenciadas por questões temporais e geográficas, passaram a ser mais abstratas, assíncronas e virtuais sendo mediadas por *smartphones*, computadores, *tablets* e demais aparelhos utilizando a Internet. Assim, as redes sociais se tornaram “o novo” lugar para a reunião de habilidades, intelecto, experiências, interações profissionais e entretenimento, passando a se consolidar como um dos canais mais comuns para as relações cotidianas além de representar diversas facetas sociais e éticas ligadas não só a comunicação, mas de todo o comportamento humano. (DIZARD JUNIOR, 2000).

Com a difusão desta cibercultura, tornou-se comum o relacionamento ocorrido exclusivamente através da Internet. Pessoas que nunca se encontraram fisicamente, passaram a colaborar, se comunicar, e se expressar pela rede com maior frequência dada às facilidades promovidas pelas ferramentas *on-line* que facilitam a publicação de conteúdo nestes canais (RECUERO, 2009). A consequência dessa mudança de cultura se traduz no aumento exponencial da informação criada e acessada por meio da rede, o que resultou no surgimento de problemas ligados ao uso e propriedade destes dados. Assim, se tornou necessário estabelecer formas para avaliar, classificar e gerenciar as informações publicadas na Internet para que estas se tornassem realmente úteis e que fossem utilizadas de modo ético e responsável.

Entretanto, parte considerável deste conteúdo exposto na rede se refere às avaliações, opiniões e sentimentos expressos voluntariamente pelos próprios interagentes. Estas publicações representam uma expressão sincera e sem filtros que são entregues deliberadamente pelos integrantes do mundo virtual. Assim, não demorou tanto para que estes dados passassem a ser percebidos pelas empresas de *marketing* como uma fonte valiosa de expressões de necessidades pessoais, bem como, uma espécie de *feedback* enviado por usuários dos produtos e serviços por eles consumidos. As empresas passaram a reconhecer a rede como uma nova mídia interativa onde era fácil identificar os interesses dos clientes e, silenciosamente, também era possível descobrir o comportamento e interesses das pessoas. Assim, era possível estabelecer

novos parâmetros para avaliar a conduta, sociabilidade e interesse das empresas para com os seus clientes (CARRANO, 2017). Dessa forma se desencadeia uma nova era nas relações de organizações e consumidores, tudo em volta das informações fornecidas.

Críticas, comentários, curtidas, compartilhamentos e outras formas de expressão feitas por usuários e que estão espalhadas por toda a rede se revelam como uma opinião, ou até mesmo uma recomendação, ainda que informal, e que tem se tornado objeto de interesse para corporações e governos. Esse conteúdo publicado nas mídias sociais, se mostra valioso pois reflete o sentimento real de uma pessoa com pouca, ou nenhuma, censura ou moderação sobre determinado assunto. Estas opiniões por serem espontâneas, gratuitas e públicas, logo passaram a ser coletadas por profissionais ou empresas para que pudessem ser realizadas análises, e assim surge uma espécie de pesquisa de mercado ou de avaliação da satisfação do usuário *online*, de modo silencioso, sem a percepção do usuário.

Hoje, conhecemos este tipo de avaliação como Análise de Sentimento (AS) e esta pode ser realizada por qualquer indivíduo ou organização que tenha acesso a dados públicos expressos nas mídias sociais. Esta técnica surge da clara demonstração da necessidade em compreender o cliente da forma como ele é, e como ele realmente pensa, e não apenas a partir do que ele consome. Aconteceu inclusive a fundação de empresas especializadas em identificar e captar as informações de interagentes para determinar seu comportamento a partir dos dados dispostos na rede utilizando a análise de sentimento como método de análise. Dessa forma, as organizações passaram a compreender os níveis de satisfação do cliente e passaram a desenvolver estratégias para o relacionamento *online* de mão dupla, não se configurando mais o modo passivo onde o cliente simplesmente recebe as recomendações das organizações. Agora, esse cliente presente nas redes sociais, tem a oportunidade de participar de um novo processo de qualificação para as organizações a partir do seu *feedback* (SANTOS, 2010).

Essa necessidade de compreensão do usuário, materializada na captura de expressões, opiniões, sensações, emoções e sentimento sobre um determinado produto, pessoa, ou mesmo ambiente através de texto é abordada por Bing Liu (2010, p.23) ao afirmar que “com a análise de sentimento a informação pode ser processada e classificada através de ferramentas de linguagem”. Foi passado o tempo de que as

mídias sociais eram apenas um canal de expressão para dar lugar a uma arena farta de classificação, quantificação e qualificação de interagentes. A Internet se tornou uma extensa máquina de pesquisa de opinião onde os agentes podem estar conectados vinte e quatro horas por dia.

Esse ciclo é retroalimentado a partir da notificação de que existem novas atividades sendo realizadas por outros interagentes que estão conectados em redes sociais, catalisando o debate em rede que no final do processo nada mais é do que a disponibilização de novos insumos para a classificação do interagente. É a contribuição do participante para o esquema panóptico, que segundo Foucault (1996) se dá

pela extensão de seu alcance a todos os elementos da comunidade envolvida, pela multiplicação dos controlados e pela utilização, inconscientemente voluntária, da contribuição dos próprios esforços dos vigiados para a obtenção do resultado final. (FERREIRA JUNIOR, 1996, p. 12).

Essa abordagem observa as ligações relacionais entre atores sociais, que são analisados tanto individual quanto coletivamente em uma estrutura de rede. Devido ao imenso volume de dados disponíveis nas mídias sociais, se fez necessário desenvolver uma espécie de processamento de dados automatizado, obviamente realizada por máquinas, e esta maneira de analisar informações se tornou uma tendência que só aumentou nos últimos 20 anos.

A principal questão a ser observada até aqui é que o resultado desta classificação realizada por agentes de software tem como principal objetivo atender as necessidades de uma organização ou governo, que quase sempre não estão alinhadas com os valores e desejos dos indivíduos. Assim, esta classificação automática, desprovida de elementos de humanização dos indivíduos, tende tornar as pessoas rótulos ou números que serão submetidos às ações de interesse de tais corporações.

1.1 Motivação

Toffler (2014) afirma que estamos em uma era intitulada a “sociedade da informação”, onde os meios de comunicação colaboram com a criação deste imenso volume de informação que são produzidas e disseminadas pelos indivíduos. Nessa

perspectiva, a economia passa a girar em torno de diversas políticas de informação que se materializam em normas, regras, processos e tomadas de decisões que são baseadas nos dados úteis disponíveis.

Entretanto, para se chegar até aqui, houve mudanças ao longo da história em que os meios de comunicação em massa tradicionais, outrora dominados por pequenos grupos, foram sendo substituídos pelas mídias sociais, que deram voz a uma grande quantidade de indivíduos. Devido ao caráter mais pessoal e intimista trazido por estes meios, pessoas comuns se sentiram atraídas por tais ferramentas o que iniciou uma mudança cultural/social e disseminação de informação (MACIEL; ALBAGLI, 2011).

Morgado (2003) sugere elementos desta transformação ao afirmar que devido a maior diversidade de formas de comunicação, comercialização e globalização implementadas pelas redes sociais, as pessoas incorporaram em seu cotidiano as atividades em ambientes *online* na comodidade dos lares.

Observando a sociedade a partir da forma como essa lida com a informação e seu repentino e exponencial interesse pelas mídias sociais, percebeu-se que as estratégias de marketing das organizações passaram a consistir em compreender o perfil do seu usuário a partir de uma espécie de análise do comportamento padrão de grupos de pessoas em relação ao produto concebido pela marca. A ação destas empresas, passaram a ser muito mais intensas através das plataformas de redes sociais digitais (GABRIEL, 2010).

Dessa forma, o aumento do engajamento dos usuários em mídias sociais incentivou a reflexão de cunho social em contextos informacionais já validados pela Ciência da Informação (CI), e justamente, uma destas aplicações é este escrutínio automatizado das informações do usuário realizados pela Análise de Sentimento. Já existem na Ciência da Informação alguns olhares para a AS, entretanto, estes ainda são incipientes, tratam de aspectos pontuais e tendem a ser mais otimistas e menos reflexivos com os resultados do processo.

Um exemplo é trazido por Malini, Ciarelli e Medeiros (2017) que afirmam que é intensa a criação de conteúdo informacional dos usuários da rede para bancos de dados digitais. Os autores alertam que com esta enorme parcela da população utilizando diariamente as redes sociais, estas acabam por deixar rastros de opinião que podem

auxiliar organizações a melhor compreender seus clientes. Sabe-se que a retenção, agrupamento e processamento de dados, analógicos ou digitais, manual ou automaticamente, são atividades relevantes para a Ciência da Informação, pois, segundo os autores, é nessa ciência que se encontram aqueles que investigam e tem acesso a variados tipos de recursos informacionais que podem desenvolver ferramentas para apoiar as diversas etapas do processo de análise de sentimento.

Porto *et al.* (2012) também apresentam um visão mais tecnicista/processual ao apontar as principais questões informacionais que envolvem a AS:

O novo desafio, entretanto, está em como processar e interpretar essa grande quantidade de informações. Esse desafio é objeto de pesquisa da disciplina denominada "análise de sentimento e mineração de opiniões" [...] considera-se a seguinte definição para o conceito "análise de sentimento e mineração de opiniões": a identificação, extração e o estudo de opiniões, sentimentos e emoções expressas em textos da web (BARACHO *et. al*, 2012, p. 3).

Lima (2015) apresenta uma visão mais utilitária da AS ao afirmar que esta possui um objetivo muito específico: identificar e polarizar opiniões. Assim, os sistemas de informação têm auxiliado a desenvolver estratégias que se adaptam às necessidades da Análise de sentimento uma vez que os requisitos quantitativos e qualitativos são previamente definidos. Neste sentido se observa o número cada vez maior de ferramentas para automatizar ações como coleta, extração, classificação e análise.

Afonso (2017) externaliza uma das escassas opiniões que envolve o contexto social da análise de sentimento. Segundo o autor, devido ao grande acúmulo de informações nos variados meios, bem como, a maior necessidade de expressão dos usuários, cabe à Ciência da Informação a elaboração de técnicas para obtenção de quadros de opinião, a partir das postagens realizadas em mídias sociais digitais, sendo que não só se deve observar as questões técnicas, mas também, aspectos éticos, morais e sociais que são discutidos no meio acadêmico.

Embora os desafios da compreensão e uso da análise de sentimento como objeto de estudo da Ciência da Informação já tenham sido lançados, cabe observar outras facetas desse mesmo objeto. Kennedy (2012) observa, sob a perspectiva das humanidades digitais, questões éticas ligadas à AS. A autora alerta que já é possível que

cada aspecto de um cidadão seja analisado, classificado e quantificado a partir de seus dados coletados na rede. Andriotis e Oikonomu (2015) apontam abertamente os problemas de privacidade de dados ligados a análise de sentimento, e como este processo pode identificar questões sensíveis aos usuários autores das mensagens. Wulff e Hardt (2014) apresentam aspectos de confiabilidade ligados a Análise de Sentimento, e que nem sempre este processo pode ser considerado “confiável” e que o resultado da AS pode depender de questões culturais e locais.

Neste sentido, é possível observar que a análise de sentimento se mostra como uma espécie de “juiz” que classifica pessoas, ou grupos, mediante a um determinado contexto mas que, ocorre em uma escala maquínica crescente e é realizado por meio de um algoritmo “classificador” de pessoas e que, certamente, podem ser questionados certos aspectos morais, sociais, éticos e até mesmo de ordem íntima-pessoal.

Nesta pesquisa será abraçado um princípio básico da Ciência da Informação que é a transdisciplinaridade do saber, baseando-se em que alguns dos assuntos abordados são de caráter tecnológico, ético e psicológico que possuem maior profundidade epistemológica em bases de dados não necessariamente convencionais à CI, sempre preocupando-se em buscar pontos de aproximação das demais ciências fazendo a correlação de suas características. Espera-se assim ampliar o debate dando oportunidade à experimentação de teorias, conhecimentos e práticas obtendo novas reflexões.

1.2 Objetivos

Neste sentido, esta pesquisa tem como objetivo propor um processo unificado de Análise de sentimento. Discorrer sobre as questões sociais ligadas ao uso da Análise de Sentimento para classificar pessoas e grupos na rede sob a ótica da teoria das máquinas morais (*moral machines*). O que se pretende discorrer são os aspectos éticos, de privacidade, segurança, confiança, propriedade das informações, acessibilidade, assertividade e utilidade da Análise de Sentimento e que já são impactantes no cotidiano

individual e que podem vir a ser, em um futuro ainda especulativo, determinantes para a construção da sociedade.

Para tanto são vislumbrados os seguintes objetivos específicos:

- Identificar os elementos utilitaristas, processuais e estruturais da Análise de Sentimento. Ou seja, quais são as finalidades da análise de sentimento e como ela é executada nos dias atuais. O intuito deste objetivo é apresentar as possibilidades técnicas e gerenciais da AS e quais as limitações e problemas que envolvem as mesmas considerando estes elementos;
- Elaborar um ciclo de vida unificado para se investigar a Análise de Sentimento. Após investigar os elementos processuais e estruturais da AS, será concebido um ciclo de vida unificado que condense, e que ao mesmo seja o padrão e bússola para este trabalho dos termos e atividades ligadas a AS. Já foram observadas divergências nas nomenclaturas e atividades propostas para Análise de Sentimento por diversos autores, e neste sentido, se espera normatizar a AS em uma base única e que serve de modelo para o uso neste trabalho.
- Identificar os elementos sociais que envolvem a análise de sentimento. Aqui, espera-se não só compilar, mas, enriquecer o debate sobre as implicações da adoção da AS em larga escala considerando a visão destes elementos sociais;
- Analisar as implicações encontradas no objetivo anterior e a partir desta análise e sugerir questões sociais que precisam de uma maior atenção por parte da comunidade acadêmica e que devem ser levadas a sociedade, indústria e governos.

Não se espera nesta dissertação esgotar a discussão sobre os aspectos sociais que estão envolvidos na Análise de Sentimento, muito pelo contrário, o que aqui se propõe é alavancar um debate, que ainda se encontra incipiente, sobre a implicação do tratamento automatizado de informação na vida das pessoas comuns. Então, considera-se que este trabalho é um possível ponto de partida para o fomento desta discussão.

2 REFERENCIAL TEÓRICO

2.1 Aspectos Sociais da Informação

A sociedade contemporânea já apresenta alguns aspectos de transformação no comportamento em relação ao manuseio de dados e informações que no dia a dia passam despercebidos pelos interagentes, mas que, na maioria das vezes, usurpa um lugar de privacidade e segurança que beira um monopólio onde não se apresenta nem legislação, nem responsabilização, nem ética.

Iniciamos esta reflexão a partir de Brookes (1980) que afirma que a informação é um elemento que provoca transformações nas estruturas. Uma mensagem é enviada a partir de um conjunto de códigos que pode ser compreendido pelo emissor e pelo receptor para que seja interpretada e ganha sentido. Utilizar essa informação faz do sujeito um agente transformador de seu ambiente, ele sai do seu estado atual de inércia e constrói uma relação entre informação e conhecimento. Essa identificação e correlação é construída através dos conceitos e ligações com a visão de mundo ao qual se encontra. Antes de ter a informação havia o estado anômalo do conhecimento, a busca pela informação é a necessidade que se tem de sair deste estado.

O desenvolvimento de habilidades específicas relacionadas ao trabalho informacional pode fazer parte de um esforço para proporcionar ao usuário os recursos necessários para lidar com a informação necessária para atender os seus anseios informacionais. Tal lacuna de conhecimento pode ser descoberta quando comunicado pela pessoa que detém, por meio de dedução, ou através do comportamento dos indivíduos. Wilson (1981, p. 6) tipifica as necessidades em cognitivas, afetivas e fisiológicas e assinala que, no caso das necessidades informacionais, existem também “motivos” na origem dos comportamentos informacionais: “por qualquer razão a necessidade de informação deve ter um motivo que ocasiona esse comportamento”.

Tendo este ponto fundamentado observa-se o comportamento do indivíduo no processo das práticas informacionais, recuperação, geração, transferência e recepção de informação, ações que ocorrem no interior de um sujeito cognitivo (ARAÚJO, 1998). Na atualidade as inovações e tecnologias apresentam um impacto na estrutura da sociedade contribuindo com a transformação e articulação da informação de acordo com

a necessidade de cada indivíduo. A informação então se tornou uma base articuladora sociocultural acentuando ainda mais a necessidade e uso informacional (PINTO; FIDELIS, 2012).

Em um contexto mais atual e voltado ao tema, destacamos Shoshana Zuboff (2019), em seu livro *Capitalismo de Vigilância*, onde esta se refere à prática do tratamento de informações pessoais de forma comercial. Destacando a relevância da coleta e do tratamento de informações pessoais e de como isso impacta em toda a sociedade a partir da transição de dados do meio digital para o mundo real. Inicialmente o tratamento de dados focava em identificar o indivíduo, mas agora máquinas inteligentes buscam compreender o comportamento populacional em rede.

Zuboff (2019) sugere que a poderosa conjuntura política, econômica e tecnológica interfere na tentativa de uma elaboração ao direito de privacidade, uma vez que, é interesse das grandes marcas que os dados pessoais estejam completamente disponíveis à elas sem qualquer legislação que regule o manuseio. Inicialmente feito de forma muito sutil há duas décadas atrás pelo Google com o uso de palavras chave, padrões nos termos de pesquisa, ortografia, padrões de cliques, localização, tempo de permanência na página, entre outros. Cada plataforma foi elaborando o método de monitoramento que lhe desse maior competitividade e ainda ajudando no processo de aprendizagem de padrões para a máquina que teria melhoria contínua, quanto mais utilizá-lo, mais relevante ele se tornará.

Até os anúncios com o passar das décadas foram se tornando cada vez mais relevantes para os usuários, a publicidade é direcionada a um indivíduo, garantindo assim a qualidade da oferta para os anunciantes. Assim as plataformas digitais passaram a monetizar a publicidade na internet, o marketing direcionado tornou-se fundamental na competição mercadológica. Para Zuboff (2019) a atual geração já é refém de uma estrutura de poder complexa de dominação e vigilância.

2.2 Análise de Sentimento

2.2.1 Histórico e conceitos

A Análise de Sentimento, também denominada como mineração de opinião, é o campo de estudo que analisa a opinião das pessoas, sentimentos, avaliações, apreciações, atitudes e emoções para com entidades como produtos, serviços, organizações, indivíduos, publicações, eventos, tópicos e seus atributos. A matéria prima a ser observada são as postagens, as informações, que os usuários tendem a compartilhar através das redes (LIU, 2012).

A Análise de Sentimento pode ser considerada também um processo de classificação que visa investigar um usuário que expresse uma opinião ou um sentimento positivo, negativo, ou neutro, considerando cada sentença como um curto documento. A correta classificação de tais considerações é, portanto, fundamental para a tomada de decisão de organizações (MEDHAT; HASSAN; KORASHI, 2014). Algumas métricas são amplamente adotadas com o propósito de analisar e comparar as diversas abordagens utilizadas na análise de sentimentos (MARTINS; PEREIRA; BEEVENUTO, 2015).

A Internet é uma ferramenta que propicia multiplicação do número de vozes e aproxima o usuário das grandes organizações, podendo assim influenciar os processos de decisão destas, ao mesmo tempo em que acelera a disseminação de informação, verdadeira ou não, sobre qualquer assunto ou evento, trazendo impacto e consequências (WESTCOTT, 2008).

Uma vez que todas essas opiniões sobre produtos estão disponíveis online, é possível afirmar que então também cada um de nós está presente, de certa forma, digitalmente. A partir dessas interações na rede, é gerada uma quantidade massiva de informação que são aproveitadas por algumas grandes empresas. Essa coleta de informação sobre os sentimentos de cada usuário ocorre hoje em dia a uma escala até há pouco tempo inimaginável. Isto representa enorme potencialidade em suprir as necessidades e superar certos desafios que vão desde uma maior eficiência e conhecimento na gestão de recursos até responder perguntas em vários níveis e dimensões de análise (CARVALHO; SANTOS, 2017).

E finalmente, o grande desafio a ser superado é responder as questões pertinentes para a definição e apropriação segura do uso da análise de sentimento da forma mais eficaz. Havia uma quantidade pequena de pesquisas sobre opiniões e sentimentos das pessoas antes do ano de 2000 e, desde então, o campo tornou-se uma área fértil para pesquisas e há algumas razões para isso. Em primeiro lugar, a grande amplitude de aplicações da AS que atinge quase todos os domínios. Segundo a indústria em torno da análise de sentimento também floresceu devido à proliferação de aplicações comerciais, isso fornece uma forte motivação para a pesquisa. Em terceiro lugar, ela oferece muitos problemas de pesquisa desafiadores, que nunca tinham sido estudados antes. E em quarto lugar, pela primeira vez na história humana, temos agora um enorme volume de dados opinativos na web (LIU, 2010).

Malini, Ciarelli e Medeiros (2017) afirmam que é intensa a criação de conteúdo informacional dos usuários da rede para os bancos de dados digitais e, uma enorme parcela da população utiliza diariamente as redes sociais deixando rastros de opinião. Neste sentido, a retenção, agrupamento e processamento desses dados são indispensáveis para as equipes de marketing e, ao mesmo tempo, tornando a AS um objeto relevante para a Ciência da Informação. Segundo os autores, é nessa ciência que estão aqueles que estudam e têm acesso a variados tipos de recursos informacionais passam a utilizar e desenvolver ferramentas para apoiar as diversas etapas desse processo.

Os sistemas para análise de sentimento de dados qualitativos não estruturados disponíveis na Web possuem arquitetura que normalmente incluem as fases de seleção, coleta, extração de conteúdo, classificação e análise e se beneficiam dos avanços em duas grandes áreas: a ciência da computação e ciência da informação. Nessas áreas alguns temas despontam, como recuperação de dados da Web, mineração de dados, mineração textual, mineração de opiniões, análise de sentimentos [...] (LIMA, 2015, p. 334).

A combinação de regras sobre o estudo das ferramentas abordadas junto com a linguagem natural é apontada como uma solução focada na melhoria dos resultados da Análise de Sentimento. Destaca-se também a importância da análise na identificação do sentido das informações (MALINI; CIARELLI; MEDEIROS, 2017).

Uma vez que o grande acúmulo de informações está cada vez maior nas redes, e cada vez mais os usuários quem se expressar, cabe à Ciência da Informação a elaboração de técnicas para obtenção de quadros de opinião, a partir das postagens em mídias sociais digitais, sendo um desafio que é proposto ao meio acadêmico (AFONSO, 2017).

Quanto ao histórico da Análise de Sentimento, iniciamos com o relato de Richmond (1998) quando este afirma que o interesse humano na opinião dos seus semelhantes é quase tão antigo quanto a própria comunicação verbal, e, segundo o autor não são incomuns os relatos históricos sobre líderes intrigados com os pensamentos de seus subordinados sobre si. A partir da descoberta e avaliação destes sentimentos, tais governantes determinavam se deveriam barrar ações oposicionistas, ou se deveriam agir para aumentar sua popularidade.

O autor retrata diversas tentativas de detecção de dissidências nos tempos da Grécia antiga e o mesmo dissertou sobre as artimanhas dos governantes para descobrir informações referentes a revoltas e traições, e elencou os mais ardilosos estratagemas para enganar e manipular as pessoas para que a posição dos poderosos fossem mantidas.

Noutra direção, Mantyla, Graziotin, e Kuutila (2018) apresentam como o sentimento das pessoas passou a ser representado pelo voto que passou a ser adotado como método para medir a opinião pública em questões políticas referentes à cidade-estado de Atenas no século 5 AC.. Neste período foi criado o conceito de “Doxa”, que remetia à crença popular e o voto era usado como meio de coletar a opinião pública e o mesmo servia para formular políticas e também para derrubar a oposição. Segundo os autores, desde essas épocas, os esforços para capturar a opinião pública quantificando e medindo-a foram evoluindo até a forma dos questionários elaborados nas primeiras décadas do século XX.

No sítio da devopedia¹, remete à 1824 a elaboração do primeiro exemplo conhecido como pesquisa de opinião que foi conduzida pelo *The Aru Pennsylvanian* para determinar quem estava na liderança na corrida para a Presidência dos Estados Unidos

¹ <https://devopedia.org/sentiment-analysis>

naquele ano. Esta pesquisa mostrou que o republicano John Quincy Adams foi ultrapassado por Andrew Jackson, que não só conquistou a Pensilvânia, mas também a presidência do país. A precisão das enquetes logo as tornou populares em toda a América e pesquisas em todo o país começaram a ser conduzidas alguns anos depois e previram corretamente os próximos vencedores presidenciais.

Pang e Lee (2008) apontam que os estudos sobre a opinião pública tiveram como ponto de partida a expressão da crença dos indivíduos e este foi evoluindo ao ponto de se interpretar metáforas, narrativas, pontos de vista, afetos e evidências presentes em textos que se consolidaram na área da linguística.

Chegando a épocas mais recentes, Hatzivassiloglou e McKeown (1997) cunharam o termo 'orientação semântica', que ao ser aplicada em opiniões ou frases, pode caracterizá-las como positiva, negativa ou neutra. Este conceito forma a base de todas as técnicas de análise de sentimento utilizadas até hoje. Pang e Lee (2004) apontam a possibilidade de classificação de sentimento com base em técnicas de aprendizado de máquina, sendo este considerado, à época, o "próximo grande avanço". Os autores propuseram um 'detector de subjetividade' que filtraria as sentenças rotuladas como 'subjetivas' em um documento e empregaria técnicas de categorização de texto nos dados resultantes. Eles implementaram alguns algoritmos como Naive Bayes (NB) e Support Vector Machine (SVM) para encontrar cortes mínimos em um gráfico. Os resultados apontados pelos próprios autores são de uma precisão de 86,4% no classificador de polaridade no Naive Bayes².

Daniel Gruhl *et al.* (2005) conduziram um dos primeiros estudos para determinar se os comentários *online* tinham relação com o número de vendas de um determinado produto. Eles obtiveram os dados de vendas de produtos da Amazon.com e usaram algoritmos de geração de consulta automatizada para prever o aumento e a queda das vendas de certos livros, com base em menções em *blogs* e conversas *online* sobre a bibliografia específica e descobriram que comentários positivos levavam ao aumento das

² : O modelo de algoritmo Naive Bayes realiza uma classificação probabilística através da observação, caracterizando-as em classes pré-definidas.

vendas. Hodiernamente, a Internet se consolida como o maior espaço público de propagação de opiniões e o mais importante canal de liberdade de expressão de indivíduos. Entretanto, o conteúdo disponibilizado pelos usuários na rede também está sendo alvo de escrutínio de grandes corporações e governos, no intuito de conhecer melhor os consumidores/eleitores e assim atingir os seus objetivos.

Tal análise, hoje denominada como Análise de Sentimento, pode ser realizada a partir dos dados públicos expressos pelos interagentes nas mídias sociais. A partir do conteúdo disponível na rede, tais organizações, seguiram o exemplo da Amazon e passaram a compreender os níveis de satisfação do cliente e passaram a desenvolver estratégias para um relacionamento *online* de mão dupla, onde uma pessoa comum tem a oportunidade de participar de um novo processo de qualificação para as organizações a partir do seu feedback (SANTOS *et al.*, 2010).

É possível perceber que, não foi na área de linguística que o termo Análise de Sentimento foi cunhado, mas sim nos estudos ligados ao uso da tecnologia que apresentavam o interesse particular em minerar opiniões na rede. Pang e Lee (2008) afirmam que alguns termos como: mineração de opinião, análise de sentimentos, análise de subjetividade estabelecem conexões com a computação afetiva, que tem como intuito fazer com que máquinas reconheçam as emoções e sentimentos humanos.

Malini, Ciarelli e Medeiros (2017) também são muito precisos ao aproximar a área da análise de sentimento do uso de tecnologias ao utilizarem o termo “análise computacional de sentimento”. Souza e Café (2018) definem a Análise de Sentimento ou Mineração de Opinião como:

uma técnica derivada da linguística e da inteligência artificial utilizada para identificar opiniões e emoções em documentos textuais, avaliando-os como positivas ou negativas em relação a algum contexto ou objeto aludido, utilizando para isso o processamento automático da linguagem natural (PLN) (SOUZA; CAFÉ, 2018, p. 275).

Li e Liu (2012) afirmam que o objetivo da AS é predizer a orientação de um sentimento (positivo, negativo ou neutro) por meio da análise de palavras ou expressões de sentimentos ou opiniões em frases e documentos. Consequentemente, esta avalia a atitude do escritor de comentários em relação a um tópico específico. Esta opinião pode

ser o julgamento ou avaliação, do estado afetivo do emissor ou a comunicação emocional pretendida. Ainda segundo os autores, existem duas maneiras para classificar os sentimentos: (i) o primeiro é voltado para classificar um grande número de opiniões em orientações bipolares (positivas, negativas ou neutras), ou (ii) na segunda onde se objetiva identificar a subjetividade/objetividade de uma opinião quantificando a assertividade da avaliação por parte da máquina.

Kucuktunc *et al.* (2012) foram os pioneiros na análise de sentimento em larga escala do Yahoo! Respostas. Eles descobriram que as respostas diferiam de acordo com os atributos dos usuários, ou seja, as respostas mais bem avaliadas tinham um tom neutro. Eles também identificaram quais sentimentos foram evocados em um usuário ao ler uma determinada questão. Essas descobertas reveladoras começaram a ser usadas em publicidade e recomendações.

Lima (2015) aponta que a AS é aplicada em diferentes domínios. Na área de finanças a análise de sentimento é utilizada para identificar fatores de risco, prever ganhos futuros, prever preços de ações, impacto na liquidez de mercado e detectar fraudes. No marketing, a AS é utilizada para identificar satisfação de clientes, recomendação produtos. Por fim, esta também é utilizada na política para mapear intenções de voto. Souza e Café (2018) relatam aplicações da AS ligadas a área de educação com o intuito de determinar o ânimo de alunos em ambientes EAD e também para a criar categorias afetivas entre alunos.

Em junho de 2014, um pesquisador chamado Aleksandr Kogan publicou um banco de dados contendo informações de cerca de 87 milhões de usuários do Facebook para a empresa de criação de perfis eleitorais Cambridge Analytica. A empresa usou tais informações para fazer 30 milhões de perfis “psicográficos” sobre os eleitores. Os dados teriam sido usados para tentar influenciar a opinião do eleitor em nome dos políticos que os contrataram.

O próximo avanço veio quando Poria *et al.* (2017) perceberam que estes foram capazes de identificar os sentimentos presentes em vídeos online. Eles usaram a análise de sentimento no nível de sentença. Uma 'sentença' é uma unidade de fala limitada por uma pausa, e cada enunciado foi classificado subjetivamente. Ou seja, a partir do trabalho deles seria possível analisar a polaridade de diversas sentenças dentro de um

mesmo documento ou mesmo de uma frase. A evolução da Análise de Sentimento se dava de forma muito rápida, embora a ideia básica de analisar a opinião pública continue a mesma, as ferramentas disponíveis e as áreas de uso cresceram consideravelmente.

Considerando as definições para a Análise de Sentimento podemos revisitar a visão processual trazida por Porto *et al.* (2012) quando estes afirmam que a AS consiste na identificação, extração e o estudo de opiniões, sentimentos e emoções expressas em textos da web. Podemos revisitar também a visão conceitual de Liu (2012) ao afirmar que a Análise de Sentimento, também denominada como mineração de opinião, é o campo de estudo que analisa a opinião das pessoas, sentimentos, avaliações, apreciações, atitudes e emoções para com entidades como produtos, serviços, organizações, indivíduos, publicações, eventos, tópicos e seus atributos.

Pang e Lee (2008) apresentam uma definição mais tecnicista ao afirmar que a Análise de sentimento ou mineração de opinião é um ramo da mineração de textos preocupado em classificar textos não por tópicos, e sim pelo sentimento ou opinião contida em determinado documento. Geralmente associado à classificação binária entre sentimentos positivos e negativos, o termo é usado de uma forma mais abrangente para significar o tratamento computacional de opinião, sentimento e subjetividade em textos. Liu (2010) também apresenta uma visão tecnicista da AS ao afirmar que esta é o estudo computacional das opiniões, avaliações e emoções das pessoas, expressas em texto, relativas a entidades, eventos e seus atributos.

2.2.2 Utilizações da Análise de Sentimento

Com a popularização das redes sociais, cresceu o uso da Análise de Sentimento e com isso várias foram as ferramentas desenvolvidas para atender as demandas existentes para o uso dessa abordagem para análise de dados e informações sobre diferentes tópicos e perspectivas. Sistemas que realizam a análise de sentimento combinam linguagem natural e técnicas de aprendizado de máquina capacitando a inteligência artificial a identificar tais sentimentos em opiniões expressas em redes sociais.

Classificar sentenças opinativas recolhidas das redes sociais em positivo, negativo ou neutro é o meio mais clássico de obter resultados, mas não o único. É possível reconhecer outros padrões passíveis de análise como em *emoticons*, memes, gifs e imagens postados em relação ao objeto em análise, e pode ser aplicado até mesmo na análise de entonação de áudios. Além disso, não apenas as redes sociais são usadas como fonte de matéria prima, mas também portais de serviço de atendimento ao consumidor, sistemas de *reviews* e avaliações, fóruns e outros.

Entre as ferramentas mais comuns está o “Alerts” do Google, ela é gratuita e permite que sejam rastreados e monitorados conversas, postagens, compartilhamento sobre uma palavra ou negócio. É possível acompanhar a presença do termo escolhido online e observar em primeira mão o que os interagentes têm comentado sobre ele, além de verificar possíveis potencializadores do negócio, entre queixas e elogios.

Existe também o *Facebook Insights* que é uma ferramenta feita para uso exclusivo da rede social, apresentando acréscimo ou perda dos números de fãs, likes, dados demográficos, visualizações, curtidas e descurtidas. Elaborando métricas completas para uma análise de sentimento, no grande volume de informações proveniente desta rede social.

Já para a rede social Twitter existe o *Tweetstats* ferramenta que ajuda a rastrear e medir parâmetros exclusivos como alcance do tweet por dia, taxa do crescimento de seguidores, quantidade de tweets sobre o assunto por hora. Sendo possível inclusive comparar diferentes contas para uma visão geral do progresso dos perfis e de seus concorrentes.

De modo geral, uma ferramenta de análise de sentimento precisa possuir dois métodos base para execução, um de entrada para estabelecer o item a ser analisado, e outro de saída para exportar os dados analisados. Ao escolher um instrumento, conceda as palavras-chave, os termos, e o recorte de período e região desejado, é importante saber que esses dados serão processados por um algoritmo que realizará a análise no *big data*, retornando com uma classificação prévia para que seja feita a execução e tomada de decisão de acordo com a apresentação do resultado.

As redes sociais foram as primeiras plataformas a serem alvo das ferramentas de análise de sentimento, mas estas já são utilizadas em outros nichos. Por exemplo, na

educação, para auxiliar professores e alunos, na política para predição de eleições, na saúde para diagnósticos clínicos entre outros.

Por ter sua importância validada nas mais diversas áreas, muito esforço tem sido feito para uma melhoria do método e da acurácia da análise, para que a máquina se aproxime do parecer de um ser humano, ou chegue a fazer um trabalho ainda melhor. Segundo Moreira *et al.* (2016) grande parte das dificuldades atualmente existentes na análise de sentimento diz respeito à linguagem como a detecção de ironia e subjetividade.

2.2.3 Facetas da Análise de Sentimento

A análise de sentimento vem ganhando força em diversas áreas do conhecimento, característica forte da sua multidisciplinaridade. Por ser uma ferramenta que apresenta diversas formas e faces a serem estudadas, faz-se necessário apresentar um panorama geral de suas facetas. A análise de sentimentos surgiu na tentativa de compreender as interações e expressões que são compartilhadas nas redes sociais sobre pessoas e suas experiências com produtos, serviços, organizações, indivíduos e seus atributos.

Nessa perspectiva, para o perfil da tecnologia o principal objetivo é definir técnicas automáticas capazes de extrair informações subjetivas de textos em linguagem natural a fim de criar conhecimento estruturado que possa ser utilizado por um sistema de apoio ou tomador de decisão (BENEVENUTO; RIBEIRO; ARAÚJO, 2015). O gerenciamento do *big data* é um dos grandes desafios, para isso, a rotulação automática é a aposta no aprendizado de máquina, criando significado para as informações.

A intenção do método é utilizar uma abordagem baseada em algoritmos de aprendizado de máquina que é denominado inteligência artificial, no qual combina a predição de um conjunto de algoritmos e define um valor pretendido como o mais votado entre eles, considerando que o voto entre eles tem peso. Para executar esse processo, são realizados alguns testes com uma base de dados já rotulada com: positivo, negativo e neutro. Para avaliar o desempenho da técnica é utilizado métricas de desempenho: acurácia, precisão, recall, pontuação e erro. A base na qual a máquina será testada é classificada usando os algoritmos e analisados os resultados de acordo com as métricas propostas, e como se intenta que seja o resultado de cada caso (AGUIAR, 2017).

Para as relações sociais apresentou-se como um novo jeito de interação. As redes sociais são a maior revolução da comunicação social (BENEVENUTO; RIBEIRO; ARAÚJO, 2015), nela se representa de diversas maneiras as interações, intenções e relações entre os indivíduos que a utilizam. São nelas que as pessoas passaram a discutir sobre tudo, desde opiniões políticas, religiosas à experiências com marcas, produtos, serviços e lugares. A Análise de Sentimento é o processo de explorar dados à procura de padrões consistentes. Nas redes sociais, esses padrões descrevem como os indivíduos interagem ou as características/padrões que dão origem às classificações (SILVA; SILVA, 2019).

Isso significa que até mesmo o ato de relacionar-se com outros seres humanos, perante essa sociedade da informação onde toda interação é proposta por mídias sociais, torna-se um ato que sobressai ao intencionado. O protocolo de monitoramento não é só um termo cabível às organizações, mas também às relações humanas, uma vez que se espera ver e participar da vida do outro interagente através de suas avaliações e compartilhamentos nas redes sociais.

Já como objeto de gestão o propósito da análise de sentimento é criar um conhecimento estruturado que possa auxiliar na tomada de decisão. Devido a sua importância para as empresas e para a sociedade, esse tema não tem ficado restrito apenas a uma área da computação e tem atingido outras áreas tais como psicologia e ciências sociais. Nos últimos anos, as atividades industriais que envolvem análise de sentimentos têm apresentado significativo crescimento sendo objeto de desenvolvimento em grandes empresas (BENEVENUTO; RIBEIRO; ARAÚJO, 2015).

Ouvir a opinião do cliente se tornou essencial para o crescimento estratégico das organizações. Isso porque a voz da sua experiência é potencializada pelos compartilhamentos nas redes sociais. Análise de sentimento é a ferramenta de apoio para o processo de tomada de decisão. Por ser de natureza dinâmica, o *big data* pode expressar diferentes aspectos e apresentar mudanças na resposta de dados, é um fenômeno contínuo.

E por fim, como método informacional, a análise de sentimentos se tornou uma ferramenta muito importante não só para a compreensão das informações que são transmitidas em redes sociais, mas em diferentes âmbitos. Informação, em seu conceito

mais puro e simples, é a mensagem entre emissor e receptor. Porém, segundo Capurro e Hjørland (2007) o conceito de informação na sociedade contemporânea desempenha um papel fundamental para o desenvolvimento e disseminação de conhecimento, e seus impactos globais nos transformaram na sociedade da informação.

2.3 Máquinas Morais

Máquinas Morais, também conhecido como agentes éticos artificiais, é um conceito que gera dilemas e capta respostas com relação às decisões para seu próprio aprendizado, para que a máquina, se utilizando de inteligência artificial, tome suas próprias decisões em situações. Considerando esse contexto, se passou a observar diferentes tipos de aprendizagens de máquina para diferentes países, de acordo com as soluções encontradas em cada senso ético de diferentes sociedades. O senso, as vivências, a intuição, valores sutilmente formados a partir de viés cultural, são um conjunto de fatores úteis para a determinação de um julgamento, fazendo assim com que máquinas morais sejam relevantes, mas não infalíveis.

Para Chiappin e Laeister (2016) a ciência da moral usa o conhecimento do seu objeto de estudo, no caso, o homem e suas ações, para desenvolver máquinas morais com o objetivo de propiciar um ambiente capaz de proporcionar e promover a cooperação e a paz entre os indivíduos. Assim possibilitando à máquina o papel de identificar fatores que valorizamos para que ela tome decisões éticas.

Em três estados do Estados Unidos: Califórnia, Flórida e Nevada, os carros sem motoristas da Google já são permitidos. E em algum momento, muito provavelmente esses carros serão capazes de dirigir melhor do que qualquer ser humano, sem distrações, sem embriaguez e sem miopia, com maior obediência às leis de trânsito e com melhores reflexos. Com o aprimoramento dessa tecnologia, dentro de alguns anos, seria até imoral permitir que um ser humano dirija qualquer carro, pondo em risco a sua vida e a de outros à sua volta. Mas, em algum momento, esse sistema perfeito também precisará tomar decisões éticas em casos de priorizar a vida de seus passageiros em detrimento de pedestres.

Um outro grande dilema de máquinas morais são os soldados robôs, que não apenas são mais fortes, mais rápidos e mais confiáveis, mas totalmente desprovidos de medo, sono ou ansiedade, porém, seriam completamente livres do sentimento de compaixão ou empatia por um ser humano. Ainda assim, qual pai seria capaz de enviar seu filho ao invés dos robôs para uma guerra? Essas são reflexões trazidas por Marcus (2012) que trazem a tona o tipo de “bem-estar ético social” uma vez que a própria sociedade tem uma moral duvidosa que durante anos legalizou escravidão e segregação.

No momento atual se compreende que a criação de novas máquinas é fomentado para a busca do bem estar da sociedade e da diminuição dos impactos negativos trazidos pelo progresso desenfreado promovido pela humanidade. Porém, compreende-se a dificuldade, ou até mesmo a impossibilidade, de programar tais máquinas sem observar um viés ético ou moral baseado nos valores de quem o constrói.

Há uma extensa literatura sobre o mito da neutralidade nas mais diversas áreas de conhecimento. Em grande parte, nesta teoria se defende a ideia que é quase impossível existir algum julgamento ou tomada de decisão neutra. Chander e Krishnamuthy (2018) desenvolvem esta teoria considerando a realidade das plataformas de software e dos algoritmos que fazem parte do nosso dia a dia. A partir do raciocínios dos autores podemos destacar a inteligência artificial, onde a partir do processamento de um conjunto de dados a máquina irá “proferir” seu próprio veredito.

Em muitos casos, os sistemas baseados em IA formularão milhares ou milhões de tentativas para apresentar a solução para um determinado problema. E isso significa que esta irá errar, provavelmente muito mais vezes do que acertar, até que ela aprenda a partir dos exemplos, e passe a descartar todas as soluções baseadas em casos desfavoráveis, e passe a tomar decisões melhores se tornando proficiente naquele tipo de problema. Este foi o caso da *IA Deep Blue* que ganhou uma partida de xadrez para Garry Kasparov, campeão mundial de xadrez em 1985 e também considerado o maior jogador de todos os tempos.

Um caso concreto aconteceu com a *IA Compas*, que prevê o índice de reincidência em crimes. O algoritmo foi acusado de ser racista por ter um alto índice de falso positivo

para pessoas negras, e um baixo índice de falso negativo para pessoas brancas³. Aqui no Brasil ficou conhecido o caso envolvendo o algoritmo de reposicionamento de fotos do Twitter, que priorizava mostrar a imagem de pessoas brancas ao de rostos negros. Vários usuários passaram a testar a ferramenta e o resultado era sempre o mesmo. Em resposta, o Twitter disse que testou vieses raciais e de gênero durante o desenvolvimento do algoritmo, mas nenhuma evidência de preconceito foi demonstrada pelo modelo, mas pediu desculpas pelo ocorrido e garantiu que iria analisar o que estava acontecendo, inclusive abrindo o código para que outros pudessem revisar⁴.

Em todos esses casos surge a dúvida sobre quem deve ser responsabilizado por tal viés. A máquina, o programador, a empresa criadora, ou até mesmo o usuário do produto. Acontece que quando se trata do aprendizado de máquinas, os dados disponíveis para o processamento do algoritmo, isso inclui o histórico de uso do utilizador, são elementos importantes para o resultado final. No caso, existe a possibilidade da máquina apresentar um viés preconceituoso de acordo com quais dados estão sendo fornecidos para seu aprendizado ou calibração. Mas, será que as informações foram inseridas de maneira aleatória? Ou vivemos em uma sociedade que realmente acredita que pessoas negras estão mais propensas a reincidência no crime? Ou pior, uma sociedade que sempre vai priorizar um rosto branco ao de uma pessoa negra.

Considerando esta realidade, já surgiram problemas relativos a quando tais máquinas inteligentes passam a decidir sobre questões ligadas à indivíduos e tomam decisões questionáveis. Estas podem ser acusadas de racismo, machismo, xenofobia, dentre outros. Então podemos nos fazer o seguinte questionamento: Pode uma máquina, livre do pré-julgamento de seu criador, ter valores morais?

A princípio, o aprendizado das máquinas inteligentes costumam ser caixas pretas, do termo em inglês *Artificial Intelligence Black Box*, não se mostrando claro, nem aos próprios programadores, como foi que ela processou e inferiu suas decisões, baseada apenas nos dados que lhe foram fornecidos. No entanto, máquinas inteligentes podem estar constantemente aprendendo e mudando sua percepção, ou processo de decisão.

³ <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

⁴ <https://www.uol.com.br/tilt/noticias/redacao/2020/09/21/algoritmo-racista-twitter-pode-estar-escondendo-rostos-negros-nos-posts.htm>

E assim, mesmo que tenha tomado decisões tendenciosas, pode ser impossível identificar se essa parcialidade foi processada pela interação com o comportamento social humano ou se ela foi treinada para tal.

A ideia de criar máquinas morais que tragam o bem estar à sociedade não é ruim. Ruim é a ideia de moral que a sociedade tem. É preciso ensinar às máquinas um moral sem viés limitado. Para construir uma consciência ética satisfatória é necessário um esforço coordenado de diferentes áreas do conhecimento científico que varie entre filósofos, cientistas da informação e da computação, legistas e muitos outros.

É desafiador regulamentar máquinas inteligentes e transformá-las em máquinas morais. Entre os grandes problemas aparentes, um é que as áreas que deveriam estar envolvidas nesse processo não costumam se comunicar. E outro é que, ao que tudo indica, embora os aspectos morais pesem na normatização desses serviços, a maioria das organizações tem como real intenção o comportamento de compra do consumidor.

Ou seja, a partir disso se conclui que para que o resultado gerado pela máquina moral seja mais justo, é preciso construir um novo modelo de sociedade. Se a polícia continuamente prende mais negros, e se são os influenciadores brancos que recebem o maior número de curtidas, a máquina vai continuar com seu viés preconceituoso pois o algoritmo está sendo alimentado dessa forma, como se a cor da pele fosse algo realmente relevante nesse aspecto. Este é o reflexo de uma sociedade problemática.

Na Análise de Sentimento o aprendizado de máquina é utilizado como motor da máquina que processa os dados para inferir o sentimento das pessoas nas mais diversas situações. Dessa forma, a AS é um técnica usada para compreender atitudes e pensamentos humanos expressos nas redes sociais e a partir daí estas tomam decisões categorizando, ou mesmo julgando, os indivíduos. Então é a partir daí que podemos encaixar a Análise de Sentimento no contexto das máquinas morais.

Embora tudo isso pareça ainda muito utópico, inteligências artificiais já estão sendo treinadas de forma significativa em ambientes de vida real. Máquinas morais já vêm revolucionando ambientes mercadológicos e realizando tarefas ao lado de seres humanos, prometendo ainda muito mais vantagem competitiva, sociais e econômicas com a automatização, além de, indiscutivelmente, poupar milhões de vidas ou esforço

físico. E isso marca uma mudança nas escolhas e nas consequências que se deseja alcançar.

Por exemplo, na indústria do turismo a experiência do cliente é crucial em um contexto de rica retroalimentação para o seu desenvolvimento. Os setores que envolvem esta indústria se adaptaram com relação à evolução das redes sociais e intensificaram o significado de estar presente nelas. A maioria dos serviços turísticos precisa estar disponível na Internet através de sites para reservas online. Além disso, as imagens precisam ser atraentes para serem amplamente divulgadas. Por isso, não surpreende que o turismo tenha sido reconhecido como o setor número um em termos de engajamento online (MACK; BLOSE; PAN, 2008).

Turistas agora acessam diferentes fontes de informação a qualquer momento e em qualquer lugar, geram seus próprios conteúdos e compartilham suas experiências. Qualquer conteúdo compartilhado nas redes sociais torna-se influente e pode impactar o desempenho do local ou setor. E o volume de dados na internet atingiu um nível que torna o processamento manual quase impossível, sendo necessário uma abordagem automatizada. A análise de sentimento se apresenta como uma solução analítica eficaz e em ascensão para esta questão e uma abordagem eficaz perante o acúmulo de informações (MACK; BLOSE; PAN, 2008).

Trazendo a Análise de Sentimento para o contexto das Máquinas Morais, observamos que as avaliações são baseadas em opiniões, sentimentos, crenças e julgamentos pessoais. Assim, a percepção do consumidor retrata felicidade, frustração, decepção, alegria, entre outros sentimentos. Mas, tecnicamente, A Análise de Sentimento representa um problema de classificação de polaridade que pode ser conceituada como uma classificação binária (positivo ou negativo). Que no contexto do turismo pode referir-se como uma pessoa satisfeita ou insatisfeita com a experiência.

Para essa abordagem é utilizado o aprendizado de máquina para a mineração de dados, separando por termos, localização e data. Os recursos da análise de sentimento são treinados para classificar os termos da amostragem e projetados para automação da classificação das avaliações. Na prática, não se conseguiria extrair e processar com a velocidade necessária o grande volume de dados dos clientes e se tornou necessária uma abordagem baseada em máquinas inteligentes para realizar tal análise. O setor do

turismo apresenta a ponta do iceberg para um paradigma complexo do uso do *big data*, tendo em vista informações com dados especiais como clima, transporte, meio ambiente e eventos especiais, tendo o nível de satisfação moderado por qualquer um desses fatores, conjunto que precisaria ser investigado e ponderado por Máquinas Morais.

Outro contexto do mundo real é o sistema de crédito social chinês, ou, como diria Mambu (2019), a confiabilidade dos cidadãos na implementação de um crédito social mediado por uma inteligência artificial implementado pela China. Esse sistema é importante pelos efeitos sociais de governança e vigilância que a inteligência artificial irá dispor sobre os cidadãos. O primeiro fator a ser observado é o nível de confiabilidade que a população apresenta sobre seu governo, e na China inúmeras pesquisas apresentam níveis estáveis de confiança política entre o público, pela alta taxa do crescimento econômico nos últimos 20 anos, onde a maioria dos cidadãos chineses têm se beneficiado (WANG, 2005).

O sucesso da confiança dos cidadãos chineses em seu governo é baseado no desempenho institucional, e o apoio dos residentes do país é indispensável para a legitimação e implementação de um sistema de crédito social obrigatório. A fim de criar uma “cultura da sinceridade” no qual os cidadãos chineses sejam leais ao estado na prestação de suas contas. De acordo com o governo chinês, o principal objetivo é beneficiar os cidadãos confiáveis e honestos e restringir as operações daqueles que não são (YOKOTA; SEYHAN, 2018).

Os objetivos do sistema de crédito social são aprimorar as leis, regulamentos e concessões de crédito, produzindo uma plataforma pública compartilhada e construindo mercados de crédito de serviço, além de criar um mecanismo de recompensa (ou punição) para fazer cumprir a disciplina. Todos os cidadãos chineses e empresas privadas serão avaliados por esse sistema, desde suas atividades comerciais até seu comportamento social, e se estão em conformidade com suas posses. É possível observar que o sistema de crédito chinês armazena uma grande quantidade de dados, e o big data precisa ser estruturado de forma automatizada pois seria impossível ser feito por mão de obra humana dado ao diversificado volume e velocidade com que chegam essas informações.

Máquinas Morais treinadas em um sistema como esse seriam capazes de descobrir padrões ocultos nas atividades financeiras dos interagentes. E é através do uso e das redes sociais que o governo chinês controla esse fluxo de informações para sua governança. São dados coletados através da vigilância. Em resumo, o sistema de crédito social chinês permite e facilita a troca de informações e compartilhamento de dados, resultando em uma rede que controla as informações financeiras de um país inteiro. E manter esse sistema que monitora a conduta dos interagentes em atividade faz com que seus cidadãos sejam recompensados.

Um outro exemplo de sistema de vigilância por máquinas inteligentes que não recebeu tanta aceitação assim foi o do centro de atendimento da *Amazon*. Empresa com uma vasta rede de entrega que inclui mais de 180 depósitos, ou centros de atendimento, que funcionam como um coração bombeando o sangue pelas veias que são suas esteiras e funcionários que rastreiam, embalam, classificam e processam cada pedido para ser entregue na porta dos clientes. Leacher (2019) denuncia que os funcionários do centros de atendimento trabalham sob péssimas condições e são pressionados a cumprir metas excessivamente altas, embalando centenas de caixas por hora sob ameaça de demissão se não o fizerem rápido o suficiente, inclusive sem ter tempo de ir ao banheiro, alguns funcionários passaram a usar garrafas para fazer suas necessidades, pois o banheiro fica muito longe e eles têm medo de serem punidos.

O medo imputado aos trabalhadores da *Amazon* vai além de câmeras nos armazéns, ou da proibição do uso de itens como moletons (para evitar furto), óculos escuros e manuseio de celulares. Foi instaurado um sistema que avalia a taxa de eficiência de cada funcionário do centro de distribuição que pode emitir automaticamente uma carta de aviso prévio ou demissão dado a qualidade do serviço ou improdutividade segundo os padrões desejáveis. A *Amazon* se defendeu afirmando que um supervisor humano pode reverter ou substituir a decisão do sistema, e que só será emitida a carta de demissão automática para o funcionário que recebe seis advertências (também geradas automaticamente) num período de 12 meses. Porém se o sistema identifica que uma pausa para descanso foi muito longa, considerando o padrão intensivo da meta a ser alcançada, o aviso prévio é gerado.

Podemos identificar dois problemas primordiais nesse exemplo. O primeiro é que não existe um mínimo de base de confiabilidade entre o empregador e funcionário, que foi o grande sucesso no caso do sistema de crédito chinês. E o segundo é que quanto mais os trabalhadores se esforçarem para atingir a meta, mais o sistema vai recalculá-la para uma meta maior, e reprimir a improdutividade. O uso da tecnologia da vigilância pela máquina inteligente faz com os funcionários sejam reprimidos e maltratados. Sem que haja uma moral imputada à máquina, como a empatia, que observe se o funcionário está doente ou abalado. Sem que a perversidade do sistema seja avaliada.

Existem diversas consequências com relação ao uso da inteligência artificial, essas consequências podem ser minimizadas se aprimorado o senso de moral das máquinas. Sistemas que operem em espaços públicos precisam se adequar à sociedade pois já estão afetando a vida das pessoas. É preciso destacar o conceito da confiabilidade e principalmente a legitimação de que a máquina moral pode desempenhar um papel com o propósito de colaborar com a humanidade.

Liu e Chen (2017) acreditam que o uso da Análise de Sentimento pode ser benéfico nesse propósito em diversas áreas, inclusive na abordagem e aplicação da lei. Utilizar o *big data* para cruzar e validar dados jurídicos desenvolve a previsão das sentenças, colaborando com o trabalho dos magistrados. Estimular a disponibilização das informações jurídicas na rede e a utilização da abordagem de análise de sentimento com a mineração de dados apresenta uma inovação determinante na utilização de documentos legais.

A aplicação tradicional da análise de sentimento é usada para determinar emoções do interagente em um texto. E o aprendizado de máquina tem sido amplamente utilizado para averiguar sentimentos diversos como felicidade, tristeza, raiva, alegria, representados em forma de texto, *emoticons*, fotos e muito mais. A análise de sentimento agrupa e classifica esses termos baseando-se nas relações estabelecidas. Os documentos jurídicos também possuem seu próprio big data, e a análise de sentimento pode poupar o esforço satisfazendo a correta recuperação e classificação de acordo com a estrutura e abordagem utilizada em julgamentos parecidos. A intenção aqui seria, portanto, fornecer previsões de julgamentos para cada caso.

Mais uma vez a máquina interferindo na vida dos seres humanos, emitir um julgamento justo é crucial para a vida em questão. Devido à falta de ferramentas, capacitação ou assistência, o número de processos acumulados é muito grande. Para superar essa dificuldade a automação pode ser utilizada, porém a imputação de uma moral às máquinas continua inquestionável.

Patil, Chaudhari, Bhavsar e Pawar (2020) avaliam que a pesquisa em análise de sentimento atingiu um avançado nível de maturidade, mas a comunidade científica ainda explora suas possibilidades em diferentes tipos de domínio, quando existem enormes quantidades de dados descritivos disponíveis nas mais diversas áreas. Apesar de ter sido um ferramenta para, a princípio, avaliar revisões de filmes, políticos, produtos e serviços, essa tecnologia vem ganhando força também na área médica e na predição de diagnósticos e prescrições de remédios identificando padrões e comportamentos.

Aprendizado de máquina e processamento de linguagem natural são duas correntes da inteligência artificial, enquanto a análise de sentimento é multidisciplinar e determina computacionalmente atitudes e avaliações do interagente (PATIL *et al.*, 2020). No domínio da medicina, a análise de sentimento é aplicada para detectar o efeito de medicamentos nos pacientes, medição da qualidade de vida, e até determinação de resultado clínico, e etc.

Capturar a experiência emocional das pessoas é importante e pode ser monitorada, que inclusive podem servir para identificar e avaliar problemas psicológicos e emocionais, uma vez que expressos e avaliados com a perspectiva da análise de sentimento.

É grande o interesse e pesquisa relacionados à saúde, condições médicas, tratamento de doenças, sobre hospitais ou planos de saúde. O desafio é fornecer uma análise rápida e coerente. Estima-se que 7% das pesquisas diárias no Google sejam sobre saúde (PATIL *et al.*, 2020). Nesse momento, já apresentando em texto sentimentos relacionados à saúde, estado, condição física, ou tratamento por qual o interagente esteja passando.

Ronand *et al.* (2017) afirma que muitas pessoas que sofrem de doenças psiquiátricas permanecem não tratadas, mesmo que exista tratamento, e se esse problemas fossem identificados logo na fase inicial poderiam ter sido prevenidos. A

maioria dos problemas mentais pode ser identificado na infância com uma coleta de dados eficiente. Garcia-Ceja *et al.* (2018) ousou coletar esses dados em sua pesquisa sobre transtornos mentais como depressão, ansiedade, transtorno bipolar e estresse. Dados provenientes de smartphones, smartwatches, pulseiras fitness com sensores de frequência cardíaca, pressão ambiental, temperatura corporal, e até mesmo a análise de seu conteúdo nas redes sociais, que juntos podem oferecer significativas informações sobre o estado de saúde atual de uma pessoa. O pesquisador apresenta três aspectos sobre esse estudo: associação, detecção e previsão.

Ortigosa, Martín e Carro (2013), também afirmam que a análise de sentimento há algum tempo já deixou de ser utilizada somente no contexto do marketing, onde o uso de informações específicas coletadas e armazenadas para objetivos e-learning adaptados como as preferências e estilos de aprendizagem, estado emocional de alunos e professores, ações anteriores e posteriores, fatores afetivos, entre outros, parecem motivar os alunos no processo de aprendizagem.

No contexto da educação, ser capaz de detectar e gerenciar informações sobre as emoções dos alunos em determinado momento pode contribuir para conhecer suas necessidades e suas aptidões. O ambiente e-learning pode ser adaptado para atender o estado de cada aluno e ainda passar avaliações psicológicas e emocionais para os professores, essencial para um momento em que não se pode haver interação face a face.

Portanto, não se trata mais de discutir a utilidade da análise de sentimento, e sim, a moral imputada à inteligência artificial que passa a ter e controlar todas as nossas informações. O armazenamento de informações, o *big data*, pode e deve ser manipulado através de análise de sentimento. Porém, precisa ser mantido certo padrão de confiabilidade e comprometimento moral no modelo instruído.

2.4 Ciclos de Vida de Análise de Sentimento

Uma vez observada a Análise de Sentimento (AS) como objeto de estudo, entende-se que é importante ressaltar que a Análise de Sentimento possui alguns *modus operandis* próprios que variam de contexto para contexto e que resultam em uma

diversidade de maneiras em como esta pode ser executada. Essa variedade de formas existente na realização da Análise de sentimento pode ser mais facilmente observada a partir de ciclos de vida, onde cada um considera as particularidades e objetivos propostos pelo contexto em questão, bem como, pelas possibilidades dispostas por esta área do conhecimento. Desta forma, é possível explorar tais ciclos a partir das etapas propostas por diferentes autores e assim estabelecer uma espécie de “funcionamento padrão” da Análise de Sentimento.

A partir disso, é possível observar os diversos ciclos de vida de Análise de Sentimento já existentes, pois, a partir deles, se compreende como as diferentes visões a respeito do “como” a AS se materializam na prática. Assim, necessidades diferentes implicam em finalidades, e conseqüentemente atividades distintas que se tornam mais ou menos importantes em contextos específicos. Assim, essa pluralidade de visões nos auxilia a montar um panorama mais amplo das possibilidades de uso, aprimoramento e lições aprendidas sobre a AS.

O que podemos observar é que qualquer indivíduo ou organização pode se utilizar da Análise de Sentimento para observar alguns dos aspectos que estão atrelados ao seu ambiente. Clientes, usuários, interagentes ou qualquer outro ator que possa conceber a sua opinião sobre aquela determinada organização é passível de ser avaliado, mensurado ou classificado a partir do que este expressa a respeito de suas experiências e impressões.

Assim, a partir da representação visual de cada ciclo de vida aqui explanado, se torna possível compreender o seu funcionamento e objetivos, e conseqüentemente como este deve ser aplicado e avaliado a partir de pontos fortes e falhos de cada ciclo. O detalhamento de cada um deles auxilia no entendimento “micro” da Análise de Sentimento e como cada ciclo apresenta seu próprio modo operacional e como este funciona em cada meio específico. E por isso, serão apresentados oito ciclos de vida voltados à análise de sentimento iniciando por aquele proposto por Lima (2015).

Ciclo 1 – (LIMA, 2015)

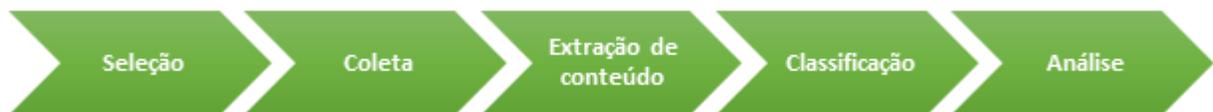
O primeiro ciclo de vida voltado para a Análise de Sentimento aqui apresentado foi proposto por Lima (2015) que em uma pesquisa na área da CI, observa se as

ontologias poderiam contribuir para a execução da Análise de Sentimento em dados não estruturados presentes na *web*. A partir de uma revisão sistemática da literatura, o autor aproxima os campos de ontologia e Análise de Sentimento. No trabalho, o autor afirma que:

Os sistemas para análise de sentimento de dados qualitativos não estruturados disponíveis na *Web* possuem uma arquitetura que normalmente incluem as fases de seleção, coleta, extração de conteúdo, classificação e análise e se beneficiam dos avanços em duas grandes áreas: a ciência da computação e ciência da informação (LIMA, 2015, p. 334).

O autor não detalha os objetivos e funcionamento das atividades existentes no ciclo de vida proposto, e sugere que tal ciclo é generalista, podendo ser utilizado para qualquer tipo de análise de dados qualitativos disponíveis na *web*. Para facilitar o entendimento deste ciclo de vida foi elaborada a Figura 1 que apresenta em um esquema visual as etapas propostas no ciclo de vida.

Figura 1 - Ciclo de Vida de Análise de Sentimento de Lima



Fonte: Baseado em Lima (2015).

Ainda segundo o autor, os dados qualitativos não precisam ser estruturados para que este ciclo tenha eficácia, uma vez que a falta de estrutura presente na *web* faz com que a análise de sentimento possa ocorrer sem que exista uma organização ou classificação prévia dos dados o que resulta em se propor ciclos de vida mais flexíveis. As postagens encontradas nas redes sociais são publicadas de forma espontânea de modo que cada usuário pode expressar sua opinião sem nenhum tipo de censura ou controle. Entretanto, esta liberdade implica na desestruturação dos dados o que dificulta a recuperação e análise, o que justifica a proposição de modelos mais flexíveis.

Embora o autor não faça nenhum detalhamento das atividades propostas no ciclo, é possível encontrar em outros trabalhos acadêmicos o que cada atividade, utilizando a mesma nomenclatura, se refere. Assim, cada etapa deste ciclo possui uma finalidade

bem específica e iniciando pela primeira (**seleção**), é onde ocorre a delimitação do escopo da Análise de Sentimento. Nesta etapa também é especificado o assunto de interesse por parte de quem demanda a AS, quais mídias sociais serão utilizadas, qual é o público alvo a ser avaliado e até mesmo os tipos de postagens publicadas na rede que devem ser analisadas. Em suma, é nesta etapa onde são escolhidas as fontes de informação, usuários e o documento (postagens) alvo da análise.

Em outro trabalho, Teixeira (2015) aponta que esta etapa está mais voltada ao planejamento e desenvolvimento de estratégias para a execução da Análise de Sentimento. Segundo a autora, esta é uma atividade típica de qualquer atividade de descoberta de conhecimento em base de dados (KDD - *Knowledge Discovery Database*), onde a Análise de Sentimento está sedimentada.

Em seguida ocorre a segunda etapa (**coleta**) onde são recuperadas postagens que estejam de acordo com a especificação produzida na etapa de seleção. O recolhimento destes dados pode ocorrer de forma automática ou manual e tudo é armazenado em uma estrutura de dados que podem ir desde planilhas eletrônicas ou até mesmo em grandes armazéns de dados sofisticados. Esta etapa pode envolver uma série de modos diferentes para se recuperar a informação sem que seja executado nenhum tipo de tratamento da mesma.

Teixeira (2015) aponta que é nesta etapa que os dados são encontrados, armazenados e agrupados em uma única base de dados. Segundo a autora, quando se trata de uma base de dados textual, esta pode ser chamada de *Corpus* ou *Corpora*. O que se destaca nesta etapa é a ausência de algum tipo de tratamento prévio tais como limpeza ou normalização de dados, o que reforça a ideia de que a coleta, *per si*, já traz muitos desafios em sua execução, sendo este um problema amplamente observado na área de recuperação da informação.

Na etapa seguinte (**extração de conteúdo**) ocorre o tratamento inicial dos dados para que estes possam vir a ser analisados. Aqui ocorre a retirada de *stopwords* (palavras sem conteúdo semântico tais como preposições e artigos), bem como a observação de dados em branco (*missing values*), dados inesperados (*outliers*) ou qualquer outra anomalia que poderia implicar na diminuição da qualidade dos dados extraídos ou trazer dados que não sejam relevantes para a análise.

Nascimento (2014) aponta que esta nomenclatura provém de algumas técnicas específicas de KDD, que são conhecidas como pré-processamento. Existem uma grande variedade de técnicas envolvidas no tratamento de dados, mas, a autora aponta que a extração de conteúdo indica que uma série de técnicas podem ser utilizadas para tornar os dados brutos obtidos na coleta prontos para serem classificados.

Na quarta etapa (**classificação**) são utilizadas técnicas, manuais ou automatizadas, de análise de sentimento. O objetivo desta etapa pode ser: (i) especificar a polaridade (positivo ou negativo), (ii) se utilizar de algum tipo de classificação (por exemplo, uma escala Likert entre péssimo e ótimo), ou (iii) até mesmo uma pontuação (*score*) em cada postagem extraída. Aqui ocorre o que é conhecido por análise da linguagem, seja está escrita (análise de sentimento textual), seja esta por codificação de categorias (EX: as reações do Facebook: curti, amei, Grrr entre outras), ou até mesmo a análise de outros tipos de mídias como imagens e vídeos.

Tanto Teixeira (2015) como Nascimento (2014) apontam que é nesta atividade onde o conteúdo textual, ou de qualquer outro formato, é compreendido e mensurado pelas técnicas de Análise de Sentimento. Nesta etapa, se identifica a polaridade de uma postagem (positiva, negativa ou neutra) e a intensidade da opinião (ex: Likert entre muito forte e muito fraco). Entretanto, as opiniões ainda não foram colocadas em um contexto (seja coletivo, ou de plano de ação), e para isso é necessário mais uma etapa.

Por fim, a etapa de **análise** é realizada e nela são realizadas avaliações quantitativas sobre as polaridades em cada postagem. Aqui, é criado um amplo panorama sobre o que foi postado, em qual frequência, com qual intensidade, o alcance da mensagem e a viralidade do conteúdo em si. Normalmente são utilizadas técnicas estatísticas para realizar tais análises e o resultado pode indicar uma série de questões sobre a percepção das pessoas sobre o conteúdo na *web*. Essa percepção pode se dar a partir de diversos olhares, por exemplo, é possível analisar a classificação geral do sentimento de uma população (rede) sobre um único objeto, como também, pode ser conduzida uma análise da opinião de uma única pessoa a partir de diversas postagens sobre um mesmo tema.

Segundo Teixeira (2015) esta etapa se caracteriza por adotar uma abordagem mais matemática, seja estatística (probabilística), seja discreta (uso de técnicas de

inteligência artificial). Não são incentivadas a nenhum tipo de análise realizada por pessoas, diferente da etapa de coleta e classificação, onde são considerados o uso de humanos para organizar, classificar ou interpretar tais fatos.

É nesta etapa que se apresentam os problemas ligados a automação da análise de sentimento, pois ainda existe uma discussão, cada vez mais efervescente, sobre a interpretabilidade da Análise de Sentimento (LIU; COCEA; GEGOV, 2016). De modo geral, existem duas questões frequentes sobre a relação entre “máquinas inteligentes” e a interpretação dos dados: (i) o quanto as máquinas são capazes de interpretar o mundo real para fazer a sua análise e (ii) o quanto nós, seres humanos, podemos compreender por que uma inteligência artificial chegou a um determinado resultado. A interpretabilidade foca apenas na segunda questão, e ainda que existam problemas ligados à não compreensão, por parte dos seres humanos, de como estes são classificados por uma máquina. Neste sentido, as máquinas não só podem “errar” em uma análise de sentimento, bem como, os seres humanos podem não perceber/compreender o erro.

Se considerarmos a visão de Lima (2015), este ciclo de vida não tem a intenção de ser uma proposição técnica para aplicação em algum contexto específico, e sim, ser uma representação, sucinta, das macro atividades existentes na Análise de Sentimento e por isto ele apresenta uma conotação muito mais didática e exploratória do que algo mais voltado a assertividade e eficiência. Entretanto, estas características são suficientes para representar algum tipo de funcionamento da análise de sentimento, e ao mesmo tempo, observar como esta está muito ligada aos objetos de estudo da CI e os objetivos de cada etapa bem como seus produtos intermediários.

Em um ciclo de vida voltado à aplicações específicas, seria esperado que existissem mais etapas no processo de Análise de Sentimento, entretanto, talvez para se tornar mais didático, faz sentido que o ciclo de Lima seja assim pois o propósito dela é destacar a ontologia nos artigos em CI e não focar na análise de sentimento.

Ciclo 2 – Devopedia (2020)

O sítio especializado em tecnologia da informação Devopedia⁵, uma espécie de enciclopédia para desenvolvedores de sistemas de informação, também apresenta um ciclo de vida para Análise de Sentimento que possui seis atividades e se apresenta um pouco mais detalhado do que o ciclo anterior. Este ciclo também é generalista não se propondo a atender uma necessidade específica, entretanto, o próprio sítio enfatiza que este compreende se adequa melhor às análises de texto, sendo igualmente didático e exploratório com o foco na compreensão da análise de sentimento. Para facilitar a sua compreensão, o ciclo foi representado na Figura 2.

Figura 2 – Ciclo de Vida de Análise de Sentimento da Devopedia



Fonte: Traduzido de Devopedia (2021)

O processo se inicia pela **Aquisição de dados**, que compreende o processo de coletar os dados que serão documentados ou analisados. Esta é uma das fases mais importantes deste ciclo de vida uma vez que, pelo próprio sítio, é necessário definir um conjunto de dados adequado, ou seja, dados oriundos de fontes confiáveis, que apresentem qualidade aceitável e que sejam relevantes para se atingir os objetivos. Este conjunto de dados serão analisados e classificados, normalmente a partir de alguma técnica de mineração de textos, e se utilizam de processos automatizados para o recolhimento de informações. Diferentemente do Processo de Lima (2015), este não se

⁵ <https://devopedia.org/sentiment-analysis>

inicia com uma etapa de planejamento e curadoria das fontes, chamada por ele de seleção, e sim pela aquisição de dados, chamado por Lima de coleta.

Em seguida ocorre o **Processamento de texto** a partir do processamento de linguagem natural onde são removidos as palavras desnecessárias, palavras repetidas, conectivos, *emoticons*, URLs e outros dados que não são úteis para a Análise de Sentimento. Esta se assemelha a etapa de “extração dos dados” proposta por Lima (2015), o que indica que não há uma padronização da nomenclatura das etapas ligadas a Análise de Sentimento, tais diferenças não serão apresentadas em todas as etapas observadas daqui para a frente, entretanto, cabe a ressalva de que esta falta de padronização é comum. E, para equalizar as visões trazidas por diversos ciclos propostos mais do que olhar a denominação das atividades, se faz necessário compreender qual o objetivo dela é o que nela é realizado.

Seleção e Extração de Recursos: a seleção e a extração de dados dispostos em modelos adequados desempenham um papel fundamental em aumentar a efetividade do modelo utilizado para Análise de Sentimento. Portanto, devem ser escolhidos aqueles mais apropriados para os tipos de documentos que serão analisados, por exemplo, em uma postagem no Twitter (análise do texto), ou o conjunto de indicações (*likes*) no Instagram, ou mesmo um documento de texto corrido? Para cada tipo de dado a ser analisado, existe uma melhor forma de compreendê-lo (modelo), e se faz necessário escolher o mais adequado para se atingir os objetivos. Utilizar técnicas de transformação de dados, ou seja, traduzir os dados da fonte original para o formato que é melhor compreendido por aquele modelo, faz parte daquilo que é chamado em outros ciclos mais adiante de extração. Esta etapa não é explicitada no modelo de Lima (2015), pois se mostra como uma espécie de segunda etapa do processamento de texto, mas, está mais voltada para formatar os dados para o passo seguinte.

Classificação dos sentimentos: Nesta fase, podem ser utilizadas algumas técnicas de classificação de sentimentos para analisar texto. Tais técnicas poderiam ser executadas a partir de técnicas de aprendizado de máquina ou até mesmo pelo discernimento humano em reconhecer palavras chave e detectar a forma como cada um se expressa perante o documento analisado. Esta etapa também está presente no ciclo de Lima (2015).

Detecção de polaridade: Após a classificação inicial dos sentimentos, se faz necessário determinar a polaridade da postagem. O objetivo da detecção de polaridade é decidir se um texto expressa um sentimento positivo, negativo ou neutro e agrupar tais postagens de acordo com cada polaridade. Esta etapa também não é explicitada no modelo de Lima (2015), e se subentende que esta se encontra inserida na etapa de classificação.

Validação e avaliação: Finalmente, na validação e avaliação os resultados obtidos são analisados de modo a se determinar a precisão geral das técnicas usadas para análise de sentimento. Estas ocorrem no momento em que são fornecidas informações para tomada de decisão, elaboração de relatórios, ou acompanhamento da popularidade do produto ou item avaliado. Esta etapa é um pouco diferente da etapa final do modelo de Lima (2015) que é a análise pois enquanto que no modelo de Lima a análise é o momento de sumarizar os resultados, aqui a validação e avaliação serve como uma única etapa onde humanos podem verificar se o resultado da análise de sentimento está coerente, se tornando auditável. Então, compreendemos que a ideia geral é tornar os resultados do processo AS verificáveis.

Este ciclo traz alguns elementos um pouco diferentes daqueles sugeridos por Lima (2015) tais como a transformação de dados e a validação da AS, mas o principal elemento é a concepção do porquê usar ciclos de AS que é compreender o que é informado pelos usuários. Para aproximar este ciclo das ideias da CI, podemos nos basear em Becker (1976) apud Sarasevic (1996), que no começo da Ciência da Informação a classificou como “O estudo pelo qual as pessoas criam, usam e comunicam informações.

O que este ciclo da Devopedia sugere é que a Análise de Sentimento é uma forma de compreender um formato diferente de comunicação de pessoas, mais coletiva, mais aberta e menos estruturada sobre algum documento. A AS pode ser compreendida como uma forma diferente de analisar informação, mas esta é observada ainda em uma visão majoritariamente tecnicista, que pouco considera fenômenos sociais emergentes que acontecem na rede.

Ciclo 3 – Assis (2014)

Este ciclo, proposto por Assis (2014), foi criado especificamente para classificar a postura em debates não ideológicos claramente tendo como base a identificação dos padrões linguísticos e ao analisar trechos textuais dispostos em redes sociais, este tenta classificar determinadas postagens. Este ciclo pode ser aplicável em outros domínios por causa da extensibilidade do padrão ao utilizar o processamento de linguagem natural. E, assim como os outros, este apenas aponta os passos de maior abstração sobre como realizar a análise de sentimento, também se mostrando mais didático e voltado a compreensão, do que, ser uma especificação técnica. Uma representação visual deste ciclo de vida pode ser vista na Figura 3, e consiste em quatro etapas que são:

Figura 3 - Ciclo de Vida de Análise de Sentimento de Assis



Fonte: Baseado em Assis (2014).

1 - Detecção de subjetividade: Nesta ocorre a identificação das sentenças subjetivas, que tratam de “pensamentos, opiniões, e percepções do personagem na narrativa” presentes no texto. Essa etapa é fundamental para o processo de AS, pois as sentenças subjetivas apresentam de forma explícita os sentimentos, crenças, e emoções sobre um determinado objeto. Dessa forma, aqui é possível determinar o tipo de informação ou opinião que o autor da postagem pretende passar, dependendo da sua narrativa, através das sentenças com base na orientação dos aspectos pré-definidos socialmente, adjetivos e atributos detectados. Esta etapa ocorre na intenção de sanar uma problemática relativa à recuperação das palavras chave uma vez que, nem todas as sentenças recuperadas irão evidenciar uma opinião ou sentimento sobre o tópico que se deseja avaliar. Assim, pode se fazer necessário o um tratamento das informações para que o *corpus* base para avaliação seja relevante. Destaca-se que esta etapa também é observada em outros ciclos que ainda serão aqui explicitados.

A análise de sentimento considera aqui um novo desafio na pesquisa relativa à recuperação de informações sobre a estrutura social no contexto da detecção de subjetividade. Ao estudar a opinião, sentimento, ou até comportamento do interagente faz-se necessário conhecer a estrutura na qual ele está inserido. Esse cuidado em compreender o contexto é explicitado por Marteleto (2001) onde esta afirma que a unidade de análise de dos participantes das redes sociais não são atributos individuais como idade e gênero, mas o conjunto de relações que os indivíduos estabelecem através de suas interações uns com os outros, essa estrutura é o fator preponderante nas escolhas, os comportamentos, opiniões e sentimentos dos interagentes.

Na **Extração de atributos**, são identificados no texto os aspectos ou características do objeto sob análise. Normalmente a palavra extração é utilizada para remover algo da sentença, mas nesse caso seria também uma espécie de destacamento dos atributos. Ao remover as sentenças das palavras irrelevantes, e destacar as palavras relevantes, se torna melhor definindo padrões de relevância para treinamento da máquina. Em outros ciclos observa-se que a detecção de subjetividade e extração de conteúdo andam acompanhadas, mas, normalmente não se apresentam em etapas separadas, porque esta também pode ser considerada como um pré-processamento de informações.

Este aspecto da Análise de sentimento também encontra respaldo na CI pois para Coll-Vinent e Bernal Cruz (1990, p. 108) o objetivo da análise documental reside em:

extrair os elementos informativos de um documento original a fim de expressar seu conteúdo de forma abreviada, resultando na conversão de um documento primário em documento secundário.

Trata-se cada postagem como um documento a ser analisado, e para os autores, esses elementos informativos darão respaldo à possíveis operações de armazenamento, condensação e indexação, reiterando a ideia desta etapa ocorrendo dois momentos conjugados: (i) a extração de atributos, que seriam os de elementos informativos, e em seguida (ii) da sua representação abreviada.

Em seguida vem a **Classificação de sentimento/polaridade** onde se busca determinar a polaridade das opiniões no texto, podendo ser positiva, negativa ou neutra.

Esta etapa está mais voltada para uma marcação semântica das polaridades e não se difere do que é preconizado pelo Ciclo da Devopedia.

Apresentação dos resultados: Esta etapa é responsável por explicitar os dados que serão exibidos, de forma simples e clara, como resultados finais para o usuário final. As informações obtidas a partir da classificação dos comentários podem ser apresentadas em diversos formatos, desde textos que podem descrever os resultados da análise, assim como gráficos que facilitam a comparação entre dois ou mais produtos do mesmo domínio.

A apresentação de resultados se compõe na forma de identificar e demonstrar a significância e a posição em que se encontra o objeto ou tópico avaliado perante os seus usuários. Não se tratando de uma forma dura ou determinista, mas que traz consigo a ideia de que esta visualização da informação serve como apoio à tomada de decisão. Assim, a análise não se constitui em um fim, ela é o próprio meio para aperfeiçoar-se.

Vale ressaltar que o autor afirma que essas são as quatro etapas principais, não excluindo a possibilidade de haver outras etapas. Este ciclo que se inicia na detecção de subjetividade, o que significa que já houve algum tipo de recuperação da informação na fonte e armazenamento de dados. Entende-se que a intenção do autor é focar no fragmento (documento, frase ou postagem) trabalhando apenas com ele isoladamente, para realizar a Análise de Sentimento.

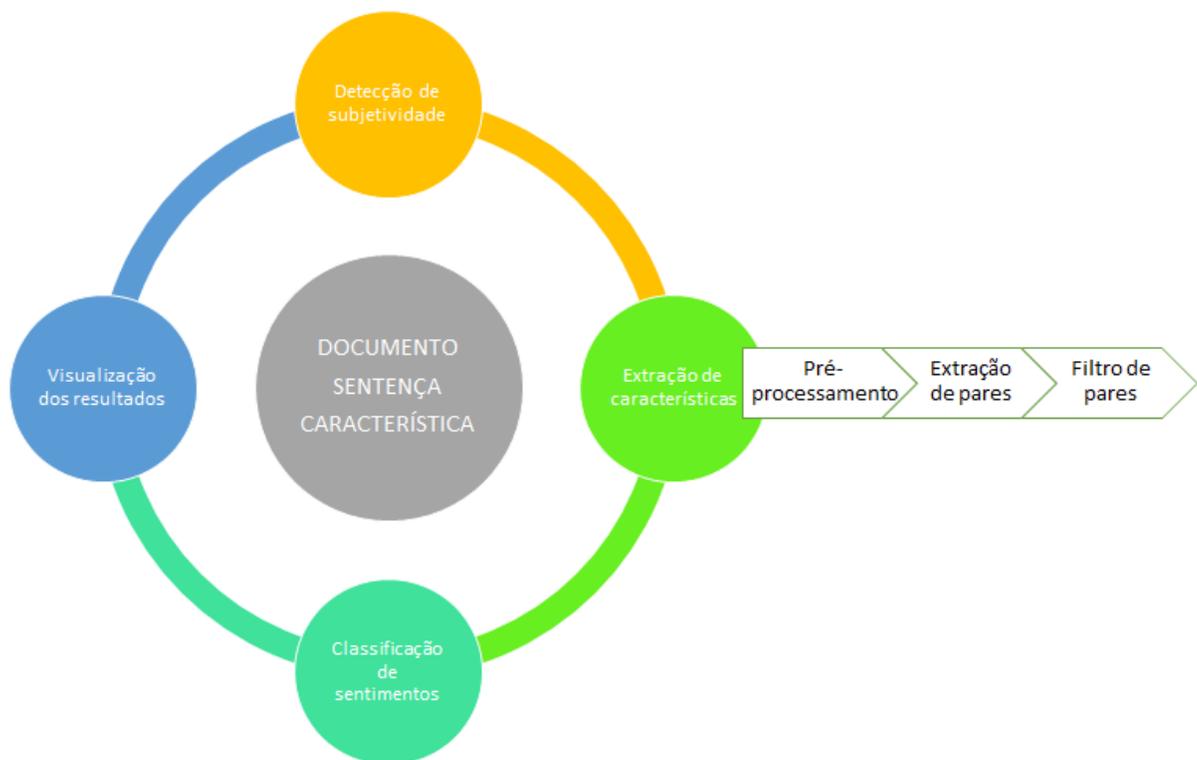
Ciclo 4 – Silva, Lima e Barros (2018)

Neste ciclo de vida para Análise de Sentimento são apresentadas duas etapas principais: (i) A extração de características, que busca identificar os aspectos sobre os quais o texto trata; e (ii) Classificação, que atribui uma polaridade, geralmente positiva ou negativa, a cada atributo com base nas palavras opinativas encontradas nas postagens. Em geral, a Análise de Sentimento pressupõe polaridades pré-definidas para as palavras opinativas, que são facilmente identificadas, como por exemplo: bom, ruim, mal, agradável dentre outras.

Contudo, os autores trazem uma ideia de que palavras opinativas podem ter uma conotação positiva em um contexto, e negativa em outro observando a palavra “gelada”

nas sentenças “cerveja gelada” e “pizza gelada” o adjetivo gelado possui significados distintos. Enquanto que a etapa extração de características é responsável por identificar características válidas a partir de um grupo representativo para o domínio escolhido, e na etapa de classificação do sentimento onde se atribui uma polaridade a cada “par” extraído. Segundo os autores o processo de AS é bastante complexo, e um sistema completo poderia ser representado em quatro etapas como pode ser observada na Figura 4:

Figura 4 – Ciclo de Vida de Análise de Sentimento de Silva, Lima e Barros



Fonte: Baseado em Silva, Lima e Barros (2018).

Antes de explicar cada uma das etapas vale ressaltar que o autor informa que outro aspecto importante para a análise de sentimento vem da observação do nível de granularidade para a análise da opinião:

- 1- no nível de documento, observando o sentimento global expresso no texto;
- 2- no nível da sentença, classificando a polaridade de cada sentença no texto; e

3- no nível de característica, analisando a polaridade das opiniões sobre características/atributos do objeto (exemplo: ‘teclas pequenas’, ‘bateria resistente’), oferecendo uma visão mais refinada das opiniões.

Análise/detecção de subjetividade – Aqui se detecta se um texto é subjetivo ou objetivo. Para este ciclo, os autores reforçam o objetivo de analisar sentimentos sobre um objeto ou tópico em textos opinativos subjetivos, em contraste com textos objetivos, que contêm apenas fatos. O texto deve expressar opiniões que possam ser classificadas em polaridades: positiva, negativa e neutra. Sendo considerada neutra também quando um mesmo texto traz opiniões positivas e negativas na mesma proporção.

Extração de características – Nesta etapa, ocorre a extração do texto das características do produto ou serviço sob análise que são alvos de comentários e grande exposição na *web*. Nesta etapa, se faz necessário realizar a AS no nível de característica, ou seja, para cada sentença contida em um documento. A etapa de extração de características recebe como entrada um grupo com opiniões de usuários dentro de um domínio particular, e retorna como resultado uma lista característica ou palavra opinativa. Em cada um desses, a palavra opinativa expressa o sentimento do interagente sobre o objeto ou tópico associado. Esta etapa conta se divide em três partes: Pré-processamento, Extração de pares e Filtro de pares. Para este autor, entende-se por “pares” as características e palavras opinativas.

No Pré-processamento um grupo de opiniões com base na linguagem do grupo de domínio escolhido são identificadas como *corpus* da AS, e são realizadas operações de “normalização” sobre os textos, formatando-os para posterior utilização. Assim, podem ocorrer substituições em abreviaturas, contradições e gírias, por alguma linguagem melhor compreendida pela máquina. Na Extração de Pares os textos normalizados são analisados de tal forma que uma lista inicial de características que correspondem a substantivos identificados previamente. Então, as palavras opinativas relacionadas aos substantivos que denotam características são identificadas por meio de regras gramaticais. E por fim, no Filtro de Pares seleciona-se da lista inicial de características apenas aquelas que possuam relevância no domínio escolhido.

Classificação de sentimentos – tem por objetivo determinar a polaridade do texto que, segundo o autor, é a principal etapa na AS, pois é nela onde ocorre a identificação

do sentimento em si. Aqui é executada a classificação no nível de documento, ou seja, postagens que podem possuir várias sentenças. Ainda segundo o autor, os classificadores atuais apresentam ótimos resultados de precisão, porém, requerem um *corpus* documental de treinamento devidamente etiquetado, e conseqüentemente, corretamente classificados. Nesta classificação ao nível de característica, a etiquetagem das informações levantadas é mais trabalhosa, pois é necessário categorizar a polaridade de expressões menores e palavras isoladas.

Visualização dos resultados – Aqui ocorre a apresentação ao usuário do resultado da análise do texto, por meio de gráficos, tabelas ou de resumos em linguagem natural. Esta etapa se mostra comum a alguns os ciclos de vida já observados, mesmo que exista uma variação na nomenclatura, porém apresenta aspectos de apresentação dos resultados obtidos através representações mais amigáveis aos usuários finais. Esta etapa remete à CI no aspecto da representação e organização da informação para uma melhor compreensão de toda a informação que foi compilada para uma mais fácil compreensão ao interagente final. Ainda neste sentido, Lima e Alvares (2012, p. 27) afirmam que:

pelo seu caráter interdisciplinar, a organização do conhecimento é estudada também em outras áreas, como antropologia, computação, filosofia, linguística, psicologia, sociologia, entre outras. Na ciência da informação, é a área de estudos voltada às atividades de organização, representação e recuperação da informação. Dentre seus limites de atuação, tenta responder a como se representa o conhecimento; se as áreas do conhecimento são representadas da mesma maneira; o que pode ser representado; e se tudo pode ser representado.

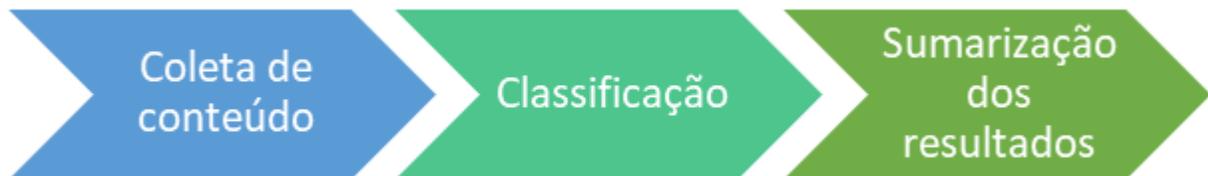
Este ciclo se mostra ser conciso ao ser composto pelas principais áreas abordadas nos demais ciclos de vida da análise de sentimento. Neste ponto, destaca-se que muitos autores estabelecem diferentes nomenclaturas às etapas, porém na descrição apresentam as práticas mais tradicionalmente utilizadas. Destaca-se também a importância da apresentação dos níveis de granularidade para que se possa aplicar a AS mais pontualmente. E, como o autor abrange o uso deste ciclo para textos que contêm opiniões sobre um objeto ou tópico de interesse é possível aplicá-lo em muitos contextos.

Ciclo 5 – Rodrigues *et al.* (2013)

O objetivo principal do ciclo de Rodrigues *et al.* (2013) é observar que o usuário final deve receber algum tipo de relatório contendo o que as pessoas pensam sobre algum item de interesse (produto, pessoa, lugar ou empresa) baseado no conteúdo disponível na Internet sem precisar analisar todas as postagens referente àquele produto.

Este ciclo não foi elaborado, especificamente, para ser executado em algum contexto real, e foram destacados alguns nichos de atuação para aplicação deste modelo, sendo sugerido pelos autores, à adoção por empresas que irão lançar novos produtos no mercado, contribuindo estrategicamente com o marketing. Este processo de mineração de opinião consiste basicamente em três etapas como mostrado na Figura 5.

Figura 5 – Ciclo de Vida de Análise de Rodrigues *et al.*



Fonte: Baseado em Rodrigues *et al.* (2013).

A primeira etapa se dá com a **coleta de conteúdo** onde ocorre uma busca em fontes diversas, tais como sítios na *web*, comentários em redes sociais, anúncios e documentos. Aqui se enfatiza a utilização de técnicas avançadas de busca, o que não fica explícito nos ciclos anteriores que se preocupam apenas com a classificação, visando identificar se o conteúdo encontrado se trata de uma opinião ou não. Os fatos, por si, devem ser descartados, porém opiniões expressas, mesmo que baseadas em fatos, devem ser mantidas. Esta tarefa da busca de conteúdo e classificação em fato ou opinião não é uma atividade trivial, se mostrando ainda hoje como um desafio a ser investigado por pesquisadores e profissionais.

Para a CI a coleta de conteúdo assemelha-se à recuperação da informação, que segundo Mooers (1951) é um campo do conhecimento científico que tem como objetivo solucionar três problemas ligados a explosão informacional: (i) a representação da

informação, (ii) a especificação da busca da informação e (iii) a criação de mecanismo para recuperação. E exatamente estes três problemas, são desafios enfrentados pela Análise de Sentimento nesta etapa de coleta de conteúdo.

A **classificação**, que segundo os autores é considerada a etapa mais importante do processo, tem como objetivo determinar a polaridade ou orientação da opinião ou sentimento. Aqui existe se pretende examinar se uma opinião é positiva, neutra ou ainda negativa. São identificadas as palavras opinativas, que têm um maior peso, e através delas, é possível determinar a polaridade expressa pelo usuário que criou o conteúdo.

Os sistemas de classificação foram um dos primeiros modelos de organização da informação e do conhecimento que se tem registro. Para Piedade (1983), o ser humano está sempre fazendo algum tipo de classificação no seu dia a dia, seja em coisas ou ideias, para que possa entender e organizar melhor suas ideias e o espaço onde vive. O autor afirma também que classificar é o ato de dividir em classes ou grupos, de acordo com suas características diferentes ou semelhantes, seguindo critérios e disposição em grupos pré-determinados, utilizando para isto uma metodologia específica.

A **sumarização** é a etapa voltada à apresentação dos resultados, que pode ocorrer de forma textual, ou gráfica. A apresentação dos resultados de forma textual não se mostra como um produto final tão atrativo, mas ainda assim, pode ser feita através de resumos e extrações de conteúdo da postagem original. Pode ser feita também através de imagens capturadas da Internet. Abordagens como esta ainda estão em desenvolvimento já que se faz necessário relacionar o que foi encontrado na análise de sentimento com uma representação imagética. A maneira mais comum e atrativa de sumarizar os resultados é utilizando gráficos (*dashboards*) onde as informações são contabilizadas e tratadas por cálculos estatísticos, padronizando sua apresentação.

A sumarização dos resultados pode ser observada como a expressa manifestação da representação da informação, que é uma área amplamente difundida na CI, onde existe a preocupação de que o acesso à informação aconteça de maneira mais eficiente/efetiva possível a partir de uma representação que não traga dificuldade ao interagente que queira absorvê-la. Para Herrera (1991) a representação da informação envolve um grande desempenho de análise por parte do profissional da informação, pois

este realiza uma tarefa de análise que abrange identificação, leitura, resumo e indicação transmitindo ao usuário para que este inicie a compreensão.

A partir da visão de Herreira (1991) verifica-se que a representação da informação tem por objetivo facilitar o acesso à informação e, por este motivo, são utilizados mecanismos e estratégias que sintetizam e condensam o conteúdo dessa informação. Assim, o processo de representação na arquivística está diretamente relacionado a qualquer tipo de informação (analógica ou digital) sendo possível ser aplicado a todo o tipo de serviço informacional, inclusive na Análise de Sentimentos em si. Pode-se dizer que os autores foram muito sucintos ao concluir que a análise de sentimento se resume a apenas três etapas. Ao mesmo tempo, foram cuidadosos em detalhar o contexto teórico que envolve o ciclo, mas, não ofereceram nenhuma perspectiva prática para a implementação do método.

De maneira geral este ciclo pode ser aplicado a qualquer nicho por apresentar uma abrangência e generalidade comuns em modelos teóricos. Ao se observar com mais cuidado a elaboração deste, os autores sugerem etapas ocultas que foram observadas em outros ciclos tais como a “extração de conteúdo” ou “detecção de polaridade”, o que evidencia o distanciamento de aplicações práticas.

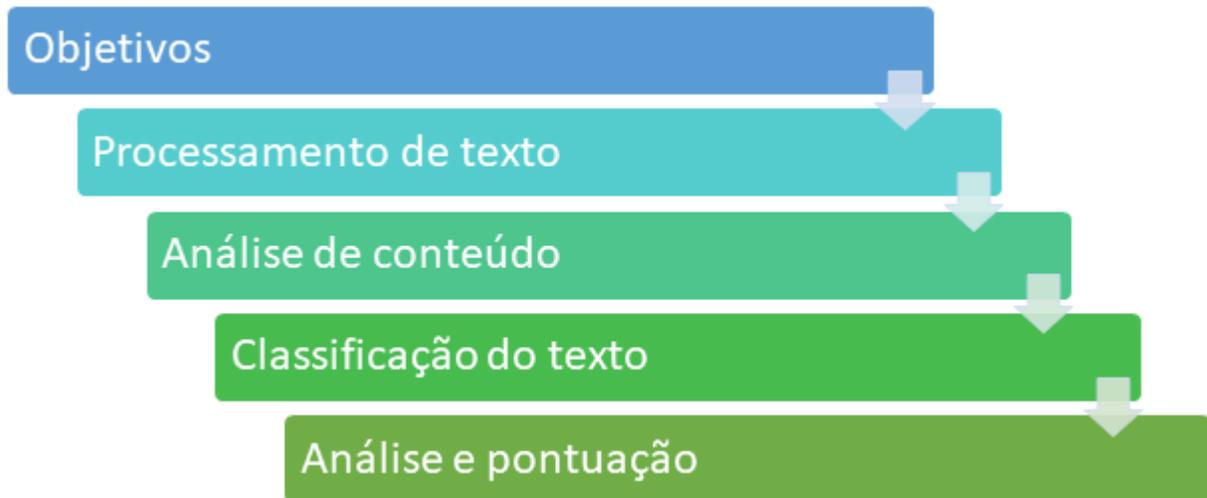
Ciclo 6 – Godsay (2015)

O ciclo de vida proposto por Godsay (2015) foi pensado para a aplicação específica em Inteligências de Negócios (*Business Intelligence*) pela quantificação de menções do produto. Uma vez uma tendência diferente pode ser detectada em cada vez que uma nova análise de sentimento for realizada, já que a quantidade menções sobre o assunto tende a aumentar. Essa medição auxilia na compreensão do sentimento geral das pessoas, permitindo ampliar a percepção, em uma perspectiva social, sobre a marca, possibilitando que empresas rastreiem e detectem sua reputação.

Segundo o autor, este ciclo utiliza o mesmo princípio aplicado na Sociologia, Psicologia e Análise Política, que permite que indivíduos obtenham opinião sobre algo numa escala global, com base em comentários. Originalmente tais modelos seriam uma maneira de se quantificar se uma determinada pessoa está feliz ou triste, alegre ou zangada e no contexto de política se pode identificar padrões de ideologia, preconceitos,

aderências e avaliações públicas. Uma visão deste modelo é fornecida na Figura 6 a seguir.

Figura 6 - Ciclo de Vida de Análise de Sentimento de Godsay



Fonte: Baseado em Godsay (2015).

Para este autor, a análise de sentimento fornece uma plataforma de investigação de experiências psicológicas, registradas na *web*, seguindo os passos apresentados abaixo. Na primeira etapa, chamada de objetivos, são determinados os “objetivos” da Análise de Sentimento. Isto inclui determinar a meta da análise de sentimento e o escopo do conteúdo do texto. Esta etapa também está presente em alguns dos ciclos de vida apresentados usando as nomenclaturas planejamento ou delimitação.

Na segunda etapa ocorre o **processamento de texto** que inicialmente envolve a determinação das fontes e informações, ou seja, se os dados estarão sendo levantados através da *web*, de microblogs ou outras mídias. A seguir, os textos encontrados devem ser carregados no sistema de processamento onde palavras indesejadas do texto são excluídas e os símbolos emocionais (*emoticons*) que as pessoas usam em textos ou palavras são organizados, além disso, é observado que para a expressão de sentimentos fortes, caracteres maiúsculos são utilizados. Comparando com ciclos de vidas já apresentados, observa-se que esta etapa funde outras três com funções mais bem definidas, são elas: armazenamento, seleção e extração de atributos.

Na etapa subsequente, chamada de **análise de conteúdo**, as palavras são classificadas com base em sua polaridade. Aqui, é realizada a marcação das partes do discurso onde estas são usadas, também ocorre a classificação das palavras (adjetivo, substantivo, etc.) e a identificação dos termos em si. Esta etapa parece ser o pilar fundamental da análise de sentimento pois, até agora, tem aparecido em todos os ciclos de vida, mesmo naqueles mais sucintos.

Em seguida é realizada a **classificação** do texto analisado onde são encontrados as *stopwords*, entretanto o autor traz um elemento novo a esta atividade ao reforçar que se deve existir a capacidade de identificar as semelhanças ou as ironias. Aqui o autor replicou a etapa que ele mesmo chamou de processamento de texto, resumindo a classificação como o reconhecimento de palavras desinteressantes à operação de análise, que ao nosso olhar, já foi realizada. Em outros ciclos, vemos que a etapa de classificação ocorre antes da análise de conteúdo auxiliando a polarização de sentimentos, mas, Godsay a coloca depois sem nenhum motivo aparente.

No último passo, portanto, o autor aponta a **análise e pontuação** que envolve determinar os sentimentos dos dados recolhidos e quantificá-los, onde a pontuação é o processo no qual a intensidade do sentimento é analisada. Mais uma vez este autor tende a se tornar redundante ao colocar a palavra análise de forma repetida em seu ciclo. Nesta segunda aparição entende-se não como análise de um resultado prévio, mas como a verdadeira análise de sentimento em si, determinando a pontuação de cada frase ou postagem.

O interessante neste ciclo é que o autor inova ao dar exemplos a respeito de como uma técnica de pontuações podem ser utilizadas para cada sentença, que podem variar de -5 a 4. Segue exemplo de pontuação no quadro a seguir:

Quadro 1 - Exemplo de Pontuação

Pontuação	Texto
2	Você é incrível e eu te amo.
- 5	Eu odeio e odeio e odeio. Que raiva. Morra!
4	Estou impressionado e maravilhado, você é inigualável e incomparável, suas conquistas nunca são medíocres.

Fonte: Traduzido de Godsay (2015).

Assim, a pontuação sugerida pelo autor resultaria em uma sumarização resultante, uma forma diferenciada para avaliar ou representar a aceitação da sentença avaliada. Vale ressaltar que também são incluídas as variações de humor representadas por *emoticons*, mas nesse caso, a avaliação tradicional de polaridade (positivo, negativo e neutro) é melhor aplicável.

Ciclo 7 – Symeonidis (2018)

Neste ciclo de vida a Análise de sentimentos está mais voltada para a interpretação de textos, tais como postagens e resenhas, enviados por usuários em plataformas de microblogs, fóruns e comércio eletrônico. Este observa as opiniões que os interagentes têm sobre um produto, serviço, evento, pessoa ou até mesmo ideias. O autor afirma que como um problema de classificação, a Análise de Sentimentos precisa de métricas de avaliação e de precisão. Além disso, o autor afirma que como métricas como macro, micro e pontuações f1-score ponderadas são úteis para problemas de várias classes. A métrica mais apropriada deve ser usada dependendo de cada classe de conjunto de dados. Uma visão do processo é promovida na Figura 7.

Figura 7 – Ciclo de Vida de Análise de Sentimento de Kdnuggets



Fonte: Baseado em Kdnuggets (2018).

A primeira etapa estabelecida pelo autor é **revisar a base de dados**, pressupondo que as informações já foram coletadas e armazenadas. Geralmente, antes mesmo de iniciar a revisão, é realizada a busca nas plataformas que se deseja alcançar de acordo com o perfil do interagente, a fim de melhor definir o escopo da pesquisa, avaliar a viabilidade e obter maior familiaridade com o tema e com a estrutura da base de dados em si.

A busca, portanto, deve ser mais criteriosa, seguindo procedimentos padronizados, possibilitando assim à revisão da base de dados uma recuperação da informação facilitada. Para a Ciência da Informação, a recuperação da informação preocupa-se com a representação, armazenamento, busca e acesso a informações relevantes aos interagentes. Com este olhar é aceitável que esta etapa assuma a primeira posição através da investigação das etapas semelhantes vistas em ciclos anteriores.

Essa preocupação de coletar os dados antes mesmo de usar, também conhecida como abordagem *data-first*, é observada na Ciência da Informação desde seus primórdios. Para Noronha e Maricato (2008) a Ciência da Informação passou a ser determinante para o desenvolvimento econômico e social de qualquer organização, nota-se, a partir da década de 60, um crescente interesse em coletar informações sobre todo o processo das atividades para que estas sejam planejadas, monitoradas e avaliadas.

A seguir vem a etapa de **pré-processamento** onde ocorre a análise textual, seguindo pelo processo de **tokenização**, termo derivado da palavra em inglês *tokenizer* que é o processo de dividir ou reduzir um texto escrito em unidades significativas como palavras ou tópicos que façam representação do texto por completo. Porém o que mais intriga é o autor ter juntado tokenização e remoção de *stopwords* e separado a **transformação** em uma etapa diferentes, quando, já observados em ciclos anteriores a comum existência de atributos das etapas, sem ter feito qualquer explicação da relevância desta separação.

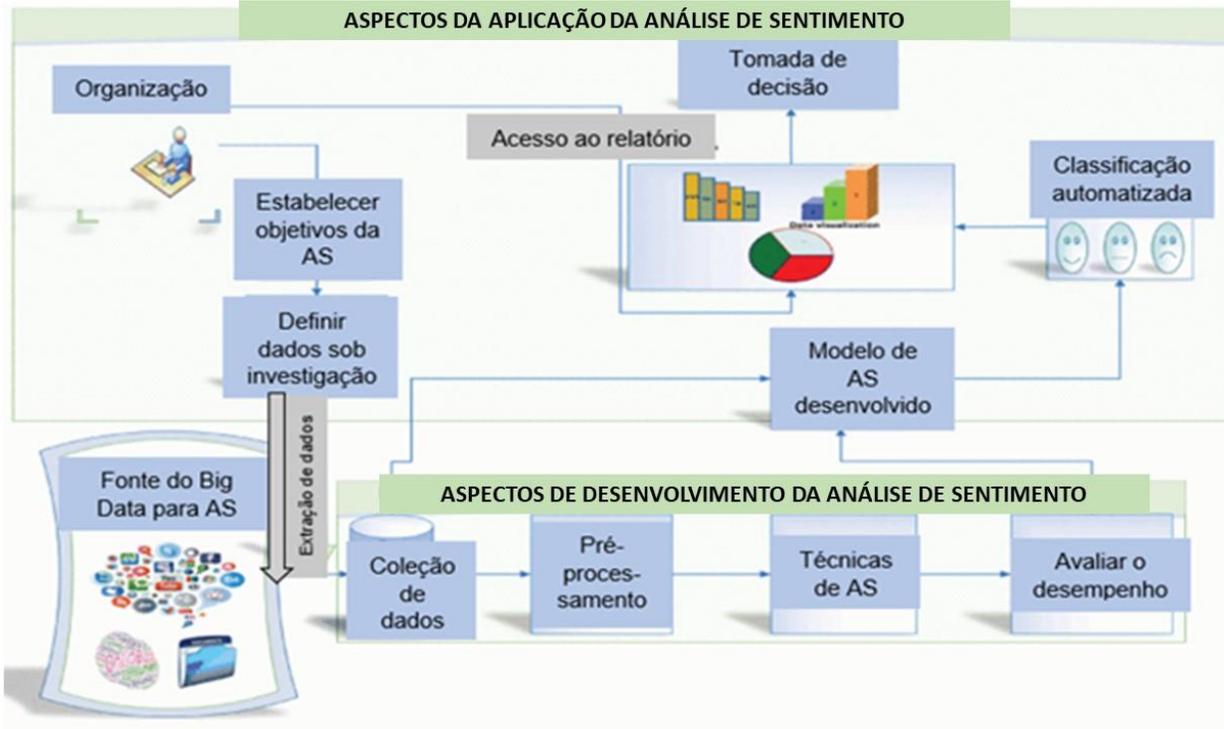
Por fim, **classificação** e **avaliação**, etapas tradicionais aos ciclos já encontrados. O autor, apesar de tentar mostrar como seria uma implementação e uso da análise de sentimento, não especifica nenhuma das etapas que se encontram em seu ciclo de vida, o que deixou muitas pontas soltas para que seja realmente utilizado.

Ciclo 8 – Shayaa *et al.* (2018)

Os autores afirmam que este ciclo de análise de sentimento é utilizado em investigações de popularidade de um produto, serviço, ocasião ou tópico. Tem sido usado em diversas áreas como assistência médica, setor financeiro, esportes, política, hotelaria e turismo e avaliação do comportamento do consumidor. Para este ciclo as declarações de indivíduos são chamadas de sentimentos e estes são compostos por pensamentos, opiniões, atitudes, pontos de vista, julgamentos e expressões.

Estes elementos são reunidos por métodos que separam opiniões e identificam a polaridade das sentenças, agrupando os sentimentos que são expressos em linguagem natural usando expressões subjetivas, o mesmo conceito de sentença subjetiva de Assis (2014). Este ciclo de vida de análise de sentimentos abrange diversas modalidades e plataformas, considerando áudios, vídeos, textos escritos, expressões em *emoticons* e etc. e se mostra apto a capturar a pluralidade de expressões relacionadas à Internet. O ciclo proposto pelos autores é apresentado na Figura 8.

Figura 8 – Processo geral para análise de sentimento



Fonte: Baseado Shayaa et. al (2018).

Na imagem, o primeiro aspecto que chama atenção é a divisão determinada entre *application aspect of sentiment and opinion analysis* (aspectos da aplicação da análise de sentimento e opiniões) e a *development aspect of sentiment and opinion analysis* (aspectos de desenvolvimento da análise de sentimento e opiniões). Na primeira é onde se inicia o processo em uma **companhia ou organização** sentindo a necessidade de acessar e extrair informações dos clientes através de tópicos ou palavras chaves relacionadas, para compreender melhor a sua imagem e auxiliar as tomadas de decisões.

Em seguida se inicia a etapa de **estabelecer o objetivo** da análise de sentimento para o determinado segmento da organização, e então vem a etapa de **definir os dados sob investigação** que considera o tipo de negócio onde será aplicado, o público alvo, o tipo de postagem que se deseja e a plataforma. Por exemplo, se a empresa deseja coletar informações sobre produtos, os dados devem ser extraídos com palavras-chave que representem tais produtos. Isso também envolve explicar a finalidade, a fim de determinar

as palavras-chave relacionadas para que em seguida aconteça a **extração** dessas informações de uma ou várias fontes.

É a partir desse momento que o processo entra no modo *development aspect of sentiment and opinion analysis* (aspecto de desenvolvimento da análise de sentimentos e opiniões), que aparenta ser o lado mais técnico da AS. Aqui, o ponto de partida se dá coleta de dados em alguma estrutura *big data*, a partir de fontes de informação na internet como por exemplo as mídias sociais.

O passo seguinte é a aplicação do **pré-processamento** de dados para remover “ruídos” e conteúdo não relacionados ao objetivo da AS a fim de minerar essas enormes massas de dados e obter informações sobre a empresa/produto no intuito de auxiliá-las nas tomadas de decisões.

Em seguida, são utilizadas **técnicas da análise de sentimento e opinião**, que são baseadas em aprendizagem de máquina, e normalmente ocorre algum tipo de transformação léxica do texto. Entretanto, tais requisitos devem ser elaborados e estabelecidos nesta etapa. Para então ser **avaliado o desempenho** do modelo. Assim termina o módulo de desenvolvimento e o ciclo volta para a área de aplicação onde o **sistema modelo** desenvolvido é posto em prática, para então acontecer a **classificação automatizada** dos dados recuperados separando as postagens em positivo, negativo e neutro, reconhecidos neste ciclo com "emojis" feliz, neutro e triste. E por fim entrega-se o **relatório** final composto por visualização gráfica apto a fornecer o apoio à decisão previamente solicitada pela organização.

Este ciclo é auto explicativo e visualmente atraente, talvez por isso os autores não se detiveram em descrever suas atividades com maiores detalhes. Em termos de aplicação porém, demonstrou-se ser o mais aprofundado pois apresentou aspectos em aspectos claros para o público em geral que é uma tarefa interdisciplinar e requer auxílio externo para desenvolvimento de sistemas e construção de modelos aplicáveis a diferentes ramos de negócio, e não lançou uma abordagem generalizada como se tivesse que ser igual para todos, respeitando assim a necessidade de cada público.

3 METODOLOGIA

Para esta pesquisa foi utilizado um conjunto de métodos que trazem evidências complementares acerca do fenômeno estudado. A variedade de manifestações observadas no mundo real propicia a criação de um quadro mais completo que envolve a execução da análise de sentimento, realizada em grande parte, por atores não humanos, e quais são os impactos sociais deste *modus operandi*, partindo de uma visão informacional. Neste sentido, esta pesquisa apresenta, em etapas distintas, o uso de métodos distintos utilizando a abordagem qualitativa (MICHEL, 2009).

Esta pesquisa se caracteriza como qualitativa no sentido trazido por Michel (2009) por existir uma “relação dinâmica, particular e temporal entre o pesquisador e o objeto de estudo”. Pois, se faz necessária a interpretação dos fatos dentro de um contexto já que o “ambiente de vida real é a fonte direta para a obtenção de dados” tendo o entendimento do pesquisador, com o máximo de imparcialidade, como chave para significar as respostas.

Segundo a mesma autora as pesquisas qualitativas são fundamentadas na discussão e correlação de dados interpessoais, e na coparticipação das situações dos informantes analisados a partir da significação que estes dão aos seus atos. Neste tipo de pesquisa a análise detalhada do fenômeno e a interpretação das evidências vão além da frieza das quantificações e da descontextualização do ambiente. O caráter qualitativo da pesquisa se dá pela interpretação das reações/comportamentos humanos diante das ações iniciadas por eles.

Quanto aos meios, esta pesquisa também apresenta uma diversidade de facetas. Iniciando pela pesquisa bibliográfica, que segundo Gil (2008, p. 45) “auxilia no aprimoramento das ideias, descobertas de intuições e construções de hipóteses a respeito de um determinado problema”. Neste trabalho, conforme a classificação de Michel (2009), foi realizada uma revisão de bibliografia que, segundo a autora, visa arregimentar informações e entender mais detalhadamente o assunto para auxiliar na proposição da pesquisa, definição de problemas e objetivos.

Além disso, as pesquisas bibliográficas podem possuir como objetivo verificar o estágio teórico em que um assunto se encontra no momento atual com o propósito de

levantar novas abordagens, visões, aplicações e atualizações a partir de material já publicado com o intuito de colocar o pesquisador em contato direto com todo material já escrito sobre o assunto da pesquisa (PRODANOV, 2013).

Desse modo, o primeiro passo desta pesquisa foi a realização de um levantamento acerca da literatura existente referente à Análise de Sentimento e seus ciclos de vida. Posteriormente foi pesquisada a literatura que trata sobre a relação entre a Análise de Sentimento e as Máquinas Morais.

Não foi realizado nenhum tipo de sistematização nesta busca pela bibliografia utilizada, sendo esta inicialmente concebida na BRAPCI⁶, verificando-se, entretanto, que não havia menções sobre questões sociais da análise de sentimento. Desta forma, a busca passou a ser realizada na base de dados Google Acadêmico onde foram encontrados alguns trabalhos que contribuíram para a pesquisa em maior ou menor grau. Esta pesquisa bibliográfica resultou nas sessões do referencial teórico, fundamentou as análises sobre as relações da AS com as Máquinas Morais e serviu de insumo para a elaboração do ciclo de vida unificado.

Ainda quanto aos meios, esta pesquisa também se caracteriza como pesquisa teórica que segundo Michel (2009) se propõe a discutir e criar teorias existentes. Segundo a mesma autora esta pesquisa se preocupa em montar e desmontar, criticar e reconstruir teorias existentes a partir de críticas e comparações sobre o tema abordado e não implica em confirmação na vida real. Ao se propor a concepção de um ciclo de vida para Análise de Sentimento considerando também aspectos regulatórios, caracterizamos a teoria que é proposta neste tipo de pesquisa.

Quanto aos fins esta pesquisa é considerada uma pesquisa exploratória, posto que o estudo realizado discorre sobre uma questão ainda pouco conhecida/explorada na área da Ciência da Informação, segundo Michel (2009, p. 40) “se caracteriza pela busca, recorrendo a documentos, de uma resposta a uma dúvida, uma lacuna do conhecimento”. Portanto, esta observação se propõe a observar/fazer relações e conexões que o ambiente exerce sobre os fenômenos. Neste tipo de pesquisa não há interferência no ambiente e o objetivo desta é explicar os fatos relacionando-os ao ambiente.

⁶ <https://brapci.inf.br/>

Quanto aos métodos de pesquisa utilizados, novamente se ressalta que vislumbra-se a utilização de diversos métodos para se encontrar evidências que se complementam e trazem um panorama mais completo do problema. O primeiro método a ser considerado para esta pesquisa foi o estudo de caso. O estudo de caso se tornou útil neste trabalho pelo fato dele, de acordo com Michael (2009), consiste na investigação de casos isolados ou de pequenos grupos, com propósito básico de entender fatos e fenômenos sociais. A intenção desta pesquisa é analisar criticamente situações reais, apresentadas em trabalhos acadêmicos, ou sítios de notícias ou sítios especializados em tecnologia, sobre a adoção da Análise de Sentimento em certos contextos, que podem trazer fragilidades sociais, bem como a proposição de um ciclo de vida unificado para AS.

O segundo método adotado é a análise de conteúdo que se compõe como um conjunto de técnicas de análise das comunicações para obter, através de procedimentos sistemáticos e objetivos de descrição do conteúdo das mensagens, indicadores (quantitativos ou não) que possibilitem a inferência de conhecimentos relativos às condições de produção e recepção dessas mensagens (BARDIN, 2006). Suas principais características metodológicas são objetividade, sistematização e inferência, e à respeito das suas modalidades, destacam-se análise lexical, análise de expressão, análise de relações, análise temática e análise de enunciação, sendo a penúltima a mais indicada para investigações qualitativas (GERHARDT; SILVEIRA, 2009).

Objetiva-se que, com o fim da análise, as questões que ligam as máquinas morais e AS estejam elencadas e que seja possível criar um arcabouço teórico de onde se possam iniciar as discussões dentro do âmbito da CI.

4 CICLO UNIFICADO DE ANÁLISE DE SENTIMENTO

Para toda e qualquer informação, criada com o intuito de atender necessidades específicas de diversos atores, existe um ciclo de vida implícito ou explícito que tende a ser dinâmico, passível de automação e que pode ser realizado de forma mais eficiente. Observando os aspectos gerenciais observa-se que se torna cada vez mais importante a gestão destes dados e o alinhamento desta atividade com os objetivos específicos da organização. Entretanto, esta etapa também apresenta uma série de desafios como é relatado por Siebra (2019) onde esta aponta que a demanda informacional atual traz uma dificuldade nítida de gerenciar e tratar informações.

Apesar de observamos propostas genéricas de ciclos de vida de Análise de Sentimento, não se pode afirmar que exista uma forma que se adeque a todos os casos possíveis para a realização e gestão de AS, uma vez que os objetivos, condições técnicas e natureza dos dados que envolvem o tratamento da informação a ser realizado são variados e isto resulta na dificuldade de se referenciar etapas, atividades e até mesmo subprodutos oriundos da análise de sentimento em uma base padrão. Por isto, se faz necessária a elaboração de um processo unificado para análise de sentimento que uniformize nomenclaturas, objetivos, atores e artefatos para que possamos referenciar neste trabalho.

Embora este ciclo aqui proposto tenha como principal função “padronizar”, não será desconsiderada a hipótese de que este possa ser uma primeira versão de um ciclo de vida AS que possa ser referenciado por outras pesquisas por ser o mais genérico, amplo e que abarque o máximo de atividades possível e também por ser apresentado em uma estrutura processual ao utilizar a notação *Business Process Management Notation* - BPMN. Vale observar que este ciclo de vida nasce voltado orientado à análise da faceta informacional conhecida como proposição opinativa encontrada em redes sociais, traçando uma rota para o uso e desenvolvimento da Análise de sentimento de um ponto de vista mais voltado ao objeto informação.

Para a elaboração deste ciclo de vida foram levados em conta todos aqueles observados na seção 2.4 deste trabalho, extraíndo as partes mais relevantes de cada um deles. Do primeiro ciclo (LIMA, 2015) foram extraídas as etapas (i) seleção, (ii) coleta, (iii)

classificação e (iv) análise. Porém, se fez necessário fazer uma reordenação destas etapas uma vez que foi observado que este autor considera a atividade de delimitação do escopo da Análise de Sentimento, que tem como objetivo determinar quais dados e estes onde serão buscados, como parte da etapa (I) seleção, para posteriormente realizar a coleta dos dados.

O que se observa em outros ciclos de vida é que normalmente também se realiza uma espécie de seleção dos dados também após coleta para admitir ou descartar sentenças que tenham sido recuperadas mas que estejam fora do contexto, atividade que este autor prefere chamar de extração de conteúdo. Desta forma, foi decidido pela não utilização deste termo, assim como outros sugeridos por este autor, por ser facilmente confundido com a etapa de coleta e por divergir da maioria dos outros ciclos de vida. A etapa de classificação foi absorvida integralmente em sua essência e nomenclatura. O autor finaliza este ciclo de vida com a etapa de análise, mas, compreende-se que análise não deveria ser a última etapa do processo uma vez que não se entrega ao cliente o conteúdo para análise e sim o resultado da análise em forma documental, o que é observado em outros ciclos de vida.

Para o segundo ciclo, elaborado pela devopedia, é orientação uma questão mais técnica. Nele percebe-se que a etapa aquisição de dados é equivalente à etapa de coleta do ciclo anterior. Entretanto, entende-se que não há como realizar qualquer tipo de coleta antes de ser planejado o que se deve coletar. Por isso, em nosso ciclo unificado, a etapa de coleta (ou aquisição de dados) ocorrerá após a criação do modelo da AS. Este ciclo de vida também usa a nomenclatura seleção após coleta de processamento de texto para eliminar as palavras desnecessárias (*stopwords*), processo este será considerado em nosso ciclo único. A etapa seguinte constante neste ciclo é a seleção e extração de recursos, mas ao observar esta etapa se observa que as atividades lá presentes não se trata de algo voltado para a ciência da informação mas sim para o desenvolvimento e aprimoração do modelo da máquina para aprimorar a inteligência e melhor se tornar mais efetiva na próxima execução.

Na etapa seguinte, chamada de classificação dos sentimentos, onde são delimitadas a serem expressões consideradas, são reconhecidas o que as palavras-chave encontradas representam. A seguir, na etapa de detecção de polaridade, os

sentimentos são classificados, grosseiramente falando, entre positivos e negativos para cada expressão, ou para o conjunto de postagens como um todo. É possível observar que esta é uma etapa presente no primeiro ciclo de vida e que estas poderiam ser consideradas como etapas que possuem o mesmo propósito. Compreende-se que este ciclo considera muitos detalhes técnicos, entretanto, é preferível que as palavras chave sejam analisadas na etapa de polarização, assim, esta não seria analisada de forma isolada, e sim em conjunto com a sentença. No ciclo de vida unificado aqui proposto, esta atividade compreende a etapa de classificação.

O ciclo é finalizado com uma etapa chamada validação e avaliação onde é realizada a análise das conclusões obtidas pela análise de sentimento e elaboração do relatório final. Considera-se que neste momento são executados dois processos diferentes, e por isso, seria mais adequado dividi-lo em duas etapas, uma para a avaliação e outra para apresentação dos resultados, elaborando-se um documento que possa ser facilmente recuperado pelo requerente da análise de sentimento em questão.

O ciclo 3, elaborado por Assis (2014), apresenta uma visão que, compreende-se, inicia em estágios tardios da AS uma vez que este começa pela etapa denominada detecção de subjetividades. Neste momento, são identificados pelo algoritmo sentimentos expressos em sentenças opinativas para assim determinar a impressão que o usuário pretende passar. Ou seja, observa-se que não são aqui consideradas etapas ligadas ao planejamento, delimitação do escopo e nem mesmo a coleta de dados é desconsiderada. Também é observado que o título dessa etapa não apresenta de forma clara o que será realizado, e na própria descrição do ciclo de vida o autor não informa como se chegou em determinados resultados deixando lacunas sobre como elaborar a detecção em si. Acredita-se que ela tem como objetivo fazer um pré-processamento para identificar se a sentença é ou opinativa, se é o ou não relevante para a avaliação do objeto em análise e isso é sim importante e encontrado também em outros ciclos, mas não com esse nome na etapa e não é interessante que inicie o ciclo.

Em seguida ocorre a etapa de Extração de Atributos onde é realizada a separação entre palavras relevantes e as palavras que não são importantes para a análise. Tal etapa já foi previamente discutida e para ciclo de vida unificada será chamada seleção. É nela que são delimitadas a importância de termos e sentenças recuperadas. Posteriormente

observa-se a etapa de classificação, também já discutida em ciclos de vida anteriores e tem como objetivo determinar a polaridade das opiniões entre positivo, negativo e neutro. E por fim, é apresentada a etapa de apresentação de resultados que tem intenção de representar as informações obtidas através da classificação. Este autor, porém, não detalha se houve qualquer tipo de análise dos dados, o que se compreende que há necessidade uma etapa específica para análise dos dados e outra para a apresentação dos resultados.

Para o ciclo de vida 4, proposto por Silva, Lima e Barros (2018) apresenta um foco mais técnico/tecnológico. Nele, os autores também iniciam a AS pela etapa de detecção de subjetividade, uma vez que se faz necessário definir se uma determinada sentença é opinativa ou apenas expressa um fato explicitado pelo interagente. Entretanto, este ciclo não trata de atividades que vem anteriormente tais como as atividades de planejamento e a própria etapa de coleta de dados em si não é abordada por este texto. A segunda atividade Em seguida ocorre a etapa extração de conteúdo que é dividida pelos autores em 3 sub etapas que são: (i) o pré processamento de texto, (ii) a extração de pares e (iii) o filtro de pares que compreendem o aperfeiçoamento da máquina e por isso não serão incluídos no ciclo voltado para a ciência da informação.

A etapa seguinte é a classificação de sentimentos, que tem como objetivo determinar a polarização das mensagens e fazer a classificação. Esta etapa é a única que está presente em todos os ciclos investigados até aqui e se mostra como indispensável ao processo de análise de sentimento. Para finalizar, os autores sugerem a etapa de visualização dos resultados da análise, embora os mesmos não apresentem tantos detalhes sobre como isso pode acontecer. No ciclo unificado aqui proposto serão distintas as etapas de análise e divulgação dos resultados. A única observação a ser feita para este ciclo proposto por Silva, Lima e Barros (2018) é que os autores deste ciclo abordam também o que eles denominam níveis de granularidade que resolvemos não abordar pois isso pode ser realizado na etapa inicial de Delimitação do Conteúdo. Segundo os autores os níveis de granularidade diferenciam-se como “documento” que observa o sentimento global expresso no texto, “sentença” que classifica a polaridade de cada sentença do texto, e “característica” que analisa a polaridade das opiniões sobre as características do objeto. Compreendemos que esses três níveis já estão intrínsecos a

qualquer Análise de Sentimento durante o processo de seleção e análise e não é exatamente uma etapa e provavelmente uma explicação de da complexidade.

No ciclo 5, de Rodrigues *et al.* (2013), os autores foram sucintos ao enxugar a análise de sentimento em apenas três etapas essenciais. No trabalho se observa uma descrição sucinta do ciclo de vida como um todo, mas não se observa recomendações específicas para cada uma destas etapas. A primeira se refere à coleta de conteúdo e é voltada para ser executada em ambientes web. E no texto existe uma maior preocupação em exaltar a importância das ferramentas de recuperação da informação em detrimento de detalhar sobre o que coletar e como coletar.

Na segunda etapa, intitulada de classificação, são observadas as polaridades das sentenças e identificadas as partes opinativas das postagens. Posteriormente, tudo o que foi classificado é dividido em grupos de acordo com os critérios pré-estabelecidos. Esta etapa de classificação apresenta alguns apontamentos inéditos que serão incorporados ao ciclo de vida unificado. Entretanto, deve-se registrar a lacuna de que não há o apontamento dos autores sobre o momento em que momento tais critérios devam ser estabelecidos. Assim entendemos que esta atividade deve ficar na etapa Análise de Sentimento. Por fim, temos a etapa final de sumarização que é o equivalente a apresentação dos resultados em outros ciclos, e que não há neste momento nenhuma inovação em relação ao que já foi observado anteriormente.

Para o ciclo 6, Godsay (2015), temos como primeira etapa aquela intitulada de objetivos, e é ela que compõem a base para a etapa inicial do ciclo de vida unificado que será chamada de delimitação de conteúdo. Compreendemos que é uma boa prática iniciar o processo de análise de sentimento por alguma etapa anterior a coleta de dados em si uma vez que se mostra necessário determinar metas de análise e delimitar com a maior precisão possível o escopo do conteúdo a ser extraído. A segunda etapa do ciclo de Godsay (2015) é intitulada processamento de texto. Aqui o autor é bastante analítico ao assemelhar esta etapa à coleta dos dados realizados em outros ciclos de vida, entretanto, este o complementa com processos de seleção e extração de atributos.

Para as etapas 3 e 4 deste ciclo, percebe-se que o autor utiliza atributos da classificação mas trata essas atividades como análise de conteúdo e seleção, respectivamente, mas a essa junção de atividades ele denomina como análise de

conteúdo. Para o ciclo de vida unificado a ordem vislumbrada é: seleção -> classificação -> análise. E por fim o autor apresenta a etapa análise da pontuação, onde este descreve um método que tem a finalidade de quantificar a intensidade do sentimento expresso na sentença.

Para o ciclo 7, Kdnuggets (2018) sugere seu início a partir da etapa revisar base de dados, onde se leva em consideração que já existe uma plataforma que funcionaria como armazém de dados (datawarehouse) e que esta é constantemente alimentada por um sistema de coleta de opiniões publicadas em mídias sociais. É importante observar que a coleta foi suprimida, o que não é interessante pois faz parte do processo de aquisição de dados. É aceitável dizer também que após a informação ter sido utilizada ela pode ser descartada pois existe uma delimitação de relevância temporal. Não achamos interessante esse meio pois após a utilização da informação ela vira estatística e documento em nossa etapa de análise e não precisa mais ocupar espaço no armazém de dados. Entendemos que o mais propício seria mesmo a coleta ser sempre atualizada onde os dados e informações serão descartados após o uso. Porém, consideramos que esta etapa pode ser facultativa e fica a critério dos gestores uma vez que pode substituir a coleta e também pode ser utilizado para treinamento de máquina.

Nas três etapas seguintes o autor apresenta de forma detalhada o que outros chamam de processamento ou tratamento da informação. A primeira destas etapas é o pré-processamento que é responsável por uma espécie de análise textual que indica as palavras relevantes a serem avaliadas. A seguir é realizada a tokenização, que tem como propósito reduzir a sentença em unidades significativas removendo palavras irrelevantes e atribuindo tokens as palavras relevantes, realizando uma espécie de classificação. A quarta etapa é chamada por ele de transformação e o que se compreende dela é que pode ser necessário colocar as palavras e sentenças rotuladas e classificadas em algum tipo de formato próprio para ser analisada pela máquina. Neste sentido, observa-se que a transformação é uma espécie de tradução do conteúdo produzido nesta etapa de processamento para aquele que será utilizado pela classificação. Por fim, nas duas últimas etapas de classificação e avaliação, ocorrem os processos já explanados em outros ciclos de vida que apresentam a mesma nomenclatura.

No último ciclo de vida, proposto por Shayaa *et al.* (2018), observa-se um esquema bem mais elaborado do que nos demais ciclos. Este é dividido em duas fases e se inicia com Aplicação que tem um caráter mais gerencial e de negócio da análise de sentimento, enquanto que na segunda (Desenvolvimento) é mais voltado ao lado técnico da AS. A ideia de se dividir a AS em fases distintas trazidas por este ciclo nos trouxe a inspiração de apresentar o ciclo de vida unificado aqui proposto em um formato claro e intuitivo para quem está interessado em momentos específicos da AS.

Considerando as etapas, na primeira fase que se inicia com a etapa de organização, já é observado um foco diferente do que foi compilado até aqui nas etapas de planejamento e delimitação do conteúdo, pois esta estabelece objetivos não só técnicos, mas também organizacionais, que são materializadas na definição dos dados de investigação. Em seguida é apresentada uma etapa de extração de dados, considerando uma estrutura de big data como fonte de informação, a qual também será incorporada ao ciclo de vida unificado e lá será chamado de Coleta.

A seguir o ciclo passa para uma fase de Desenvolvimento que aborda aspectos mais técnicos ligados a manutenção de uma base para a coleção de dados. Nela, percebemos que a etapa de coleção de dados envolve disponibilizar os dados necessários à realização da AS.

Esse processo pode ser observado inserido nas atividades do agente de software, lá é feito o pré-processamento, que chamamos de Análise, com as técnicas baseadas no aprendizado de máquinas e avaliação do desempenho do modelo, essas atividades são compreendidas pela CI, porém, executadas pela equipe de tecnologia. Voltando para a fase de Aplicação, o modelo é posto em ação classificando, etapa já adquirida, e gerando relatório o qual chamamos de etapa de Análise e por fim apresenta a Tomada de Decisão que também será incorporada pois é o encerramento do ciclo informacional logo após a etapa que denominamos de apresentação dos resultados que contribuirá para gerar a base do conhecimento.

A construção do novo ciclo absorveu primeiramente as etapas mais importantes da análise de sentimento, observadas por autores da área que as destacam, e que por conveniência aparecem na maioria dos ciclos. Uma outra observação foi remover dualidades, etapas que pudessem ter nomes similares ou processos repetitivos. E só

então remover etapas que não agreguem, mais confundem o ciclo do que aprimoram. Assim, utilizando as técnicas consagradas pela CI, conseguimos elaborar um ciclo de Análise de sentimento completamente voltado para a faceta da informação, desde a ideia e concepção até a reta final.

Este novo ciclo possui a vantagem de poder ser aplicado a qualquer meio ou processo decisório que se necessite, produto, tópico, organização, empresa, pessoa e muito mais. Pode ser usado de maneira genérica mas também é passível de modificações específicas para melhor atender as necessidades do cliente. Considerando todos estes aspectos, foi decidido não somente elaborar um modelo um modelo amplo e unificado, considerando o máximo de atividades possível, mas também organizado por papéis e responsabilidades, baseado na ideia do modelo de Shayaa *et al.* (2018).

Além disso, neste novo ciclo são sugeridas atividades que não são explicitadas nos modelos anteriores. As atividades propostas se referem a regulamentação e controle do uso dos dados de cidadãos em processos de Análise de Sentimento.

Também se decidiu que a representação visual do modelo utilizaria uma notação padronizada e de entendimento comum para diversas áreas de conhecimento, já que, as apresentadas até aqui, apresentam uma visão particular dos autores, ou mesmo, uma interpretação minha para o ciclo de vida. Para minimizar as subjetividades na interpretação de modelos visuais, foi decidido usar a notação BPMN (*Business Process Model Notation*) para representar o ciclo de vida unificado para Análise de Sentimento. Assim, a própria notação já possui seus significados que tornam o ciclo de vida melhor compreendido. E por fim, como forma de permitir a contribuição da comunidade, decidimos deixar os arquivos BPM disponíveis para o público geral para que além de reutilizar o modelo, estes possam complementá-los com suas ideias e a medida em que forem recebidos os feedbacks, estes serão incorporados a uma base comum do modelo.

A versão proposta para o modelo pode ser vista na Figura 9 a seguir. Aqui sugere-se a participação de seis atores:

Governo: Ator que regulamenta as políticas voltadas à Análise de Sentimento e fiscaliza o uso de tal técnica de processamento de dados. Na União Europeia isso está sendo regulamentado na GDPR (Lei geral de proteção de dados europeia).

Organização Demandante: É o verdadeiro cliente do processo, ou seja, a organização, pessoa física ou terceiro setor que demanda o serviço que utiliza a Análise de Sentimento. É importante ressaltar que o usuário final, em raros casos, será o demandante de tal perspectiva. Então, será mais frequente que os dados de diversos indivíduos sejam utilizados para propósitos de organizações, e não dele próprio.

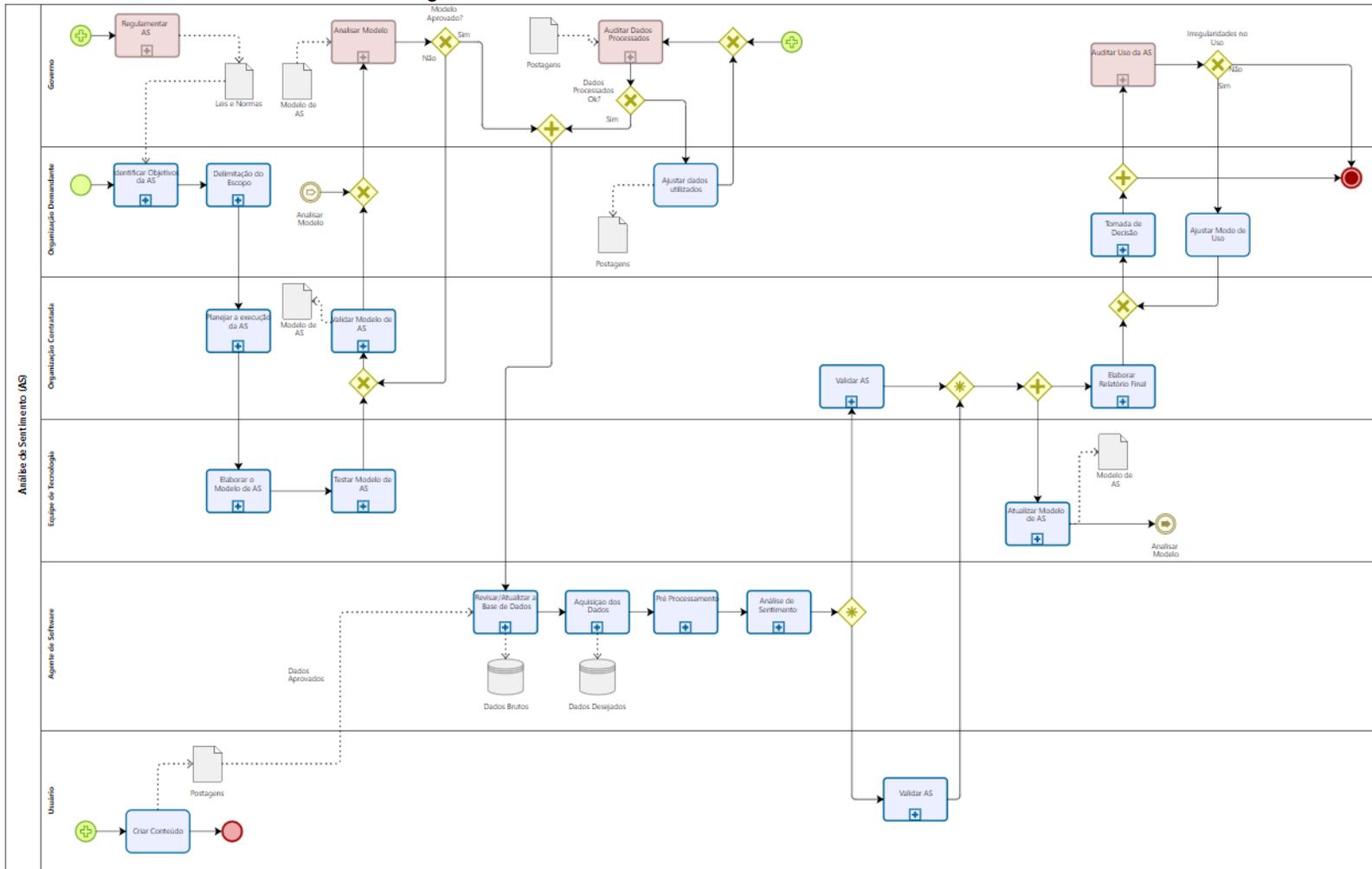
Organização Contratada: Se trata da empresa de TIC (em seu CNPJ) que tem suas políticas e sua missão, visão e valores. É ela quem pode responder juridicamente pelos seus produtos e quem pode se manifestar, como organização, pelo produto criado.

Equipe de Tecnologia: Essa é a equipe que irá conceber, treinar e disponibilizar o agente de software que irá realizar o processo Análise de Sentimento. Esta equipe responde por questões técnicas e tecnológicas a respeito do produto.

Agente de Software: Este é o agente de software em si e é a entidade que realizará o processamento de informação e disponibilizará os resultados finais para os usuários e interessados na informação.

Usuário: Estes são os usuários dos sistemas que fornecerão os dados que serão analisados pelo agente de software. É sobre os sentimentos destes indivíduos que a organização demandante irá obter seus resultados.

Figura 9 - Processo Unificado de Análise de Sentimento



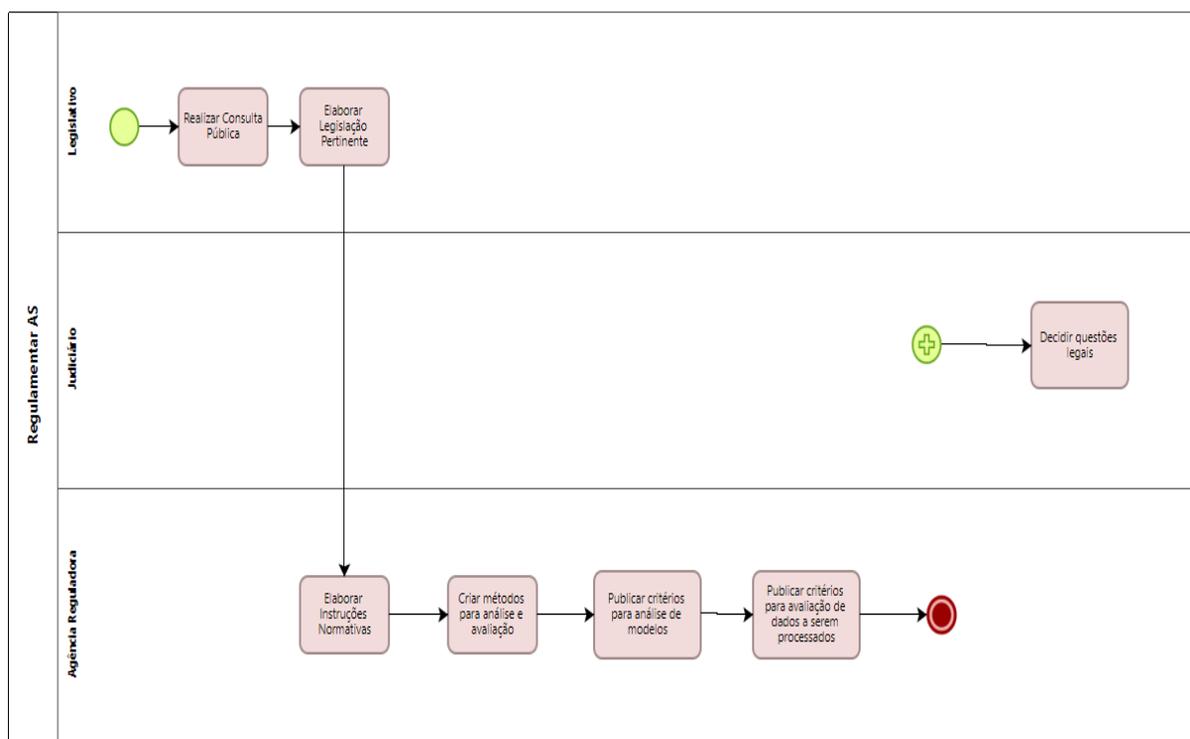
Fonte: A autora (2021).

O Modelo Unificado para Análise de Sentimento propõe vinte e duas atividades, que são distribuídas pelos atores, que explanadas a seguir:

4.1 Regulamentar Análise de Sentimento

Esta etapa se refere a ação inicial governamental de regulamentar as atividades de Análise de Sentimento detalhado na figura a seguir.

Figura 10 - Processo Unificado de Análise de Sentimento



Fonte: A autora (2011).

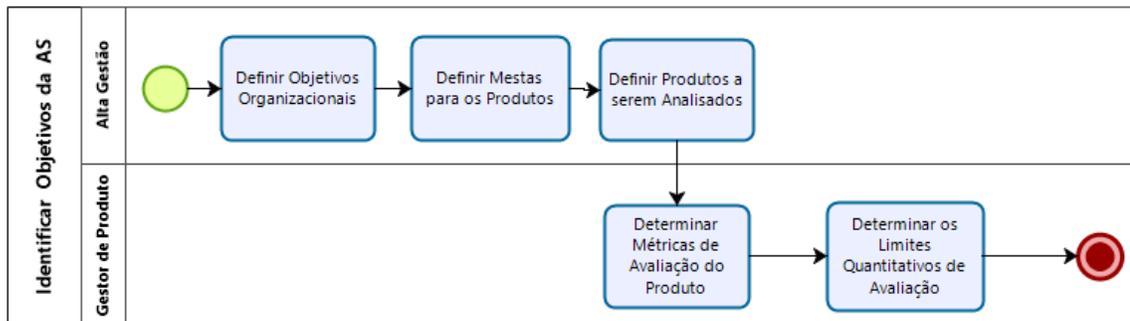
O processo se inicia com a realização da consulta pública para em seguida seja realizada a elaboração das leis que irão regulamentar o uso de Análise de Sentimento. Após a aprovação da lei, cabe à agência reguladora elaborar as instruções normativas e critérios que irão implementar o texto da lei. Após a elaboração destas definições cabe ao judiciário verificar os excessos e ajustes à lei ou instrumentos normativos que não estão em consonância com as legislações de mais alto nível.

Aqui é criada a estrutura de regulamentação a partir de uma lei e do detalhamento da regulamentação por uma agência.

4.2 Identificar objetivos da Análise de Sentimento

Esta etapa se refere a ação inicial do órgão demandante onde este formaliza as demandas que o levaram a iniciar uma ação que utilizará técnicas de Análise de Sentimento para processar dados de indivíduos disponibilizados na web. O detalhamento deste processo é apresentado na figura a seguir.

Figura 11 - Detalhamento da Atividade Identificar Objetivos da AS



Fonte: A autora (2021).

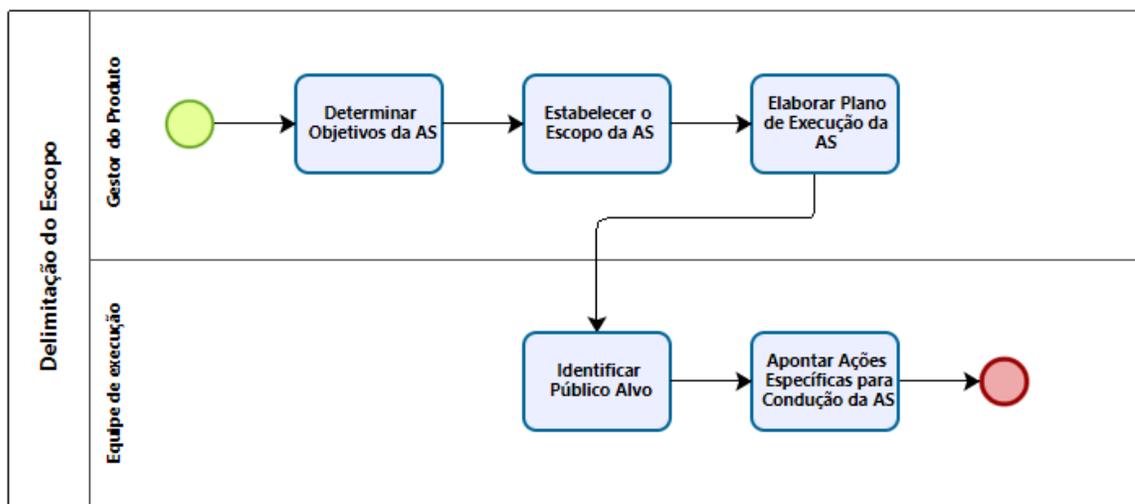
Esta ação se inicia no alinhamento estratégico da organização quando esta define seus objetivos organizacionais por parte da alta gestão. Em seguida, a organização define como se dará o desenvolvimento de seus produtos através da elaboração de métricas estratégicas e da escolha de quais produtos terão a percepção e sentimento dos consumidores analisados. Em seguida, os gestores dos produtos definirão quais aspectos quantitativos e qualitativos serão considerados e quais os resultados esperados.

4.3 Delimitar o escopo

A etapa se refere às especificações mais detalhadas por parte da organização demandante dos objetivos e resultados esperados da Análise de

Sentimento a ser executada. O detalhamento deste processo é apresentado na figura a seguir.

Figura 12 - Detalhamento da Atividade Delimitação do Escopo



Fonte: A autora (2021).

Sant'Ana (2016) afirma que antes de se fazer qualquer coleta de informação é preciso obter os dados que serão utilizados para atender a necessidade específica ou demanda prevista de informações sobre um determinado contexto.

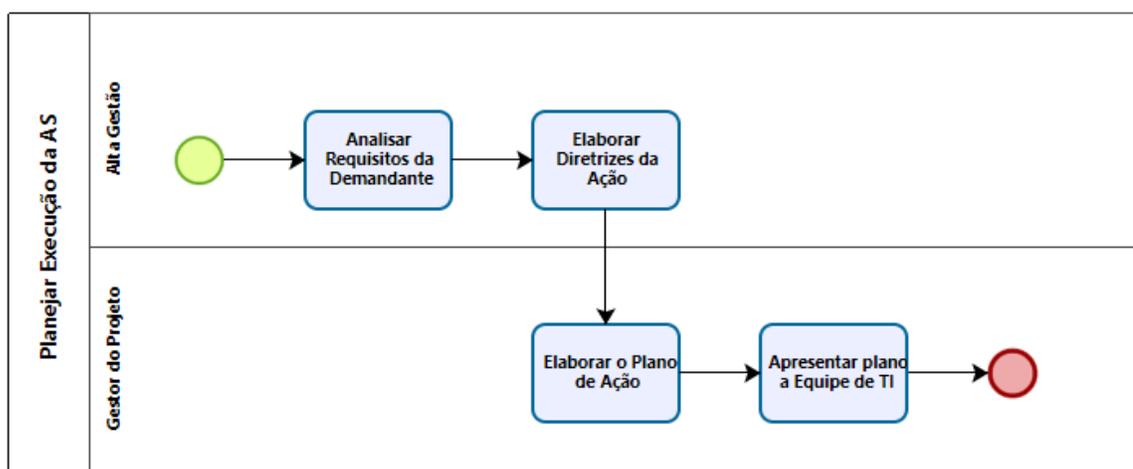
Assim, para esta etapa, compreende-se a necessidade de identificar os objetivos da AS, delimitação do escopo dos dados, elaboração do projeto de execução, identificar o público alvo e fornecer informações específicas como ações de vida real ou outros dados relevantes que podem impactar na relação do consumidor com o produto em um período específico.

Destaca-se a elaboração do plano de execução que deve detalhar o resultado esperado, definir as características da informação que se deseja obter, de quais plataformas específicas os dados serão coletados, formato em que se encontram as informações, se há riscos para algum quesito da privacidade do usuário, como os dados serão classificados e avaliados, e como se dará o acesso a tais informações. É possível sumarizar esta etapa introdutória como o desenho do que se deseja recuperar e quais áreas se deseja atingir.

4.4 Planejar Execução da AS

Na etapa anterior a organização demandante realizou o seu planejamento para maximizar a eficácia da ação. Agora, se faz necessário que a empresa de tecnologia garanta que o projeto seja executado de forma adequada e que se atinja os objetivos propostos pela organização demandante. O detalhamento deste processo é apresentado na figura a seguir.

Figura 13 - Detalhamento da Atividade Planejar Execução da AS



Fonte: A autora (2021).

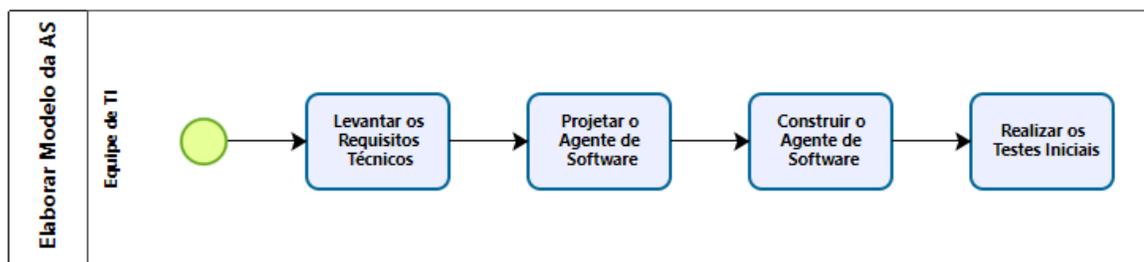
Esta se inicia com a análise do que é solicitado pela organização demandante e em seguida são elaboradas as diretrizes da ação que envolverá o uso da Análise de Sentimento. A partir destas orientações institucionais, é elaborado o plano de ação contendo todo o andamento detalhado de como será realizado o trabalho. E por fim, este plano é apresentado à equipe de TI onde pode sofrer pequenos ajustes devido a aspectos técnicos.

4.5 Elaborar Modelo de AS

Esta etapa se refere a concepção do agente de software que irá realizar a Análise de Sentimento. Ressalta-se que esta é uma atividade que ocorre sob

demanda e que o agente inteligente criado é único e exclusivo. O detalhamento deste processo é apresentado na figura a seguir.

Figura 14 - Detalhamento da Atividade Elaborar Modelo da AS



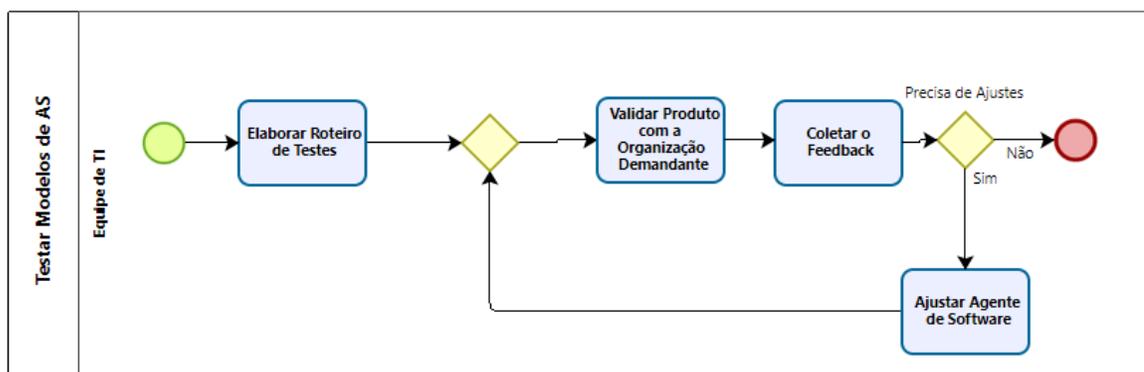
Fonte: A autora (2021).

O processo se inicia com a definição dos requisitos técnicos do agente a ser concebido. Aqui é definido o que o agente deve fazer e os requisitos técnicos para sua construção. Em seguida a Equipe de TI realiza o projeto do software, com o intuito de minimizar o erros durante a etapa seguinte que é a construção. Após concluído o desenvolvimento do agente, este deve ser testado para se verificar se há erros de ordem técnica.

4.6 Testar Modelo de AS

Esta etapa se refere a apresentação formal do agente de software à Organização Demandante. Os testes de natureza técnica foram realizados na etapa anterior, então cabe nesta etapa à organização demandante validar se o que será entregue pelo software é o que ela espera. O detalhamento deste processo é apresentado na figura a seguir.

Figura 15 - Detalhamento da Atividade Testar Modelo de AS



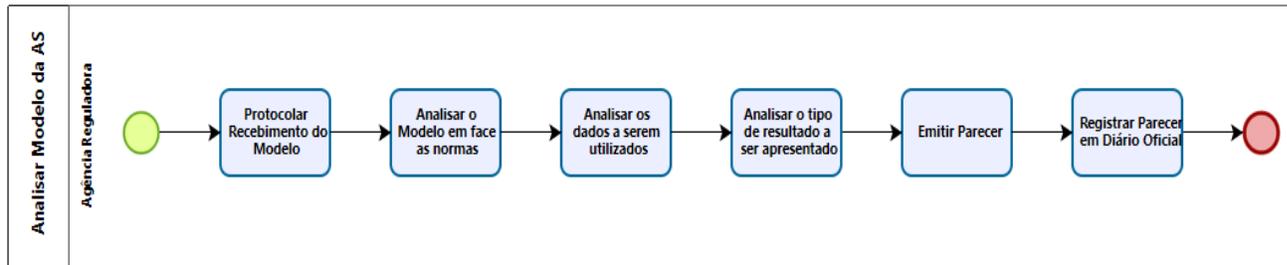
Fonte: A autora (2021).

O processo é iniciado pela elaboração do roteiro de testes com o objetivo de organizar as atividades e potencializar os efeitos deste processo. Em seguida, o software é apresentado à organização demandante em uma versão ainda não calibrada, e conseqüentemente, não pronta do agente de software. Mas, já se é possível indicar se o que o agente irá entregar é de interesse para a organização demandante. Após esta validação, a organização demandante provê o feedback sob a demonstração e os ajustes lá recomendados, caso haja, são realizados. Este processo continua até que não haja mais ajustes solicitados pela organização demandante.

4.7 Analisar Modelo de AS

Esta etapa se refere a avaliação do modelo criado pela agência reguladora. O intuito desta ação é analisar se o modelo proposto segue todos os aspectos legais e regulatórios, sendo este uma chancela para autorização do uso. O detalhamento deste processo é apresentado na figura a seguir.

Figura 16 - Detalhamento da Atividade Analisar Modelo de AS



Fonte: A autora (2021).

Esta etapa se inicia com o protocolo de recebimento do modelo e sua documentação por parte da agência reguladora. Em seguida será analisado se o modelo segue as restrições regulatórias, se os dados a serem utilizados são permitidos e se o resultado que será entregue pelo modelo é condizente com o que permite a legislação. Em seguida o resultado do parecer detalhado é enviado a Organização Demandante e Organização Contratada e um resumo deste é publicado em canal de comunicação público do governo, Diário Oficial da União, explicitando a aprovação ou reprovação.

4.8 Criar Conteúdo

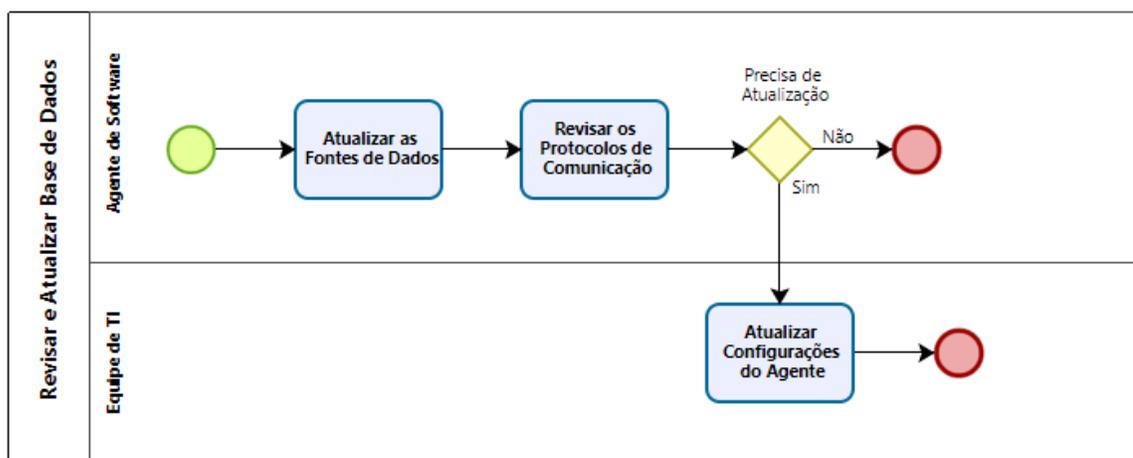
Como é possível ver no diagrama geral, Figura 16, a criação de conteúdo não é um subprocesso e sim uma atividade que ocorre de maneira orgânica e descentralizada a partir das postagens dos usuários nos mais diversos canais de mídias sociais na Internet. Por isso, não há uma maneira de realizar o detalhamento desta atividade, mas deve-se ressaltar que este processo realizado pelos bilhões de usuários é que possibilita a realização deste processo.

4.9 Revisar e Atualizar Bases de Dados

Esta atividade não deverá ocorrer para a concepção de todos agentes inteligentes que utilizam de técnicas de Análise de Sentimento, mas, apenas para aqueles que armazenam dados em um repositório interno, *offline*, dos dados

existentes na rede. Em alguns casos, pode ser mais interessante para o agente de software, que este colete os dados disponíveis na Internet frequentemente, uma vez por dia por exemplo, e o agente passe a utilizar estes dados. Para esses casos, esta etapa também se mostra necessária, ainda que em uma frequência menor do que a coleta de dados em si. O detalhamento deste processo é apresentado na figura a seguir.

Figura 17 - Revisar e Atualizar Base de Dados



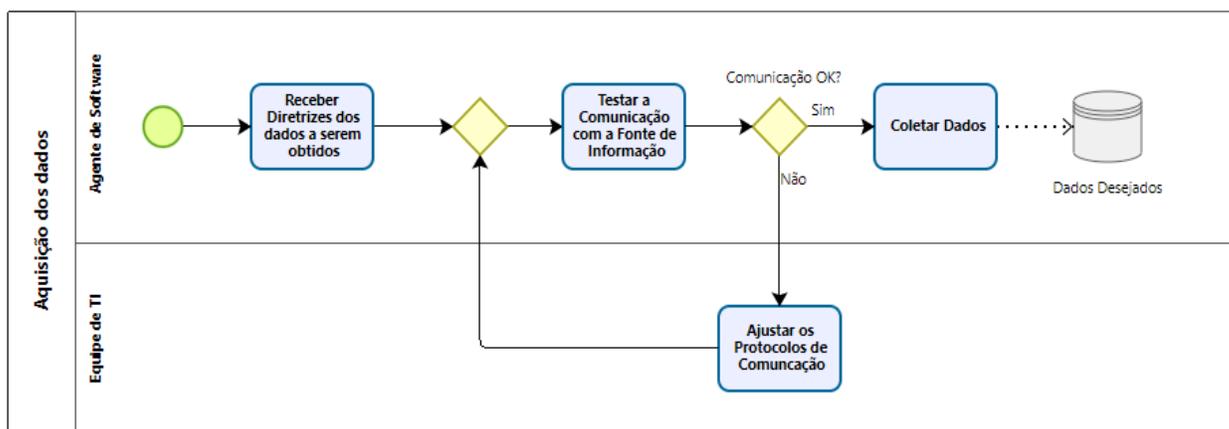
Fonte: A autora (2021).

Para tanto, o agente de software atualiza a lista das fontes de informações que são válidas e testa os protocolos de comunicação com cada uma delas. Se houver algum problema de comunicação, o que impossibilitaria a aquisição dos dados, a equipe de TI deve ajustar os protocolos de comunicação.

4.10 Aquisição dos Dados

Este é o processo que adquire os dados que serão utilizados na Análise de Sentimento e considerando o contexto de Big Data o destaque desta coleta se dá na convergência dos fatores delimitados na etapa definir modelo da AS. A ideia aqui é trabalhar com a ideia de recortes de tempo, adjetivos, e de marcações (folksonomia ou *hashtags*). O detalhamento deste processo é apresentado na figura a seguir.

Figura 18 - Aquisição dos Dados



Fonte: A autora (2021).

O processo se inicia com a definição das diretrizes dos dados a serem coletados. Quais mídias sociais? Quais personas? Postagens de quando? Com quais características? Dentre outras informações que auxiliam o agente a coletar os dados que serão mais úteis na análise. Em seguida é realizado um teste da comunicação com as fontes de dados e se tudo estiver correto, ocorre a coleta dos dados. Caso necessário, a equipe de TI pode realizar ajustes no agente para reparar alguma falha.

4.11 Pré-Processamento

O tratamento da informação consiste e depende de um sistema complexo e uma de suas facetas é a seleção dos dados, que se encontra como subárea da recuperação da informação. A seleção é um dos pilares de qualquer sistema de recuperação da informação, sendo uma função fundamental à CI. Se não houver uma seleção no tratamento de um determinado conjunto de documentos, ou informação, então não podemos falar de ciclo de vida da informação ou de recuperação da informação, já que uma de suas funções principais não foi devidamente suprida (DIAS, 2016).

Acontecerá o recorte no Big Data. O tratamento selecionado da informação não surgiu primeiramente no contexto digital, mas se tornou essencial para uma

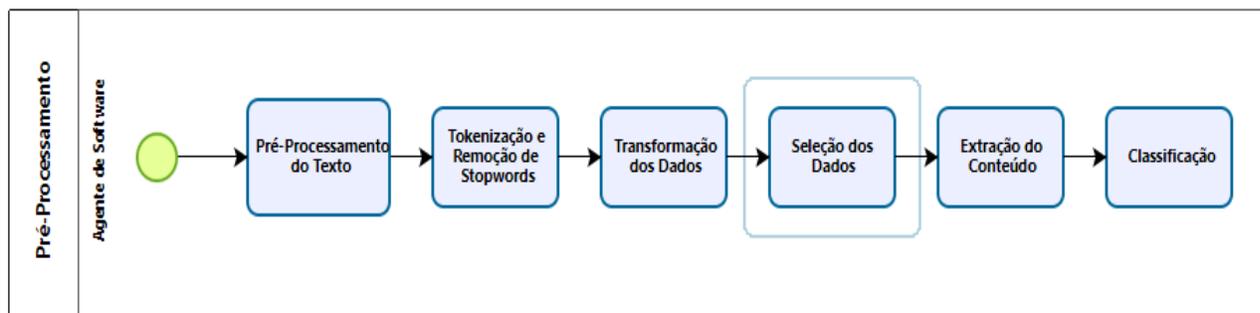
melhor definição da natureza e características das diversas plataformas digitais, atendendo a necessidades informacionais específicas trazendo soluções eficientes. Dessa forma, os termos são escolhidos dos textos selecionados, agrupados, e, então, as facetas são definidas a partir dos grupos. As postagens são então classificadas, e os termos, em cada postagem, de acordo com as necessidades das organizações (TRISTÃO, 2004).

Serão utilizados também termos e vocabulário específicos também conhecidos como tesouros, que nada mais é do que uma seleção de termos, baseados em análise de conceitos, na qual se define o termo geral, de maior abrangência, e sua relação com termos mais específicos, que representam os conceitos menores. Para a seleção é preciso intencionalidade no conceito, sempre respectivo a um grupo de informações em que há uma seleção das características relevantes para aquela área e para os propósitos do trabalho. Assim, conforme a área e conforme o ponto de vista abordado, mudam as características e conseqüentemente as relações entre os conceitos, e é por esse motivo que os tesouros são desenvolvidos de forma específica a cada área do conhecimento, sendo um conjunto de características de um conceito específico.

Considerando as características necessárias para utilização do conceito e considerando a importância do sistema de conceitos a ser construído, bem como a expectativa do usuário. Extensão e intenção são as formas de captação e identificação das informações (nesse caso as postagens) e influenciam na elaboração da seleção. A extensão e quantidade de conceitos selecionados que um grupo abarca é agregado por características relacionadas e com a delimitação das características. Portanto existe é o conjunto de características de um conceito. Quanto mais características possuir um conceito, melhor e mais específica se torna a seleção (TRISTÃO, 2004).

Todos estes aspectos ficam englobados neste Ciclo de Vida Unificado na etapa de pré-processamento, uma vez que é aqui onde o conteúdo que será submetido ao algoritmo de análise de sentimento é filtrado, selecionado e preparado. O detalhamento deste processo é apresentado na figura a seguir.

Figura 19 - Pré-Processamento



Fonte: A autora (2021).

Aqui consideramos que a primeira etapa do processo é a realização de um tratamento de dados mais geral dispendo os dados brutos em uma estrutura mínima para passar pelos processos posteriores. Afinal, o uso de linguagem natural dificulta o processamento do texto e normalmente, textos dispostos neste código, são pré-processados para um formato mais amigável para o tratamento posterior. Em seguida são identificadas as palavras-chave (tokens) de cada sentença e são removidas preposições, artigos e outras palavras que não tem significado em si (*stopwords*).

Em seguida ocorre uma nova transformação nos dados para auxiliar na etapa de extração do conteúdo. Esta transformação tem o intuito de facilitar a compreensão, por parte do agente de software, a partir de uma nova estrutura dos termos, esta etapa também pode ocorrer após a etapa de seleção de dados (seguinte).

Na seleção dos dados, será definido quais sentenças e palavras chave serão utilizadas para se passar adiante e depois se dá a extração do conteúdo que a disposição do conteúdo selecionado em uma estrutura de dados ótima para que o algoritmo de análise de sentimento seja executado. Por fim, os termos extraídos são classificados e agrupados para finalmente passar pelo processo de AS.

4.12 Análise de Sentimento

Na CI o conceito de classificação consiste em um processo definido como “dividir em grupos ou classes, segundo as diferenças e semelhanças”. É dispor os conceitos, segundo suas semelhanças e diferenças, em certo número de grupos metodicamente distribuídos” (PIEDADE, 1977, p. 9). Para a recuperação da informação, o tratamento da informação é definido como a função de descrever os documentos, tanto do ponto de vista físico (características físicas dos documentos) quanto do ponto de vista temático (ou de descrição do conteúdo). Para as definições deste trabalho trabalharemos também com o termo hashtag, que consiste em associar temas a tópicos, especificamente como uma só palavra ou frase muito curta.

Essa etapa resulta em determinações ou representações documentais, mais comumente encontradas na Ci como fichas de catálogo, referências bibliográficas, resumos, termos de indexação e etc, tentando trazer uma certa facilitação para a manipulação e recuperação de determinado documento ou sistema informacional, mas também é usado para uma representação simples e de fácil avaliação desenvolvido para meios específicos. É o caso do código de avaliação utilizado pelo sistema de classificação “positivo, negativo ou neutro” para delimitar a postagem recuperada em sua categoria uma vez analisado o contexto da postagem em sua totalidade.

Cada postagem pode ser tratada do ponto de vista da totalidade, dos elementos e das relações considerando:

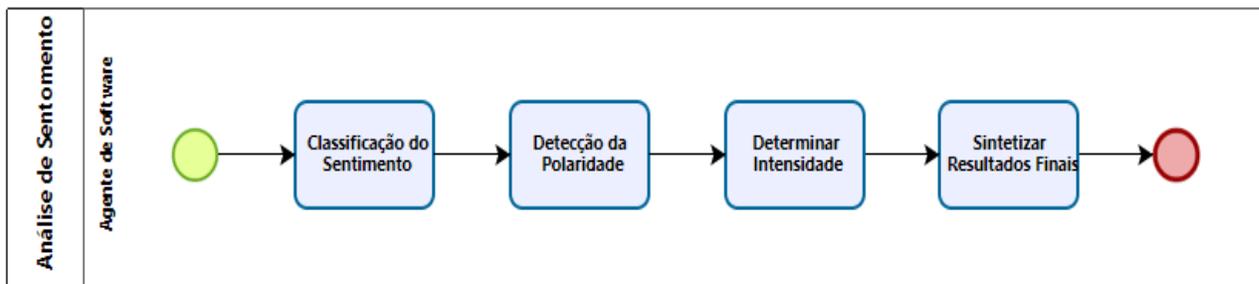
- 1) do ponto de vista da totalidade, enquanto podem ser estruturados mediante o uso do método dedutivo, como até agora se costumou fazer com os sistemas de classificação tradicionais;
- 2) do ponto de vista dos elementos enquanto é utilizado o método indutivo, praxe geralmente seguida na elaboração dos tesouros;
- 3) do ponto de vista das relações, enquanto é posto em relevo fundamentalmente o relacionamento categórico das classes. Por este último processo o problema da classificação não recai sobre a divisão ou subdivisão de classes, mas sobre o relacionamento das mesmas classes, podendo-se lembrar -a propósito a definição de classificação apresentada pelo comitê FID/CR: "Por classificação deve-se entender qualquer processo de criar relações, genéricas ou outras, entre as unidades semânticas [...]"

Em síntese, a chamada "teoria analítica dos conceitos" fornece elementos para: a formulação dos conceitos; a compreensão dos conceitos como sínteses de características; a origem das características dos conceitos; as relações conceituais; a construção de sistemas de conceitos de acordo com critérios objetivos. Levando em consideração também os tesouros constituídos (DAHLBERG apud DIEMER, 1978).

O processo de análise consiste em pôr em observação os detalhes obtidos. Corresponde a interpretar o que está sendo apresentado nas entrelinhas. Estuda as características e possíveis aplicações de tratamento da informação. Faz-se referência a um exame qualitativo do objeto em questão de acordo com os métodos especializados da área visando descobrir elementos. Cabe destacar que a análise contempla diferentes dimensões que permitam relações direta ou indiretamente relacionadas ao texto ou produto em questão, inclusive envolvendo o aspecto individual e coletivo do objeto em questão perante aos usuários e a sociedade.

Na Ciência da Informação a Análise de Informação é apreciada através dos estudos bibliométricos e infométricos que representam sistemas automáticos que, através de cálculos, informam quantas vezes determinada palavra foi utilizada apontando para uma tendência na área contemplada que consta no texto, na base de dados ou até mesmo na classificação anteriormente abordada (TARAPANOFF; ARAÚJO, 2000). É na etapa de análise que a informação agrega valor quando permite a percepção de oportunidades e ameaças à sua operacionalidade, inclusive detectando tendências ou problemas, inclusive apresentando novos produtos e novas maneiras de fazer negócio, desta forma a análise é fator fundamental para a sustentação e desenvolvimento do ciclo de vida informacional da análise de sentimento (FERREIRA, 2003). O detalhamento deste processo é apresentado na figura a seguir.

Figura 20 - Análise de Sentimento



Fonte: A autora (2021).

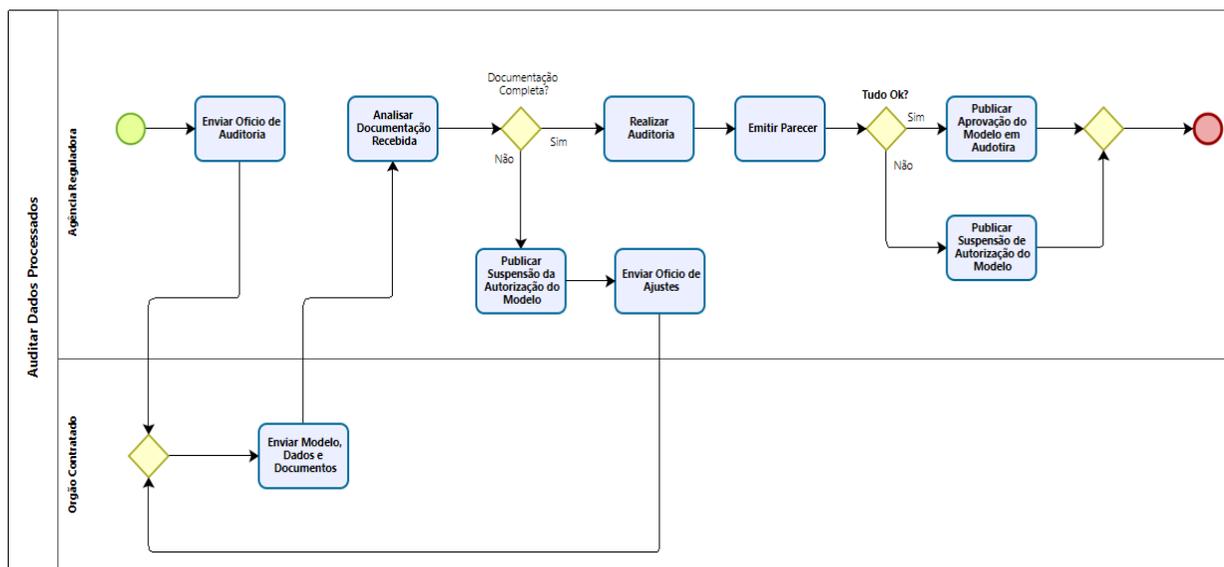
A Análise de Sentimento se inicia com a identificação dos sentimentos envolvidos naquelas postagens. Em seguida tenta se detectar se estes sentimentos são positivos, negativos ou neutros, bem como a intensidade do sentimento. Por fim, o agente pode sintetizar os resultados da análise a partir de relatórios ou outro formato pré-definido.

4.13 Auditar dados processados

Todo modelo de AS aprovado pela agência reguladora deve passar por alguma espécie de auditoria para verificar os dados utilizados e se eles estão de acordo com a legislação, por isso, pelo menos uma vez enquanto está em atividade, o ciclo de vida deve ser auditado. O detalhamento deste processo é apresentado na figura a seguir.

O processo se inicia com o aviso formal por parte da Agência Reguladora de que o modelo e os dados utilizados serão auditados. Em seguida a organização contratada (TI) enviará toda a documentação para agência. Caso esta não venha de forma adequada, a agência suspenderá preventivamente a autorização do modelo e este só será recuperado caso aprovação da auditoria. Este processo se repete até que a documentação seja enviada de forma esperada pela agência, que realiza a auditoria e emite um parecer. Caso haja problemas no modelo um ofício publicado em diário oficial suspendendo a autorização de uso do modelo será emitido, caso contrário, a aprovação do modelo e autorização para uso será publicada em diário oficial.

Figura 21 - Auditar Modelo



Fonte: A autora (2021).

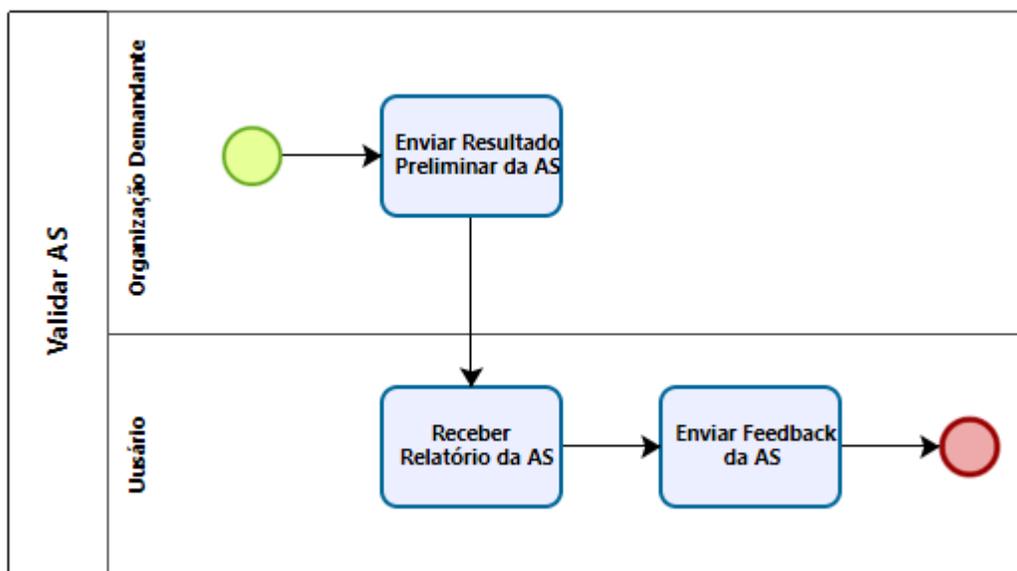
4.14 Utilizar os Dados Ajustados

Caso o modelo apresente problemas na auditoria, ajustes devem ser realizados no modelo e para conduzir tal modificação, se deve seguir o mesmo processo de Elaborar o Modelo da AS, onde a recomendação da agência reguladora deve estar nas especificações do modelo.

4.15 Validar AS (Usuário)

Esta etapa é uma inovação deste ciclo de vida unificado pois, não é comum que o indivíduo sequer saiba que o sentimento dele foi analisado e muito menos que ele promova o *feedback* disto. Aqui a intenção não é só o usuário validar os resultados do modelo, mas ser comunicado sobre que análise foi conduzida e quais os resultados gerados a partir dela. O detalhamento deste processo é apresentado na figura a seguir.

Figura 22 - Validar AS (Usuário)



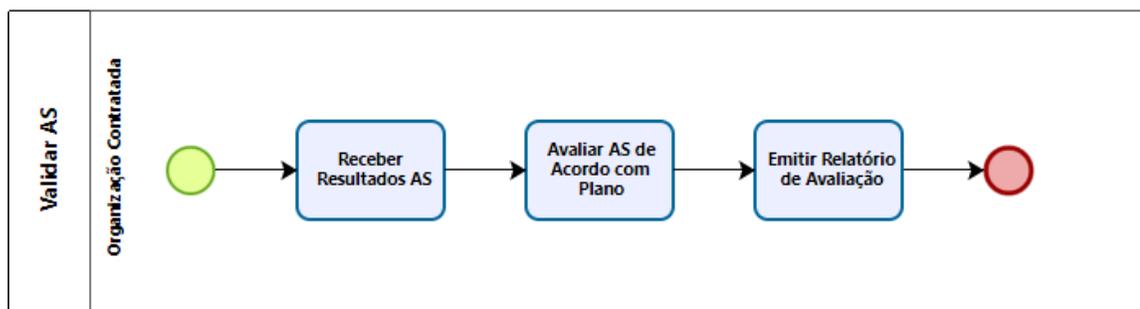
Fonte: A autora (2021).

O processo se inicia com o envio por parte da organização demandante da AS para o usuário. Este recebe o relatório com os resultados da AS e este irá promover um feedback, inclusive com a possibilidade de este não querer mais ser incluído nas próximas análises.

4.16 Validar AS (Organização Contratada)

Esta etapa se refere a avaliação do modelo criado pela organização contratada quanto a efetividade do realizado frente ao planejado. Esta atividade serve como uma avaliação formal dos resultados do modelo O detalhamento deste processo é apresentado na figura a seguir.

Figura 23 - Validar AS (Organização Contratada)



Fonte: A autora (2021).

O processo se inicia com o recebimento por parte da organização demandante dos resultados preliminares da AS. Em seguida, a própria organização conduz uma avaliação perante ao que foi planejado e dispõe os resultados em um relatório que pode servir de insumos para revisões e alterações do modelo.

4.17 Atualizar Modelo da AS

Após a análise tanto interna da organização contratada e do usuário, caso se faça necessário realizar ajustes no modelo, deve ser seguido o mesmo processo de Elaborar o Modelo da AS, onde estas novas informações devem constar nas especificações do modelo.

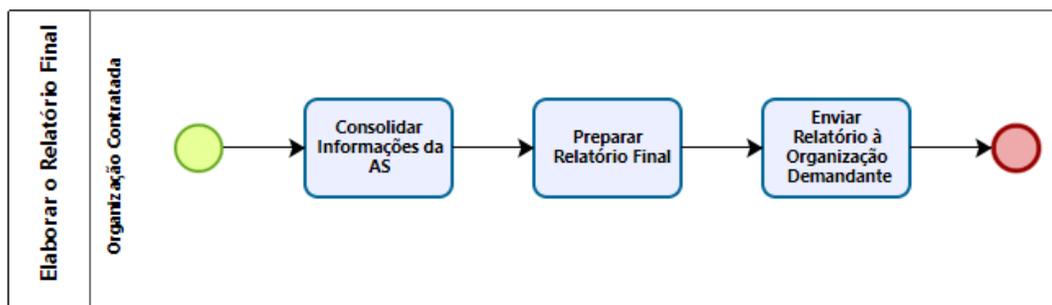
4.18 Elaborar Relatório Final

Esta etapa tem como objetivo consolidar e apresentar os resultados obtidos pelas etapas anteriores. A representação de forma agradável é um componente indispensável para a aceitação, por parte da organização demandante, de todo o processo da Análise de Sentimento. Afinal, o requerente não deve compreender como trabalhar em todas as etapas do ciclo do ponto de vista técnico, ele precisa de uma resposta aos seus anseios de negócio.

Para a Ciência da Informação, a representação da informação nada mais é do que um documento. Briet (2006, p. 7) afirma que documento é “qualquer signo físico ou simbólico, preservado ou registrado com a intenção de representar, reconstruir ou demonstrar um fenômeno físico ou abstrato”. Apoiando-se em diferentes tipos de representação de resultados, esta pode ser apresentada na forma descritiva: padroniza pontos de acesso facilitando a busca e recuperação pois agrupa o conjunto de documentos ou informações semelhantes ou de série específica. Já a representação temática torna mais fácil a recuperação do material de acordo com o tema, nesse caso são elaborados resumos breves e coerentes com a essência do documento, pode ser feito também extrações de frases ou termos mais encontrados durante a avaliação da análise.

O ponto de vista da representação da informação é dar ênfase à organização do conhecimento, que na CI diz respeito à avaliação e análise de determinado conteúdo visando elaborar métodos de representação das informações geradas. Sendo assim a transferência de um ponto de vista da informação (NOVELLINO, 1996). A principal característica da etapa de resultados é representar a informação substituindo o processo longo e complexo. Não só utilizando a linguagem escrita mas desenvolvendo gráficos e imagens que colaborem com a assimilação dos resultados. O detalhamento deste processo é apresentado na figura a seguir.

Figura 24 - Elaborar Relatório Final



Fonte: A autora (2021).

O processo se inicia com a consolidação dos resultados da AS e em seguida é elaborado o relatório final contendo tabelas, gráficos e imagens dos resultados do processo de AS conduzido. O término desta atividade se dá com o

envio do relatório final, e outras questões contratuais de encerramento do projeto, para a organização demandante.

4.19 Tomada de Decisão

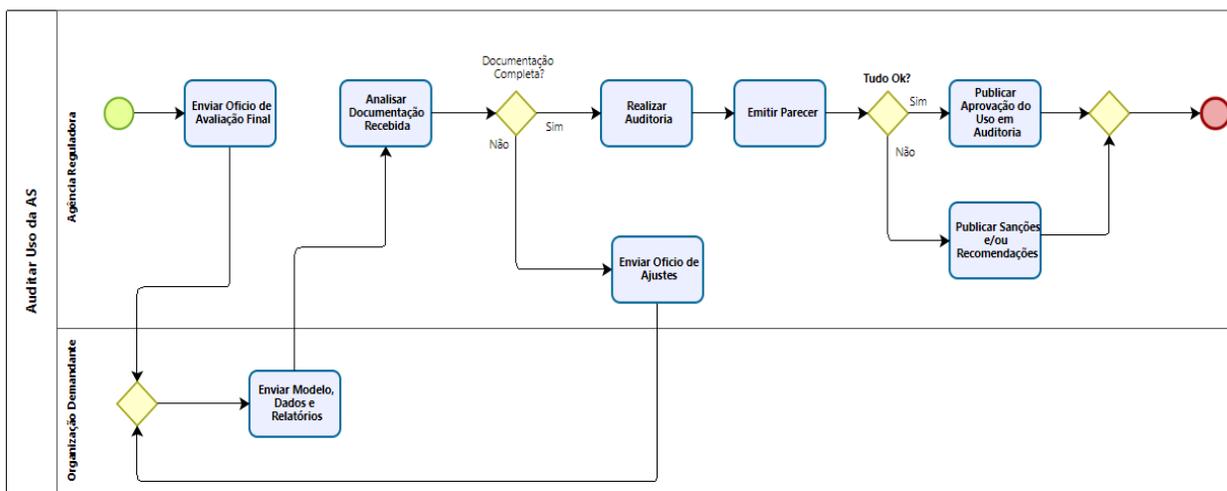
Este é o processo que se utiliza dos resultados da AS para que a empresa decida como irá proceder com o produto, marca ou público alvo investigados no processo. Devido a existência de muitos modelos para tomada de decisão organizacional, este subprocesso não será aqui modelado.

4.20 Auditar Uso da AS

Todo modelo de AS cujo as ações foram finalizadas deve passar por uma avaliação do uso para verificar se todos os aspectos legais e regulatórios foram seguidos. O detalhamento deste processo é apresentado na figura a seguir.

O processo se inicia com o aviso formal por parte da Agência Reguladora de que o modelo, os dados utilizados e forma como foram utilizados serão auditados. Em seguida, a organização demandante enviará toda a documentação para a agência. Caso esta não venha de forma adequada, a agência solicitará a documentação correta. Este processo se repete até que a documentação seja enviada de forma esperada. Em seguida, a agência realiza a auditoria e emite um parecer. Caso haja problemas no uso do modelo um ofício será enviado para a organização demandante contendo as sanções e recomendações futuras, e um resumo será publicado em diário oficial. Caso contrário, a aprovação do uso do modelo será publicada em diário oficial, embora que um ofício contendo boas práticas para melhoria possa ser enviado em separado.

Figura 25 - Auditar Modelo



Fonte: A autora (2021).

4.21 Ajustar modo de Uso

Este é o processo que se utiliza das recomendações da agência reguladora para que a empresa ajuste sua conduta em futuras ações que envolvam AS e os dados fornecidos por usuários. Como cada organização possui uma forma de ajustar seus procedimentos internos, este subprocesso não será aqui modelado.

4.22 Conclusão

A elaboração deste ciclo de vida unificado para análise de sentimento além de conter as boas práticas observadas nos modelos, também propõe um conjunto de atividades regulatórias e que envolvem o usuário no processo com o intuito de tornar mais transparente e seguro o uso dos dados dos indivíduos e que estes tenham a possibilidade de saber como estes foram utilizados. A adoção deste ciclo de vida, pode tornar menos danoso o mal uso dos dados dispostos em rede por máquinas e de organizações.

5 CONSIDERAÇÕES FINAIS

O avanço da tecnologia a princípio é assustador. Foi assim desde a chegada do fogo, não se tinha conhecimento se fazia mais mal do que bem, nem se era passível de ser controlado. Passando pelo século XV com as descobertas de Copérnico que puseram em questão o todo o conhecimento e poder do Clero. Também no século XVIII com a chegada da revolução industrial, onde as máquinas passaram a assumir as funções dos operários. Com o avanço da tecnologia da informação não seria diferente. Chegamos em uma era onde assusta saber que nossos dados estão à deriva e à disposição, e assusta mais ainda saber o quanto somos dependentes dessa troca de informações. Afinal, quem na atualidade vive sem um celular ou internet acaba por se tornar um eremita informacional.

A grande questão é? Como podemos viver em conciliação com essa tecnologia? Existem aspectos econômicos e sociais que permeiam tais considerações. Econômicos porque passamos a trabalhar e negociar por este meio, e sociais porque também passamos a nos relacionar por ele. Porém, existe um limite ético de invasão de privacidade que não pode ser ultrapassado. E é isso que desperta um dos primeiros sentidos de autopreservação da humanidade “este tal fogo é passível de ser controlado?”.

A minha resposta é sim. Claro que para que isso seja possível é necessário algum tempo de prática, estudo, habilidade, desenvolvimento. Mas sim, a humanidade aprende, evolui, cria técnicas para não ser consumida por suas descobertas que, algumas vezes, podem parecer perigosas para nossa própria existência. A história apresenta como a sociedade conseguiu se desenvolver para controlar as máquinas e não ser meramente mão de obra na era industrial, apesar do nível de dificuldade, pelo senso de sobrevivência, somos capazes de controlar e utilizar o avanço da tecnologia da informação para nosso benefício, embora tenhamos que nos queimar um pouco com esse fogo para aprender a manuseá-lo.

As consequências disso já são perceptíveis, quem nunca se sentiu invadido ao ter seus dados compartilhados e receber uma quantidade imensurável de

conteúdo ou chamadas nunca solicitadas? Será que tudo o que um indivíduo publicar poderá e será usado contra ele no tribunal do marketing informacional? Ou o fato de todos terem se tornado interagentes os obriga a cumprir uma certa cota? Ou melhor, se todas essas informações são tão valiosas para as organizações, porque é aceitável que sejam feitas de chuva de confetes que se espalham e vira aglomerado em qualquer bueiro? E com isso eu quero dizer que é inadmissível que dados pessoais estejam nas mãos (ou base de dados) de organizações nem o menor tipo de escrúpulo.

Hoje ainda somos instruídos a dizer que “li e aceito” os termos de utilização. Sendo que talvez não fosse necessário nos tornar tão susceptíveis às corporações ao aceitar tais condições. Seria essa a contribuição dos vigiados denominada por Foucault? Precisa haver um limite. A relação entre os agentes informacionais pode ser melhor elaborada através de uma abordagem analítica provocada pela Ciência da Informação. Desenvolver uma ética informacional é uma tendência necessária para já.

Nela se relacionam a observação da conduta social em meio ao uso das plataformas e mídias sociais, bem como a comercialização de dados pessoais e o exponencial interesse do marketing de organizações. Dessa forma a reflexão de como tem sido o impacto desse desenvolvimento para a sociedade contemporânea, e assim, ferramentas e métodos podem ser regulamentados e regularizados.

5.1 Contribuições

A Análise de Sentimento, amplamente abordada neste trabalho, é uma combinação de regras e ferramentas multidisciplinares que juntas apontam um resultado satisfatório de fornecimento, coleta e uso apropriado de informações fornecidas deliberadamente. Identificou-se um nicho em aberto no mercado que, capacitado pela inteligência artificial e máquinas morais, pode identificar opiniões. Discutir sobre os aspectos éticos e sobre a assertividade e utilização da AS é reconhecer seu impacto no cotidiano do indivíduo e da sociedade.

Foram identificados os principais aspectos técnicos da AS e desenvolvido um ciclo de vida que possa ser aplicado e/ou adaptado para qualquer nicho. Condensando e padronizando nomenclaturas divergentes para as mesmas propostas de atividades, por diferentes autores, normatizando assim a estrutura da AS. Dessa forma pode-se esperar uma ampliação na propagação do vislumbre e da utilização da AS.

5.2 Impactos Teóricos e Práticos

Considerados os prós e contras da aplicação da AS e do impacto de Máquinas Morais na contemporaneidade, este estudo considera que ainda é muito abrangente o universo a ser explorado. Apesar de ter encontrado um ponto de reflexão dentro da CI, ainda existem várias facetas que precisam de um olhar ético e científico mais apurado. E este estudo é apenas um ponto de partida.

A sociedade apresenta um comportamento, ao mesmo tempo, receoso com relação à proteção de sua privacidade e também conformado de que precisa interagir com as plataformas de dados. Como dito anteriormente por Brooks (1980) a informação é um elemento que provoca transformações nas estruturas. E o desenvolvimento dessa estrutura requer habilidades específicas que só se adquire através da análise e discussão, até onde se pode seguir, e onde é o limite para a utilização dos dados de outrem.

O conceito de máquinas morais e do seu poder de julgamento sobre a sociedade contemporânea leva também à reflexão sobre a conjuntura econômica e tecnológica interferindo no direito à privacidade. Na prática, os dados já estão sendo coletados, processados e utilizados. Mas na teoria ainda está muito longe de ter todas as suas questões respondidas.

REFERÊNCIAS

AFONSO, A. A referenciação em textos do YouTube: um estudo com vistas à análise de sentimentos. **Liinc em Revista**, Rio de Janeiro, v.13, n.2, p. 357-370, nov. 2017. Disponível em: <http://revista.ibict.br/liinc/article/view/3933>. Acesso em: 8 jan. 2022.

AGUIAR, E. J. **Análise de Sentimento em Redes Sociais Utilizando Combinação de classificadores**. 2017. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) – Universidade Estadual do Norte do Paraná, Bandeirantes, 2017. Disponível em: <http://200.201.11.152/handle/123456789/240>. Acesso em: 8 jan. 2022.

ANDRIOTIS, P.; OIKONOMOU, G. Messaging activity reconstruction with sentiment polarity identification. *In: INTERNATIONAL CONFERENCE ON HUMAN ASPECTS OF INFORMATION SECURITY, PRIVACY, AND TRUST*. 9., 2015, Springer. **Anais [...]** New York: Cham, 2015.

ARAÚJO, E. A. de. **A construção social da informação: práticas informacionais no contexto de organizações não-governamentais**. 1998. Tese (Doutorado em Ciência da Informação) - Universidade de Brasília, Brasília, 1998.

ASSIS, F. **ASDP: Um Processo de Análise de Sentimento em Debates Polarizados não Ideológicos**. 2014. Dissertação (Mestrado em Ciência da Computação) – Centro de Informática, Universidade Federal de Pernambuco. Recife, 2014. Disponível em: http://bdtd.ibict.br/vufind/Record/UFPE_207738333a3876843ca9a46bbaa1e8c6. Acesso em: 8 jan. 2022.

BARACHO, R. *et al.* Análise de sentimento sobre veículos em redes sociais. *In: ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO*. 13., 2012, Rio de Janeiro. **Anais [...]** Rio de Janeiro: UFRJ, 2012.

BARBIERI, F. **Um pequeno guia para F. A. Hayek**. Students for Liberty, 4 dez. 2014. Disponível em: <https://studentsforliberty.org/brazil/blog/pequeno-guia-friedrich-hayek/>. Acesso em: 7 jan. 2022.

BARDIN, L. **Análise de conteúdo**. Lisboa: Edições 70, 2006.

BECKER, J. **The first book of information science**. Washington: Energy Research and Develop, 1976.

BENEVENUTO, F.; RIBEIRO, F.; ARAÚJO, M. Métodos para Análise de Sentimentos em mídias sociais. *In: BRAZILIAN SYMPOSIUM ON MULTIMEDIA AND THE WEB*, 9., 2015, Manaus. **Anais [...]**. Manaus: UFAM, 2015.

BRIET, S. **What is Documentation**. Oxford: Scarecrow Press, 2006.

BROOKES, B.C. The Foundations of Informations Science. **Journal of Information Science**, [s. l.], v. 2, p. 209-221, 1980.

CAPURRO, R.; HJORLAND, B. O conceito de informação. **Perspect. ciênc. inf.**, Belo Horizonte, v.12, n.1, jan/abr. 2007. Disponível em: <https://doi.org/10.1590/S1413-99362007000100012>. Acesso em: 8 jan. 2022.

CARRANO, P. C. R. Redes sociais de internet numa escola de ensino médio: entre aprendizagens mútuas e conhecimentos escolares. **Perspectiva**, Florianópolis, v. 35, n. 2, p. 395-421, abr./jun. 2017. Disponível em: <https://periodicos.ufsc.br/index.php/perspectiva/article/view/2175-795X.2017v35n2p395>. Acesso em: 7 jan. 2022.

CARVALHO, A. D. S., SANTOS, S. J. A Internet como Espaço Econômico e de Segurança: espaços econômicos e espaços de segurança. **Observare**, Lisboa, p. 693-726, 2017.

CHANDER, A.; KRISHNAMURTHY, V. The myth of platform neutrality. **Georgetown Law Technology Review**, [s.l.], v. 2, n. 2, p. 400-416, 2018.

CHIAPPIN, J. R.; LEISTER, C. Experimento Mental I: O Problema Da Emergência Da Cooperação e o Modelo Da Tragédia Dos Comuns - Hobbes, Os Fundamentos Do Estado e a Emergência Do Indivíduo Como Pessoa. **EALR**, [s. l.], v. 7, n. 1, p. 291 - 315, jan-jun, 2016.

COLL-VIENTENT, R.; BERNAL CRUZ, F. J. **Curso de documentación**. Madrid: Sossat, 1990.

DAHLBERG, I. Begriffstheoretische Grundlagen der Klassifikation. Tradução de Astério Tavares Campos. **R. Bibliotecon.**, Brasília, v. 6, n. 1, jan./jun. 1978.

DIAS, E. W. Organização do conhecimento no contexto de bibliotecas tradicionais e digitais. In: NAVES, M. M. L.; KURAMOTO, H. (orgs.). **Organização da informação: princípios e tendências**. Brasília: Briquet de Lemos Livros, 2016, p. 62-75.

DIZARD JUNIOR, Wilson. **A Nova Mídia: A comunicação de massa na era da informação**. 3. ed. Rio de Janeiro: Jorge Zahar, 2000.

FERREIRA JUNIOR., H. S. Otlet realizador ou visionário? o que existe em um nome?. **Ciência da Informação**, [s. l.], v. 35, n. 2, 2006. DOI: 10.18225/ci.inf..v35i2.1136 Acesso em: 12 jan. 2022.

FERREIRA, D. T. Profissional da informação: perfil de habilidades demandadas pelo mercado de trabalho. **Ci. Inf.**, Brasília, v. 32, n. 1, jan./abr. 2003.

FOUCAULT, M. **Vigiar e punir**. Petrópolis: Vozes, 1996.

GABRIEL, M. **Marketing na era digital: conceitos, plataformas e estratégias**. São Paulo: Novatec, 2010.

- GARCIA-CEJA E. *et al.* Mental health monitoring with multimodal sensing and machine learning: a survey. **Pervasive Mob Comput.**, [s. l.], v. 51, n. 26, 2018.
- GERHARDT, T. E.; SILVEIRA, D. T. **Métodos de pesquisa**. Porto Alegre: Editora da UFRGS, 2009.
- GIL, A. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2008.
- GODSAY, M. **The Process of Sentiment Analysis**: A Study. *International Journal of Computer Applications*, [s. l.], v. 126, n. 7, set. 2015.
- GRUHL, D. *et al.* The predictive power of online chatter. *In: INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY IN DATA MINING*. 12., 2005. **Anais [...]** Chicago: ACM, ago. 2005. p. 78-87.
- HATZIVASSILOGLOU, V.; MCKEOWN, K. Predicting the semantic orientation of adjectives. *In: ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS AND 8TH CONFERENCE OF THE EUROPEAN CHAPTER OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS*. 25., Madri, 1997. **Anais [...]** Madri: ALC, 1997. p. 174-181.
- HERRERA, A. H. **Archivística general**: teoría y práctica. Sevilla: Disputación de Sevilla, 1991.
- KENNEDY, H. Perspectives on sentiment analysis. **Journal of Broadcasting & Electronic Media**, [s. l.], v. 56, n. 4, p. 435-450, 2012.
- KUCUKTUNC, O. *et al.* A large-scale sentiment analysis for Yahoo! answers. *In: INTERNATIONAL CONFERENCE ON WEB SEARCH AND DATA MINING*. 50., Seattle, 2012. **Anais [...]** Seattle: WA, 2012. p. 633-642.
- LEACHER, C. **How Amazon automatically tracks and fires warehouse workers for 'productivity'**. The Verge, 2019. Disponível em: <https://www.theverge.com/2019/4/25/18516004/amazon-warehouse-fulfillment-centersproductivity-firing-terminations>. Acesso em: 10 jan. 2022.
- LI, G.; LIU, F. Application of a clustering method on sentiment analysis. **Journal of Information Science**, [s. l.], v. 38, n. 2, p. 127-139, 2012.
- LIMA, P. Revisão sistemática sobre uso de ontologia para análise de sentimento em conteúdo da Web. **Ci.Inf.**, Brasília, v.44 n.3, p. 430-443, set./dez. 2015.
- LIU, B. Sentiment Analysis and Opinion Mining. **Synthesis Lectures On Human Language Technologies**, Colorado, v. 5, n. 1, p. 1-167, 23 mai. 2012.
- LIU, H; COCEA, M.; GEGOV, A. Interpretability of computational models for sentiment analysis. **Sentiment Analysis and Ontology Engineering**, Springer, Cham, p. 199-220, 2016.
- LIU, Y.; CHEN, Y. A two-phase sentiment analysis approach for judgement prediction. **Journal of Information Science**, [s. l.], v. 1, n. 14, 2017.

MACIEL, M. L.; ALBAGLI, S. **Informação, conhecimento e poder**: mudança tecnológica e inovação social. Rio de Janeiro: Garamond, 2011.

MACK, R. W.; BLOSE, J. B.; PAN, B. Believe It or Not: Credibility of Blogs in Tourism. **Journal of Vacation Marketing**, [s. l.], v. 14, n. 2, p. 133-44, 2008.

MALINI, F.; CIARELLI, P.; MEDEIROS, J. O sentido político em redes sociais: big data, algoritmos e as emoções nos tweets sobre o impeachment de Dilma Rousseff. **Liinc em Revista**, Rio de Janeiro, v.13, n.2, p. 323-342, nov. 2017.

MAMBU, M. **Trust and the implementation of a technologically mediated social credit system**. New York: Springer, 2019.

MANTYLA, M.; GRAZIOTIN, D.; KUUTILA, M. The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. **Computer Science Review**, v.27, fev. 2018.

MARCUS; G. **Moral Machines**. Disponível em: <https://users.manchester.edu/Facstaff/SSNaragon/Online/100-FYS-F15/Readings/Marcus,%20Moral%20Machines.pdf>. 24. nov. 2012. Acesso em: 4 jan. 2022.

MARTELETO, R. M. Análise de Redes Sociais: aplicação nos estudos de transferência da informação. **Ci. Inf.**, Brasília, v. 30, n. 1, p. 71-81, jan./abr. 2001.

MARTINS, R. F.; PEREIRA, A.; BEENEVUTO, F. An approach to sentiment analysis of Web applications in Portuguese. *In*: BRAZILIAN SYMPOSIUM ON MULTIMEDIA AND THE WEB, 5., 2015, Manaus. **[Anais]** ... New York: ACM, 2015. p. 105- 112.

MEDHAT, W.; HASSAN, A; KORASHY, H. Sentiment analysis algorithms and applications: a survey. **Ain Shams Engineering Journal**. v. 5, n. 4, p.1093-1113, 2014.

MICHEL. M. **Metodologia e Pesquisa Científica em ciências sociais**: um guia prático para acompanhamento da disciplina e elaboração de trabalhos monográficos. São Paulo: Atlas, 2009.

MOOERS, C. Zetocoding applied to mechanical organization of knowledge. **American Documentation**, [s. l.], v. 2, p. 20-32, 1951.

MOREIRA, V. S. *et al.* Análise de Sentimentos: Comparando o uso de ferramentas e a análise humana. *In*: BRAZILIAN SYMPOSIUM ON INFORMATION SYSTEMS, 12., 2016, Florianópolis. **Anais [...]** Florianópolis: UFSC, 2016 p. 441-448.

MORGADO, M. G. **Comportamento do consumidor online**: perfil, uso da Internet e atitudes. São Paulo: FGV, 2003.

NASCIMENTO, P. **Dicionário de polaridades para apoio a análise de sentimento**. Rio de Janeiro: COPRE UFRJ, 2014.

- NORONHA, D. P.; MARICATO, J. de M. Estudos métricos da informação: primeiras aproximações. **R. Eletr. Bibliotecon. Ci. Inf.**, Florianópolis, n. esp., 2008.
- NOVELLINO, M. S. F. Instrumentos e metodologias de representação da informação. **Inf. Inf.**, Londrina, v.1, n.2, p.37-45, jul./dez. 1996.
- ORTIGOSA, A.; MARTÍN, J.; CARRO, R. **Sentiment analysis in Facebook and its application to e-learning**. [S. l.]: Elsevier, 2013.
- PANG, B., LEE, L. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. **ACL**, [s. l.], p. 271-278, 2004. Disponível em: <https://arxiv.org/abs/cs/0409058>. Acesso em: 10 jan. 2022.
- PANG, B., LEE, L. Opinion mining and sentiment analysis. Foundations and Trends, **Information Retrieval**, [s.l.], v. 2, n. 2, p. 1-135, 2008.
- PATIL, M. K. *et al.* A review on sentiment analysis in psychomedical diagnosis. **OJPAS**, Maharashtra, jun. 2020.
- PIEIDADE, M. A. **Introdução à teoria da classificação**. Rio de Janeiro: Interciência, 1977.
- PIEIDADE, M. A. **Introdução à teoria da classificação**. 2. ed. rev. e aum. Rio de Janeiro: Interciência, 1983.
- PINTO, L. P.; FIDELIS, M. B. O uso social da informação como vetor de fortalecimento do mundo social da vida. ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 13., Rio de Janeiro, 2012. **Anais [...]** Rio de Janeiro: Fiocruz, 2012.
- PORIA, S. *et al.* Context-dependent sentiment analysis in user-generated videos. *In*: ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS. 55., Vancouver, 2017. **Anais [...]** Vancouver: ALC, 2017. p. 873-883.
- PORTO, R. *et al.* Análise de sentimento sobre veículos em redes sociais. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 13., Rio de Janeiro, 2012. **Anais [...]** Rio de Janeiro: Fiocruz, 2012.
- PRODANOV, C. **Metodologia do trabalho científico**: métodos e técnicas da pesquisa e do trabalho acadêmico. Novo Hamburgo: Feevale, 2013.
- RECUERO, R. Mapeando Redes Sociais na Internet através da Conversação Mediada pelo Computador. *In*: HETKOWSKI, T. M.; NASCIMENTO, A. D. (orgs.). **Educação e Contemporaneidade**: pesquisas científicas e tecnológicas. Salvador: EDUFBA, 2009, p. 251-274.
- RICHMOND, J. Spies in ancient Greece. **Greece & Rome**, [s.l.], v. 45, n.1, p. 1-18, 1998.

RODRIGUES, C. A. *et al.* **Mineração de Opinião: Análise de Sentimentos**. 2013. Disponível em:

<http://www.inf.ufsc.br/~luis.alvares/INE5644/MineracaoOpiniao.pdf>. Acesso em: 14 jan. 2022.

SANT'ANA, R. Ciclo de vida dos dados: uma perspectiva a partir da ciência da informação. **Inf. Inf.**, Londrina, v. 21, n. 2, maio/ago., 2016.

SANTOS, L. L. *et al.* **Twitter, análise de sentimento e desenvolvimento de produtos**: Quanto os usuários estão expressando suas opiniões? *Prisma.com*. n.13, 2010. Disponível em:

<http://ojs.letras.up.pt/ojs/index.php/prismaom/article/view/2005/1842>. Acesso em: 07 jan. 2022.

SHAYAA, S. *et al.* Sentiment Analysis of Big Data: Methods, Applications, and Open Challenges. **IEEE**, [s. l.], v. 6, 2018. Disponível em:

<https://ieeexplore.ieee.org/abstract/document/8399738>. Acesso em 13 jan. 2022.

SIEBRA, S de A. Curadoria Digital: uma área em expansão. **Archeion Online**, João Pessoa, v.6, n.2, p. 01-06, jan./jun. 2019. Disponível em:

<http://periodicos.ufpb.br/ojs2/index.php/archeion>. Acesso em: 11 jan. 2022.

SILVA, N. R.; LIMA, D.; BARROS, F. **SAPair**: Um Processo de Análise de Sentimento. São Paulo: USP, 2012. Disponível em:

<http://sites.labic.icmc.usp.br/wti2012/artigos/105283.pdf> . Acesso em: jan. 2022.

SILVA, P. M. S. E.; SILVA, A. K. A. Aplicação das técnicas de análise de sentimento em redes sociais informais: revisão bibliográfica sistemática. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 20., Florianópolis, 2019. **Anais [...]** Florianópolis: UFSC, 2019 p. 441-448. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/124085>. Acesso em: 10 mai. 2021.

SOUZA, R., CAFÈ, L. Análise de sentimento aplicada ao estudo de letras de música. **Informação & Sociedade**, [s. l.], v. 28, n.3, p. 275-286, 2018.

SOUZA JUNIOR, M. B. M.; MELO, M. S. T.; SANTIAGO, M. E. A análise de conteúdo como forma de tratamento dos dados numa pesquisa qualitativa em Educação Física escolar. **Movimento**, Porto Alegre, v. 16, n. 03, p. 31-49, 2010.

SYMEONIDIS, S. **5 Things You Need to Know about Sentiment Analysis and Classification**. KDnuggets, 2018. Disponível em:

<https://www.kdnuggets.com/2018/03/5-things-sentiment-analysis-classification.html> . Acesso em: 13 jun. 2019 às 13:36h.

TARAPANOFF, K.; ARAÚJO, Rogério H. A. Jr. CORMIER, Patricia M. J. Sociedade da informação e inteligência em unidades de informação. **Ci. Inf.**, Brasília, v. 29, n. 3, 2000.

TEIXEIRA, C. **Análise de sentimento dos usuários do twitter em relação à atual situação política do Brasil**. 2019. Dissertação (Mestrado) – Universidade Federal de Ouro Preto, Ouro Preto, 2019.

TOFFLER, A. **A terceira onda**. 32. ed. Rio de Janeiro: Record, 2014.

TRISTÃO, A. M. D. FACHIN, G. R. B. ALARCON. O. E. Sistema de classificação facetada e tesouros: instrumentos para organização do conhecimento. **Ci. Inf.**, Brasília, v. 33, n. 2, p. 161-171, maio/ago. 2004.

WANG, Z. **Political trust in China: Forms and causes, Legitimacy: Ambiguities of Political Success of Failure**. Michigan: University of Michigan, 2005. Disponível em: https://www.academia.edu/1093737/Political_trust_in_China_Forms_and_causes. Acesso em: 12 jan. 2022.

WESTCOTT, N. **Digital Diplomacy: The Impact of the Internet on International Relations**. Oxford: Oxford Internet Institute, 2008. Disponível em: <https://www.oii.ox.ac.uk/archive/downloads/publications/RR16.pdf>. Acesso em: 11 jan. 2022.

WILSON, T. D. **Information Needs and Uses: fifty years of progress?** London: Aslib, 1981.

WULFF, J.; HARDT, D. Can you trust online ratings? Evidence of systematic differences in user populations. *In*: CONFERÊNCIA EUROPEIA SOBRE SISTEMAS DE INFORMAÇÃO. 22., Tel Aviv-Yafo, 2014. **Anais [...]** Tel Aviv-Yafo: David Intercontinental Hotel, 2014.

YOKOTA, H; SEYHAN, S. **The emergence of the Social Credit System is linked to the government's fear of the digitalization of things**. Louvain-la-Neuve: Université catholique de Louvain, 2018. Disponível em: https://www.academia.edu/38159911/The_emergence_of_Social_Credit_System. Acesso em: 10 jan. 2022.

ZUBOFF, S. **The Age of Surveillance Capitalism**. São Paulo: Intrínseca, 2019.