

UNIVERSIDADE FEDERAL DO ABC
Programa de Pós-Graduação em Ciências Humanas e Sociais

CARLA OLIVEIRA SANTOS

**O USO DO APRENDIZADO DE MÁQUINA NAS CAMPANHAS
POLÍTICAS NO BRASIL**

São Bernardo do Campo - SP

2019

CARLA OLIVEIRA SANTOS

**O USO DO APRENDIZADO DE MÁQUINA NAS CAMPANHAS
POLÍTICAS NO BRASIL**

Dissertação apresentada ao Programa de Pós-Graduação em Ciências Humanas e Sociais da Universidade Federal do ABC, como parte dos requisitos para obtenção do título de Mestre em Ciências Humanas e Sociais.

Orientador: Professor Doutor Sérgio Amadeu da Silveira

São Bernardo do Campo - SP

2019

Sistema de Bibliotecas da Universidade Federal do ABC
Elaborada pelo Sistema de Geração de Ficha Catalográfica da UFABC
com os dados fornecidos pelo(a) autor(a).

Santos, Carla Oliveira

O uso do aprendizado de máquina nas campanhas políticas no Brasil / Carla Oliveira Santos. — 2019.

206 fls. : il.

Orientador: Sergio Amadeu da Silveira

Dissertação (Mestrado) — Universidade Federal do ABC,
Programa de Pós-Graduação em Ciências Humanas e Sociais,
São Bernardo do Campo, 2019.

1. Inteligência Artificial. 2. aprendizado de máquina. 3.
marketing. 4. campanhas políticas. 5. monitoramento. I.
Silveira, Sergio Amadeu da. II. Programa de Pós-Graduação
em Ciências Humanas e Sociais, 2019. III. Título.

Este exemplar foi revisado e alterado em relação à versão original, de acordo com as observações levantadas pela banca no dia da defesa, sob responsabilidade única do(a) autor(a) e com a anuência do(a) orientador(a).

São Bernardo do Campo/SP 05 de agosto de 2019

Assinatura do(a) autor(a): Carla Oliveira Santos


Assinatura do(a) orientador(a): 



MINISTÉRIO DA EDUCAÇÃO
Fundação Universidade Federal do ABC
Programa de Pós-Graduação em Ciências Humanas e Sociais
Avenida dos Estados, 5001 – Bairro Santa Terezinha – Santo André – SP
CEP 09210-580 · Fone: (11) 4996-0017
ppg.chs@ufabc.edu.br

FOLHA DE ASSINATURAS


Assinaturas dos membros da Banca Examinadora que avaliou e aprovou a Defesa de Dissertação de Mestrado da candidata Carla Oliveira Santos, realizada em 1 de julho de 2019:



Prof.(a) Dr.(a) **Sérgio Amadeu da Silveira** (Universidade Federal do ABC) – Presidente



Prof.(a) Dr.(a) **Lucas de Almeida Pereira** (Universidade Federal do ABC) – Membro Titular



Prof.(a) Dr.(a) **Demi Getschko** (Pontifícia Universidade Católica de São Paulo) – Membro Titular

Prof.(a) Dr.(a) **Claudio Luis de Camargo Penteado** (Universidade Federal do ABC) – Membro Suplente

Prof.(a) Dr.(a) **Henrique Parra** (Universidade Federal de São Paulo) – Membro Suplente

Esta pesquisa é dedicada a todos aqueles que, assim como eu, acreditam no poder da educação e consideram o conhecimento um bem comum.

*“Nunca se vence uma guerra lutando sozinho...”
(Raul Seixas)*

AGRADECIMENTOS

Conseguir equilibrar o mestrado e o trabalho é uma tarefa árdua que requer muita energia e jogo de cintura. Porém, ao olhar a trajetória percorrida até aqui, percebo o quanto todo o esforço valeu a pena, não só pelo conhecimento adquirido e transmitido, mas, principalmente, pelos laços de amizade (antigos e novos) fortalecidos.

Por isso, agradeço a todos que fizeram parte desta caminhada. Agradeço à Universidade Federal do ABC e ao programa de Ciências Humanas e Sociais pela oportunidade. Em especial, aos professores Demi Getschko e Lucas Pereira pelas importantes contribuições desde a banca de qualificação. Ao meu orientador, Sergio Amadeu da Silveira, por todo apoio e por ter me mostrado a importância de humanizar a tecnologia, a você Sérgio, “vida longa e próspera”.

À CAS Tecnologia, empresa onde trabalho, por ter me liberado no horário comercial para cursar as disciplinas e demais atividades complementares necessárias para a conclusão deste mestrado.

Aos colegas e amigos da UFABC e do Laboratório de Tecnologias Livres (LabLivre), em especial, à Joyce Souza, Débora Machado e Rodolfo Avelino por todo suporte, carinho e incentivo.

Aos amigos da área de tecnologia, Leonardo Arenzon e Yumi Ambriola, pelo esclarecimento de dúvidas e revisões técnicas. À minha amiga e jornalista, Lourdes Guimarães pelas dicas de escrita. Às amigas, Carmen Veiga e Vera Braz por todo o apoio na qualificação.

A todos os entrevistados pela contribuição, disponibilidade e atenção, além das informações disponibilizadas, sem isto esta pesquisa não teria tanto embasamento. Em especial ao Alexander Schmitz-Kohlitz, por todo esclarecimento sobre os subcampos da Inteligência Artificial e suas especificidades.

Aos meus pais, João Pedro e Filomena Rita, que são minha vida. Às minhas irmãs, Cassiana e Jorciane que são um pedaço de mim. Ao meu sobrinho Heitor, que me incentiva com tanta curiosidade e alegria e ao meu tio Francisco que desde criança é meu mentor.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

RESUMO

Esta dissertação discute o uso do aprendizado de máquina nas campanhas políticas no Brasil. Começa com a história da Inteligência Artificial (IA) na qual são descritos os principais atores, marcos da IA, subcampos, definição de aprendizado de máquina, tipos de aprendizado, algoritmos, tecnologias e aplicações. Em seguida, contempla um referencial teórico descrevendo a passagem das sociedades disciplinares para as sociedades de controle, assim como a relação entre as sociedades de controle, marketing e tecnologias da informação e comunicação. Debate as diferenças entre dispositivos de manipulação e dispositivos de modulação e apresenta exemplos de como o aprendizado de máquina e as pegadas digitais dos usuários podem ser utilizados como dispositivo de modulação do comportamento humano. Na sequência, traz um referencial teórico sobre marketing político, marketing eleitoral e marketing *one to one* (individualizado). Examina o papel da informação e da pesquisa em campanhas políticas e o uso da Internet e redes sociais como fontes de pesquisa. Apresenta a definição de campanha eleitoral e a utilização das mídias de massas, *Web 2.0*, segmentação e microssegmentação nessas campanhas. Aborda o uso de aprendizado de máquina nas campanhas políticas dos candidatos à presidência dos EUA Barack Obama, em 2012, e Donald Trump em 2016. Antes do resultado da pesquisa empírica e apresentação dos atores envolvidos, é feita uma síntese das últimas campanhas eleitorais no Brasil. Por fim, com base nos resultados da pesquisa, considera-se que o aprendizado de máquina está sendo utilizado em campanhas políticas no Brasil para monitoramento dos eleitores, análise de sentimento para gestão de reputação do candidato e segmentação de grande volume de dados para direcionamento de discurso personalizado.

Palavras-chave: Inteligência Artificial; aprendizado de máquina; marketing; campanhas políticas; monitoramento.

ABSTRACT

This dissertation discusses the use of machine learning in political campaigns in Brazil. It begins with the history of Artificial Intelligence (IA) where the main actors, IA milestones, subfields, definition of machine learning, types of learning, algorithms, technologies and applications are described. It then contemplates a theoretical reference describing the transition from disciplinary societies to control societies, as well as the relationship between control societies, marketing and information and communication technologies. It discusses the differences between manipulation devices and modulation devices. It also presents examples of how machine learning and user's digital footprints can be used as modulation device of human behavior. It also brings theoretical reference on political marketing, electoral marketing and marketing one to one (individualized). It discusses information and research's roles in political campaigns and the use of the Internet and social networks as sources of research. It presents the definition of electoral campaign, the use of mass media, Web 2.0, segmentation and microsegmentation in them. It addresses the use of machine learning in the US presidential races of Barack Obama in 2012 and Donald Trump in 2016. Before the result of the empirical research and presentation of the involved actors are presented, there is a synthesis about the last electoral campaigns in Brazil. Finally, based on the research's results, it is considered that machine learning is being used in the political campaigns in Brazil for voter's monitoring, sentiment analysis for candidate's reputation management and high-volume segmentation of data for discourse targeting custom.

Keywords: Artificial Intelligence; machine learning; marketing; political campaigns; monitoring.

LISTA DE FIGURAS

Figura 1 - Exemplo de Busca DFS e BFS	24
Figura 2 - Exemplo de Rede Semântica	25
Figura 3 - <i>Booms</i> da IA	28
Figura 4 - Exemplos de subcampos da IA.....	31
Figura 5 - Inteligência Artificial, <i>Machine Learning</i> e <i>Deep Learning</i>	32
Figura 6 - Imagem de um gato identificada pelo algoritmo do <i>Google</i>	39
Figura 7 - Imagens geradas pelo <i>Inceptionism</i>	40
Figura 8 - Imagem dos Sonhos Urbanos, da exposição “Consciência Cibernética [?]”	41
Figura 9 - Tipos de Aprendizado de Máquina e seus algoritmos	46
Figura 10 - Algoritmos de Aprendizado de Máquina mais utilizados	47
Figura 11 - Tecnologias de Aprendizado de Máquina mais utilizadas	49
Figura 12 - Tecnologias <i>open source</i> mais utilizadas no Aprendizado de Máquina.....	50
Figura 13 - Tecnologias do Aprendizado de Máquina	53
Figura 14 - Aplicações agrupadas por subcampos da IA	57
Figura 15 - Ferramentas de Aprendizado de Máquina do <i>Facebook</i>	74
Figura 16 - Relação entre os atores envolvidos com a <i>Cambridge Analytica</i>	102
Figura 17 - Actantes selecionados	110
Figura 18 - Actantes pesquisados.....	112
Figura 19 - Slide com os traços de personalidade OCEAN	117
Figura 20 - Slide com pessoas parecidas com resultado do OCEAN diferente	117
Figura 21 - Slide com anúncio para pessoa com alta conscienciosidade	119
Figura 22 - Slide com anúncio para pessoa com alta extroversão	119
Figura 23 - Slide Transformando <i>Big Data</i> em <i>Smart Data</i>	120
Figura 24 - Slide Mensagens segmentadas de acordo com os traços do <i>Big Five</i>	120
Figura 25 - Palestra “O uso da informação como vantagem competitiva”	143
Figura 26 - Estágios da maturidade analítica nas empresas de marketing.....	144
Figura 27 - Complexidade e ações de cada tipo de processo analítico	145
Figura 28 - Etapas do tratamento dos dados	146
Figura 29 - Etapas do processo analítico	146
Figura 30 - Soluções levantadas por empresa/ <i>startup</i>	151
Figura 31 - Soluções nacionais versus soluções de outros países.....	152
Figura 32 - Aplicações das soluções analisadas.....	154
Figura 33 - Algoritmos utilizados nas soluções	155
Figura 34 - Metodologia seguida para realizar um trabalho de monitoramento	170
Figura 35 - <i>Brandwatch</i> volume por <i>hashtags</i> na entrevista do Bolsonaro	172

Figura 36 - <i>Brandwatch</i> nuvem de palavras por frases na entrevista do Bolsonaro	173
Figura 37 - <i>Brandwatch</i> evolução de menções na entrevista do Bolsonaro	173
Figura 38 - <i>Brandwatch</i> informações demográficas na entrevista do Bolsonaro	174
Figura 39 - <i>Brandwatch</i> menções por região na entrevista do Bolsonaro.....	175
Figura 40 - <i>Brandwatch</i> volume por <i>hashtags</i> na entrevista da Marina Silva	175
Figura 41 - <i>Brandwatch</i> nuvem de palavras por frases na entrevista da Marina Silva	176
Figura 42 - <i>Brandwatch</i> evolução de menções na entrevista da Marina Silva	177
Figura 43 - <i>Brandwatch</i> informações demográficas na entrevista da Marina Silva	178
Figura 44 - <i>Brandwatch</i> menções por região na entrevista da Marina Silva	178
Figura 45 - <i>Netbase</i> principais métricas e tendências no monitoramento do Rodrigo Maia	180
Figura 46 - <i>Netbase</i> onde as pessoas estão falando no monitoramento do Rodrigo Maia .	181
Figura 47 - <i>Netbase</i> o que as pessoas estão falando no monitoramento do Rodrigo Maia	181
Figura 48 - <i>Netbase</i> quem são os autores no monitoramento do Rodrigo Maia	182
Figura 49 - <i>Netbase</i> informações demográficas no monitoramento do Rodrigo Maia.....	183
Figura 50 - Nuvem de palavras do site Polis Consulting	184
Figura 51 - Soluções da Tail	187
Figura 52 - Nuvem de palavras do site Tail	188
Figura 53 - Lista de parceiros da Tail.....	192
Figura 54 - Nova marca da Elife em 2019 e antiga de 2009.....	193
Figura 55 – Soluções da Elife	194
Figura 56 - Gráfico com resultado da análise de imagens no Carnaval	195
Figura 57 - Exemplo de monitoramento de perfis psicográficos	197
Figura 58 - Nuvem de palavras do site da Elife.....	199
Figura 59 - Nuvem de palavras do site da Stilingue	201
Figura 60 - Informações sobre a MarketData	204
Figura 61 - Soluções da MarketData.....	204
Figura 62 - Nuvem de palavras do site da MarketData	206
Figura 63 - Política de Privacidade: Compartilhamento de informações com terceiros	207

LISTA DE TABELAS

Tabela 1 - Resultados do teste do ILSVRC.....	38
Tabela 2 - Evolução das Redes Neurais Artificiais até o <i>Deep Learning</i>	42
Tabela 3 - Porcentagem de uso do Aprendizado de Máquina por segmento	55
Tabela 4 - Fases do Marketing Político no Brasil	79
Tabela 5 - Tipos de segmentação.....	93

SUMÁRIO

1	INTRODUÇÃO	15
2	HISTÓRIA DA INTELIGÊNCIA ARTIFICIAL E SUA EVOLUÇÃO	21
2.1	Inteligência Artificial e seus Subcampos	30
2.2	Aprendizado de Máquina (<i>Machine Learning</i>)	33
2.2.1	Aprendizado Profundo (<i>Deep Learning</i>)	37
2.2.2	Processamento de Linguagem Natural (<i>Natural Language Process - NLP</i>)	44
2.3	Tipos de Aprendizado de Máquina e seus Algoritmos	45
2.4	Tecnologias (<i>Hardware, Frameworks</i> e Linguagens de Programação)	49
2.5	Aplicações do Aprendizado de Máquina	54
3	SOCIEDADES DE CONTROLE, MARKETING E TECNOLOGIAS DA INFORMAÇÃO E COMUNICAÇÃO	60
3.1	Aprendizado de Máquina e Modulação do Comportamento Humano	63
4	MARKETING POLÍTICO	76
4.1	Marketing Eleitoral e Marketing <i>One to One</i> (Individualizado)	80
4.2	O Papel da Informação e da Pesquisa na Campanha Política	81
4.2.1	A Internet e as Redes Sociais como Estratégias de Pesquisa Eleitoral	83
4.3	Campanhas Eleitorais	86
4.4	Mídias de Massa e <i>Web 2.0</i> nas Campanhas Eleitorais	89
4.5	Segmentação	91
4.6	Microsegmentação	92
5	APRENDIZADO DE MÁQUINA NAS CAMPANHAS POLÍTICAS	97
5.1	Campanha Eleitoral de Barack Obama	97
5.2	Campanha Eleitoral de Donald Trump	98
6	APRENDIZADO DE MÁQUINA NAS CAMPANHAS POLÍTICAS NO BRASIL	105
6.1	Uma Síntese das Últimas Campanhas Eleitorais no Brasil	105
6.2	Metodologia da Pesquisa	108
6.2.1	Métodos de Pesquisa	110
6.3	Consultor de Marketing Político	114
6.3.1	Matérias Veiculadas sobre Torreta e a <i>Cambridge Analytica</i>	115
6.3.2	Uso de Dados Pessoais como Instrumento de Campanha Eleitoral e a Persuasão da Opinião Pública	116
6.3.3	Principais Pontos da Entrevista	121
6.3.4	O que sua marca pode aprender com as Campanhas de Trump e Bolsonaro?	127

6.4	Empresas/Startups	128
6.4.1	Polis Consulting	128
6.4.2	Tail	132
6.4.3	Elife	136
6.4.4	Stilingue.....	139
6.4.5	MarketData.....	143
6.5	Especialistas.....	146
6.5.1	Ambriola	147
6.5.2	Ikeda	150
6.6	Analise das Soluções Levantadas	151
7	CONSIDERAÇÕES FINAIS.....	156
8	REFERÊNCIAS BIBLIOGRÁFICAS.....	160
	APÊNDICE	169
	APÊNDICE A - Informações sobre as Soluções da Polis Consulting	170
	APÊNDICE B - Informações sobre as Soluções da Tail	186
	APÊNDICE C - Informações sobre as Soluções da Elife	193
	APÊNDICE D - Informações sobre as Soluções da Stilingue	201
	APÊNDICE E - Informações sobre as Soluções da MarketData	204

1 INTRODUÇÃO

As Tecnologias da Informação e Comunicação (TICs) abriram espaço para a criação de uma variedade de códigos e aparatos tecnológicos que alteraram comportamentos e geraram novas maneiras de interação entre os coletivos. Com o desenvolvimento cada vez mais acelerado das TICs a vida social está passando por mudanças significativas.

De acordo com Flusser (2007, p. 54), as informações que invadem nosso mundo são imateriais (*softwares*) e essas “não coisas” estão cada vez mais no centro da economia, da cultura, da vida social e até mesmo da política. Nessa nova ordem, os dados também se tornaram imateriais. Eles passaram do papel para o disquete e, mais recentemente, para o armazenamento em nuvem (*cloud storage*).

Para Cavalheiro e Brandão (2017), a Internet interligou o mundo e com isso “os processos tornaram-se instantâneos e a vida mais dinâmica” (2017, p. 87). Essas “mudanças nos meios e nas técnicas” modificaram não só as formas de trabalho, mas também “as relações entre as pessoas, desencadeando transformações na esfera política, nos comportamentos e desejos” (2017, p.87).

A Internet, segundo Cavalheiro e Brandão (2017), permite que atores coletivos se expressem e as mensagens circulem sem burocracia ou hierarquia contribuindo para a criação de “movimentos autônomos” (2017, p. 88). Para estes autores, essas novas tecnologias representam um papel importante na independência individual e coletiva, pois permitem que todos possam ser, não só receptores, como também atores, alterando as “relações hierárquicas de transmissão de informações, antes restrita a grandes veículos” (2017, p. 88).

Por outro lado, a rapidez com que as mensagens circulam nessa rede provoca alterações nos comportamentos e tornam as diferenças mais visíveis, principalmente no campo político. Perante essa multiplicidade de vozes, são desenvolvidas estratégias para que determinadas ideias prevaleçam no lugar de outras. Essas técnicas de persuasão, utilizadas ao longo da história, são aperfeiçoadas com o uso de algoritmos para impulsionar postagens, notícias e campanhas publicitárias. Tais estratégias, aplicadas na venda de produtos e no marketing, também vêm sendo empregadas para influenciar a opinião pública (CAVALHEIRO e BRANDÃO, 2017).

Cavalheiro e Brandão (2017) alertam que “a instantaneidade com que mensagens são visualizadas, editadas, apagadas, comentadas ou compartilhadas”

(2017, p. 97) contribui para um ambiente de incertezas, em que erros jornalísticos e notícias falsas (*fake news*) alcançam rápida repercussão e geram, principalmente no campo político, conflitos e desinformação.

Os avanços das TICs trouxeram uma novidade para o ambiente político: ferramentas capazes de minerar, analisar e agrupar dados comportamentais e estruturá-los em uma base de dados para o desenvolvimento de tudo que diz respeito à subjetividade e à emoção. Segundo Cavalheiro e Brandão (2017), “os fluxos de desejo são captados” por essas ferramentas e direcionados de modo distribuído pelos meios de comunicação, fazendo prevalecer “os interesses de poucos, mas dando a impressão de que é o desejo de muitos” (2017, p. 98). Nesse cenário, é aceito que as pessoas sejam monitoradas e seus dados transformados em matéria-prima para algoritmos empregados em inúmeras finalidades, sem qualquer conhecimento e permissão das pessoas que tiveram seus dados coletados (GUARESCHI et al, 2017).

De acordo com Levy (2011) essas ferramentas ou *softwares* “reorganizam a visão de mundo de seus usuários e modificam seus reflexos mentais” (2011, p. 54). Para este filósofo, cada inovação promovida pelas TICs abre a possibilidade de novas relações entre homens e dispositivos (*hardwares* e *softwares*) e devido à importância dessas conexões as mesmas devem ser analisadas. Howard (2015) dialoga com o pensamento de Levy (2011) ao considerar que os cientistas sociais precisam buscar novas categorias, uma vez que as categorias antigas são abrangentes e não refletem as nuances sociais; eles precisam perceber o mundo atual como um sistema de relação entre pessoas e dispositivos.

De acordo com Silva (2015), o estudo do *software* é uma nova categoria de pesquisa cujo objetivo é analisar mais detalhadamente as transformações que o *software* opera nas sociedades contemporâneas, uma vez que essa tecnologia está cada vez mais interferindo na forma como trabalhamos, ensinamos, pesquisamos e consumimos.

Para Manovich (2008), o *software* é uma camada que permeia todas as áreas das sociedades contemporâneas. E para compreender as técnicas contemporâneas de poder, controle e comunicação, a análise não pode ser completa se essa camada do *software* não for considerada. Segundo ele, é importante que as disciplinas que tratam da sociedade e da cultura contemporânea levem em conta o papel do *software* e suas consequências.

A expressão Inteligência Artificial (IA) está sendo amplamente difundida. Utilizando o jargão do mercado, IA é considerada a *buzzword* do momento, ou seja, uma palavra barulhenta e esse barulho está ocorrendo não só na academia, mas principalmente no mercado e até mesmo na política. De acordo com Sugomori (2016), empresas de tecnologia como *Google* e *Facebook* estão cada vez mais adquirindo *startups*¹ de IA. Em 2014, foi anunciado que o *Google* comprou a *Deep Mind* (HAAS, 2014). Já em 2018, foi a vez do *Facebook* investir US\$ 30 milhões na aquisição da *Bloomsbury*, *startup* especializada em processamento de linguagem natural (WAKKA, 2018). Vale ressaltar que, além dessas aquisições, vem crescendo consideravelmente os investimentos destinados a esta área da Ciência da Computação.

Para o pesquisador Pedro Domingos (2017), a “Revolução Industrial automatizou o trabalho manual e a Revolução da Informação fez o mesmo com o trabalho mental” (2017, p. 33). Assim como a “Internet e o computador pessoal”, o aprendizado de máquina irá gerar grandes “mudanças econômicas e sociais” (p. 33). “Vivemos na era dos algoritmos” (2017, p. 24), observa. Mas, o que é algoritmo?

De acordo com Brookshear (2001), algoritmo é um conjunto ordenado de passos executáveis e não ambíguos que definem um processo e seu término. Todo algoritmo tem uma entrada de dados e uma saída com o resultado do processamento desses dados. Na Ciência da Computação, os algoritmos são representados pelos programas, sendo que um conjunto de programas é denominado *software*. Neste cenário, para que um computador execute um algoritmo é necessário que um ser humano (programador), por meio de uma linguagem de programação (*Java*, *Python*), escreva previamente a sequência de passos desse algoritmo.

Segundo Brookshear (2001), o algoritmo constitui uma codificação do raciocínio necessária para a resolução de um determinado problema. Foi graças à capacidade dos algoritmos para captar e transferir inteligência que as máquinas com comportamento inteligente foram desenvolvidas. No entanto, o nível de inteligência dessas máquinas se limita à inteligência que o algoritmo é capaz de suportar

¹ De acordo com Moreira (2018), *startup* é o termo utilizado para designar um grupo de pessoas trabalhando com uma ideia diferente, cujo objetivo é obter lucro. *Startup* também é sinônimo de iniciar uma empresa e colocá-la em funcionamento.

(BROOKSHEAR, 2001). Com o passar do tempo os algoritmos foram aperfeiçoados, e a geração e o armazenamento de dados aumentaram exponencialmente, aprimorando assim a capacidade de inteligência das máquinas.

Um exemplo desse tipo de aperfeiçoamento é o *machine learning*. Domingos (2017) explica que os algoritmos de *machine learning*, também conhecidos como aprendizado de máquina, é um subconjunto da Inteligência Artificial. O aprendizado de máquina “é uma tecnologia que constrói a si própria” (2017, p. 16). Ao contrário dos algoritmos tradicionais, os algoritmos de aprendizado “são artefatos que projetam outros artefatos” (2017, p. 16). “Os aprendizes transformam dados em algoritmos” e neste caso quanto mais dados (*big data*²) melhor (2017, p. 17). Em suma, com o aprendizado de máquina, Domingos considera que os computadores são capazes de criar seus próprios algoritmos.

Segundo Domingos (2017), a essência do aprendizado de máquina é a previsão de desejos, comportamentos e de como o mundo poderá ser alterado. Para ele, o aprendizado de máquina está “recriando a ciência, a tecnologia, os negócios, a política e a guerra” (2017, p. 16). O pesquisador alerta para o fato de que não podemos controlar aquilo que não compreendemos. Por isso, é importante entender o que é aprendizado de máquina e como esse tipo de tecnologia está sendo utilizado.

Diante de todo o panorama descrito até aqui, finalmente é possível apresentar o tema desta dissertação: O Uso do Aprendizado de Máquina nas Campanhas Políticas no Brasil. Esta pesquisa tomou como base as seguintes hipóteses: a) o aprendizado de máquina permite que campanhas eleitorais tenham melhor conhecimento em relação ao comportamento dos eleitores contribuindo para a segmentação e personalização do discurso e b) o uso de aprendizado de máquina é

² Para entender o conceito de *big data*, Marquesone (2017) cita três propriedades conhecidas como 3Vs (volume, variedade e velocidade): volume está relacionado à capacidade de armazenamento dos dados (terabytes, petabytes e zettabytes); variedade diz respeito aos tipos de dados e a forma como os mesmos são organizados, podendo ser estruturados (banco de dados relacionais), semiestruturados (dados com estrutura pré-definida, como JSON e XML) e não estruturados (vídeos, imagens e alguns formatos de textos) e, por fim, a velocidade com que esses dados são coletados, armazenados e utilizados. Em síntese, "*big data* faz referência não somente ao volume, mas também à variedade e à velocidade de dados, necessitando de estratégias inovadoras e rentáveis para extração de valor dos dados e aumento da percepção" (2017, p. 16).

compreendido pelo marketing político como eficiente para formar a opinião dos eleitores e realizar a gestão de reputação dos candidatos.

Para validação destas hipóteses será investigada à seguinte questão: O aprendizado de máquina está sendo utilizado em campanhas eleitorais no Brasil? Além do questionamento principal, outras questões secundárias serão consideradas: Em caso positivo, a partir de quando e como o aprendizado de máquina está sendo utilizado?

É necessário destacar que essa temática é nova e ainda não possui discussões acadêmicas mais aprofundadas a esse respeito aqui no Brasil, principalmente no campo das campanhas políticas, em especial as campanhas eleitorais. Salienta-se também que esse estudo não visa demonstrar tecnicamente como se utiliza o aprendizado de máquina, mas apresentar, por meio de pesquisas empíricas, se essa tecnologia é utilizada em campanhas eleitorais no Brasil. A partir desse objetivo geral, busca-se os seguintes objetivos específicos: a) levantar consultores de marketing político, empresas que comercializam soluções de marketing que utilizem esse tipo de tecnologia e especialistas em aprendizado de máquina.

Esta dissertação reúne 7 capítulos. São eles: 1 Introdução; 2 História da Inteligência Artificial e sua Evolução; 3 Sociedades de Controle, Marketing e Tecnologias da Informação e Comunicação; 4 Marketing Político; 5 Aprendizado de Máquina nas Campanhas Políticas; 6 Aprendizado de Máquina nas Campanhas Políticas no Brasil e 7 Considerações Finais.

O Capítulo 2 apresenta a história da Inteligência Artificial, sua evolução, subcampos, tipos de algoritmos, definições e tecnologias utilizadas. Para tanto foram levantados os principais marcos (*booms*) da IA, suas características, atores, algoritmos e suas aplicações. Foi dada maior ênfase ao aprendizado de máquina (*machine learning*) devido a sua relevância para esta pesquisa.

O capítulo 3 contempla o referencial teórico descrevendo a passagem das sociedades disciplinares para as sociedades de controle, bem como a relação entre as sociedades de controle, marketing e as TICs. Também promove um questionamento sobre as diferenças entre dispositivos de manipulação e dispositivos de modulação e, por meio de exemplos, mostra como o aprendizado de máquina, e as pegadas digitais dos usuários podem ser utilizados como dispositivo de modulação do comportamento humano.

No capítulo 4 é apresentado um referencial teórico sobre marketing político, marketing eleitoral e marketing *one to one* (individualizado). Mostra o papel da informação e da pesquisa em campanhas políticas e o uso da Internet e redes sociais como fonte de pesquisa eleitoral. Em seguida, aborda o conceito de campanha eleitoral, o papel das mídias de massas e Web 2.0 nessas campanhas, assim como a utilização de segmentação e microsegmentação.

O capítulo 5 aborda o uso de aprendizado de máquina nas campanhas políticas de Barack Obama em 2012 e Donald Trump em 2016.

O capítulo 6 é dedicado à pesquisa empírica. Para contextualizar o tema é feita uma síntese das últimas campanhas eleitorais no Brasil. Na sequência, é apresentada a metodologia da pesquisa e os métodos adotados, assim como os resultados obtidos.

O capítulo 7 faz uma análise do que foi levantado na pesquisa empírica e tece considerações tomando como base as informações coletadas.

2 HISTÓRIA DA INTELIGÊNCIA ARTIFICIAL E SUA EVOLUÇÃO

Um dos grandes objetivos da Ciência da Computação, ao longo de vários anos, é o desenvolvimento de máquinas com a mesma capacidade sensorial dos seres humanos para realizar funções e atividades de forma autônoma e inteligente.

Tanto a percepção quanto a dedução se enquadram na categoria das atividades corriqueiras que, embora naturais para a mente humana, representam intensas dificuldades para as máquinas. (BROOKSHEAR, 2001, p. 372)

O embrião da Inteligência Artificial surgiu em 1943, com o artigo “A logical calculus of the ideas immanent in nervous activity” (McCulloch e Pitts, 1943), escrito pelo psiquiatra Warren McCulloch e pelo cientista cognitivo Walter Pitts. Este artigo teve como objetivo estabelecer uma analogia entre as células nervosas e o processo eletrônico. Pode-se considerá-lo como precursor do modelo de redes neurais artificiais.

Em 1950, Alan Turing, considerado um dos pais da Ciência da Computação e da Inteligência Artificial, publicou o artigo “Computing Machinery and Intelligence” (Turing, 1950), na revista inglesa *Mind*. Neste artigo, foi proposto o seguinte questionamento: “As máquinas podem pensar?” Devido à complexidade dos termos máquina e pensar, Turing procurou a resposta para esta dúvida por meio de um jogo, o qual batizou de Jogo da Imitação (TURING, 1950), sendo que mais tarde ficou conhecido como Teste de Turing.

De acordo com Brookshear (2001), o Teste de Turing tem por objetivo identificar inteligência em uma máquina. Ele tem a seguinte proposta: um humano denominado interrogador, por meio de uma espécie de teclado, deve manter um diálogo com base em questionamentos com um ser humano e uma máquina. O computador deve ser programado para responder aos questionamentos e o humano deve responder normalmente. Se ao final do teste o interrogador não conseguir identificar quem é o humano e quem é a máquina, conclui-se que a máquina é inteligente.

Brookshear (2001) cita que um famoso exemplo para validação do Teste de Turing foi o sistema ELIZA, desenvolvido por Joseph Weizenbaum, na década de 1960. Segundo ele, este sistema foi desenvolvido para ser interativo e representar

um analista conduzindo uma entrevista psicológica com base na Teoria Rogeriana³, na qual o computador representa o papel do analista e o usuário o do paciente. O que este sistema tinha que fazer era reestruturar com base em regras definidas as declarações realizadas pelo paciente repassando-as para o computador. O objetivo de Joseph Weizenbaum era estudar a comunicação em linguagem natural e o tema da psicoterapia assumiu um papel secundário, apesar de vários psicólogos terem de fato utilizado este sistema para psicoterapia. Pode-se considerar que o ELIZA foi o primeiro programa de processamento de linguagem natural da história.

O termo Inteligência Artificial (IA) foi apresentado pela primeira vez pelo cientista da computação John McCarthy, na 2ª Conferência de Dartmouth, em 1956, nos EUA (JÚNIOR E LIMA, 2010). Isso mostra que pesquisas relacionadas a IA vêm sendo desenvolvidas desde a década de 50. Antes desta conferência, o termo Inteligência Artificial foi descrito no estudo “A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence”, de 31 de agosto de 1955, elaborado por John McCarthy, Marvin Minsky, Nathaniel Rochester e Claude Shannon (MCCARTHY et al, 1955).

McCarthy et al (1955), propuseram nesse estudo que se o processo de aprendizagem ou qualquer outra característica da inteligência fossem descritos de forma precisa, uma máquina conseguiria simulá-los. Mas, para isso acontecer seria necessário descobrir como fazer com que uma máquina pudesse compreender a linguagem natural, as abstrações, os conceitos, resolver problemas antes restritos aos seres humanos e, principalmente, conseguir se autoaperfeiçoar.

Para Halper (2017), a ideia de que as máquinas poderiam agir de forma inteligente vem desde a Grécia antiga, contudo não há um consenso sobre o que realmente a Inteligência Artificial significa. Segundo a autora, quando John McCarthy criou esse termo, ele descreveu que uma máquina seria chamada de inteligente se se comportasse de forma semelhante a um humano.

Desde então, começou-se um debate sobre o assunto, com algumas definições mais genéricas, como a capacidade de um dispositivo computacional para “aprender, raciocinar e interagir”. Também foram apresentadas definições mais

³ De acordo com Brookshear (2001), na Teoria Rogeriana é o paciente e não o psicólogo/analista que deve conduzir o diálogo durante a terapia, por este motivo os psicólogos acreditaram que um computador poderia realizar uma entrevista tão bem quanto um terapeuta.

específicas, como a competência para realizar tarefas cognitivas feitas por humanos, reconhecimento de imagem ou compreensão da linguagem natural (HALPER, 2017, p. 7).

De acordo com Sugomori (2016), estamos vivemos o terceiro marco (*boom*) da IA. O primeiro foi registrado no final dos anos 1950 e o segundo nos anos 1980. Para ele, este último tem uma diferença significativa em relação aos anteriores e essa mudança ocorre por conta do *deep learning* (aprendizado profundo) que conseguiu alcançar o que as técnicas anteriores não foram capazes de realizar. Neste *boom* atual, a máquina é capaz de descobrir características e padrões dos dados fornecidos e aprender sozinha. Com essa conquista, há uma grande possibilidade da IA deixar de ser apenas ficção científica e se tornar uma realidade.

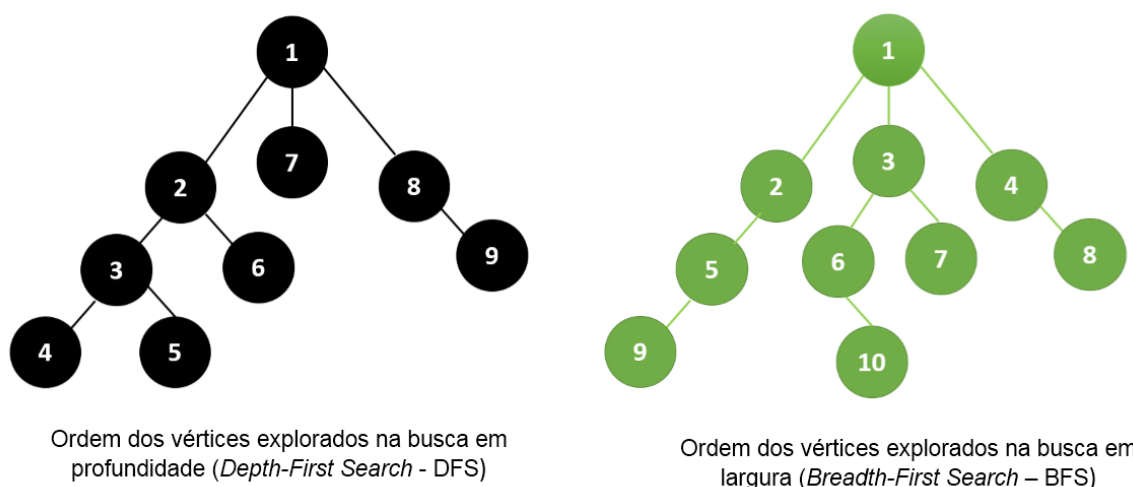
É visível o longo caminho percorrido até o estágio em que uma máquina é capaz de aprender sozinha sem a interferência humana. O primeiro *boom* da IA surgiu no final dos anos 1950 com programas de busca baseados em regras fixas. Nessa época, para que uma máquina pudesse executar qualquer processo era necessário escrever previamente um algoritmo com todos os padrões possíveis. Não há dúvidas de que uma máquina possa calcular muito mais rápido do que um ser humano. Contudo, mesmo que uma máquina possa calcular em alta velocidade, se estiver apenas procurando respostas aleatoriamente levará uma enorme quantidade de tempo. Portanto, muitos estudos foram realizados sobre como tornar essa busca mais eficiente. Os métodos mais populares desse tipo de algoritmo são a Busca em Profundidade (*Depth-First Search – DFS*) e a Busca em Largura (*Breadth-First Search – BFS*).

A busca DFS faz a varredura na profundidade dos nós⁴. Ou seja, ela começa na raiz de um grafo/árvore e explora o máximo possível de nós ao longo de cada ramificação antes do retrocesso. Por exemplo, na Figura 1 (DFS) a busca começa na raiz (1), percorre os nós (2), (3) e (4) e somente após chegar ao último nó (4) é que retrocede para o nó (3) e varre suas ramificações até encontrar o último nó (5). Esse processo se repete para os demais nós até que todos tenham sido percorridos.

⁴ De acordo com Brookshear (2001), na estrutura de dados denominada árvore, cada posição da árvore é chamada nó. O nó do topo é o nó-raiz e os nós das extremidades são denominados como nós-folhas. A linha que conecta os nós é definida como arco.

Já a busca BFS faz a varredura na largura dos nós. Ou seja, ela começa na raiz de um grafo/árvore e explora os nós vizinhos primeiro antes de passar para os nós que estão abaixo. Por exemplo, na Figura 1 (BFS) a busca começa na raiz (1), percorre os nós (2), (3) e (4) e, somente após não encontrar mais nenhum nó nesse nível, é que passa para os nós mais abaixo (5), (6), (7) e (8). Esse processo se repete para os demais nós até que todos tenham sido percorridos.

Figura 1 - Exemplo de Busca DFS e BFS



Fonte: Wikipédia

Os algoritmos de busca foram aplicados em campos específicos, como o xadrez e o jogo *Shogi* (versão japonesa do xadrez). Aliás, os jogos de tabuleiro são uma das áreas em que a máquina se destaca, pois, se for inserida uma entrada com grandes quantidades de padrões de vitória e derrota, dados de jogos anteriores e todos os movimentos permitidos de uma peça antecipadamente, a máquina pode avaliar a posição do tabuleiro e decidir o melhor movimento.

Segundo Sugomori (2016), uma máquina é excelente em processar dados em alta velocidade com base em um determinado conjunto de regras, mas não pode descobrir como agir e quais regras aplicar por si só. Os humanos avaliam, descartam muitas opções que não estão relacionadas e fazem uma escolha entre milhões de padrões de forma totalmente autônoma. Pesquisas realizadas nessa época do primeiro *boom* da IA só permitiram uma máquina seguir regras detalhadas e estabelecidas previamente por um humano. Embora os métodos de busca tivessem grande sucesso em uma área específica, como jogos de tabuleiro, estavam

longe de atingir a IA. Este foi o motivo pelo qual a primeira onda da IA perdeu força (SUGOMORI, 2016).

De acordo com Sugomori (2016), o segundo *boom* da IA ocorreu nos anos de 1980 com o movimento da Representação do Conhecimento (*Knowledge Representation* – KR). Para esse movimento:

If all the knowledge in the world was integrated into a machine and a machine could understand this knowledge, it should be able to provide the right answer even if it is given a complex task (SUGOMORI, 2016, p. 6)⁵.

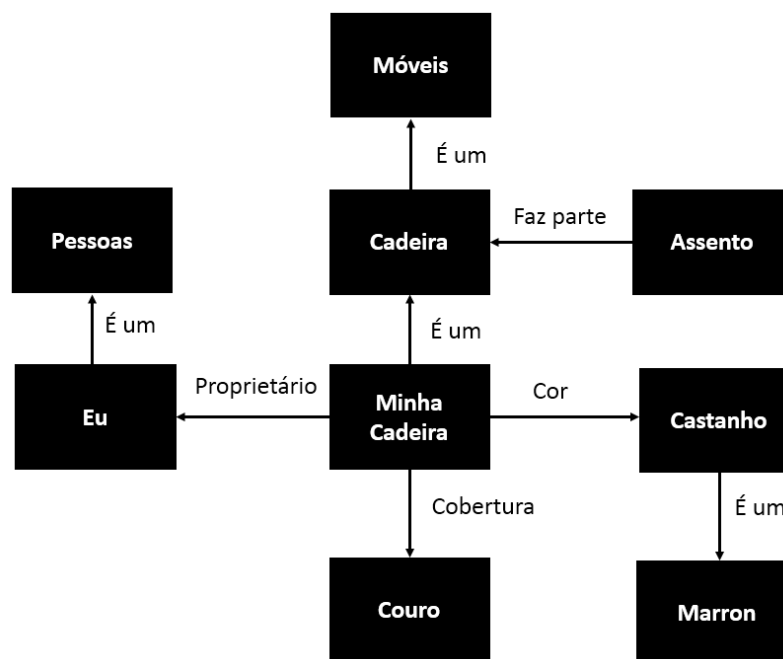
Com base nessa hipótese, vários métodos foram criados com o objetivo de introduzir conhecimento em uma máquina.

Muitas técnicas de Representação do Conhecimento têm sido estudadas e exploradas pelos pesquisadores de Inteligência Artificial. O objetivo dessas técnicas é produzir conceitos para a modelagem mais exata possível de alguns domínios do conhecimento para a criação de uma ontologia que descreva os conceitos do domínio, e isso será usado para armazenar e manipular o conhecimento para a geração de inferências, tomadas de decisões ou simplesmente para responder perguntas (ALVARENGA, 2005, p. 6).

Entre as várias técnicas de Representação do Conhecimento que foram estudadas, encontra-se a Rede Semântica. De acordo com Alvarenga (2005), uma Rede Semântica pode ser considerada como um grafo direcionado, formado por um conjunto de nós que representam os objetos e por um conjunto de arcos que representam as relações entre os objetos. Um arco é rotulado com o nome da relação que ele representa, e vários arcos podem ter o mesmo rótulo, porém, cada objeto é representado por apenas um nó.

Figura 2 - Exemplo de Rede Semântica

⁵ Tradução livre: Se todo o conhecimento do mundo fosse inserido em uma máquina e essa máquina fosse capaz de entender esse conhecimento, então ela seria capaz de fornecer a resposta certa mesmo em uma tarefa complexa.



Fonte: Alvarenga (2005)

Para Alvarenga (2005), ontologia é um termo que está intimamente ligado à Representação do Conhecimento e às Redes Semânticas. Na Ciência da Computação, ontologia é um modelo de dados que representa um conjunto de conceitos dentro de um domínio e os relacionamentos entre eles. Os estudos que envolvem ontologias procuram descrever as estruturas e os relacionamentos que são possíveis na realidade por meio de algum vocabulário comum (ALVARENGA, 2005).

Sugomori (2016) cita a *Web Semântica* como exemplo de Representação do Conhecimento. De acordo com Pickler (2007), Tim Berners-Lee, fundador da *World Wide Web*, foi o responsável pela proposta da *Web Semântica*, cuja finalidade consiste em incluir análise semântica na estrutura dos conteúdos disponíveis na Internet de forma a permitir que uma máquina possa analisar os conteúdos publicados. Segundo Sugomori (2016), a classificação em positivo e negativo, utilizada na análise de sentimentos, é um dos exemplos de como isso pode ser aplicado. Sugomori (2016) explica que a análise de sentimento é uma técnica que pode ser utilizada para avaliar postagens em redes sociais e *blogs*. A implementação dessa técnica, tomando como base as premissas da Representação do Conhecimento, se constitui da seguinte forma: define-se previamente um dicionário com o que é positivo ou negativo para cada palavra, de forma que uma máquina consiga comparar uma sentença com o dicionário, e com base nessa

confrontação possa definir se ela é positiva ou negativa (SUGOMORI, 2016). Sugomori (2016) chama a atenção para o fato de que nesse tipo de técnica mesmo o conhecimento sendo estruturado e sistematizado, a máquina continua a entendê-lo como um rótulo e esse conhecimento ainda é introduzido por um ser humano. Ou seja, a máquina apenas compara os dados e assume o significado de acordo com um dicionário previamente definido, ela ainda não consegue aprender de forma autônoma.

Outro exemplo de Representação do Conhecimento mencionado por Sugomori (2016), é o Cyc,⁶ que foi criado em 1984, por Douglas Lenat, pesquisador na área de Inteligência Artificial. Trata-se de um enorme banco de dados suportado por tecnologias semânticas que combina uma base de conhecimento comum com mecanismo de inferência. Domingos (2017) considera o Cyc como sendo o maior fracasso da história da Inteligência Artificial. Segundo ele, o objetivo era inserir em um banco de dados todo o conhecimento necessário para atingir a Inteligência Artificial e que essa meta seria alcançada dentro de uma década. Passaram-se mais de trinta anos e o Cyc continuou crescendo sem atingir o resultado esperado. Foi então que o Dr. Douglas Lenat decidiu minerar dados na *Web* e inseri-los no Cyc (DOMINGOS, 2017). Para Sugomori (2016), o fracasso inicial do Cyc ocorreu devido à inclusão de todo o conhecimento do mundo real em um banco de dados exigir uma “quantidade quase infinita de trabalho” (2016, p. 7). Para este autor, na época em que o Cyc foi concebido, não era possível minerar milhões de dados, analisá-los e inserir todo o conhecimento adquirido em uma máquina de forma automatizada. Isso se tornou viável somente agora, com a popularização da Internet, a enorme quantidade de dados disponíveis nesse meio e os vários algoritmos de *machine learning*.

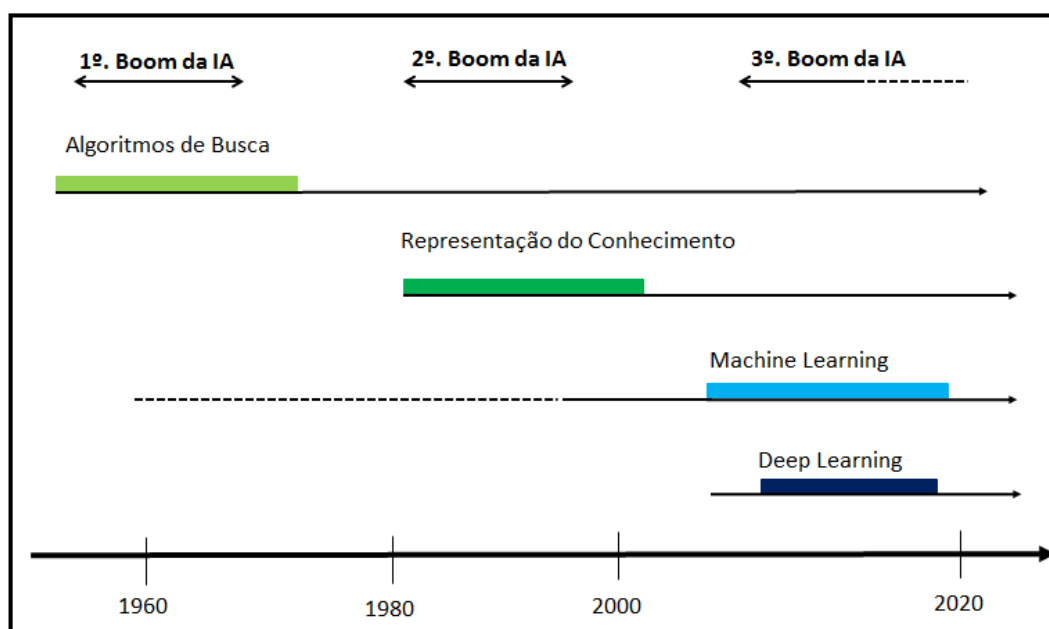
Segundo Puget (2016), o método denominado *machine learning* (aprendizado de máquina) foi definido inicialmente pelo cientista da Computação Arthur Samuel, em 1959. De acordo com Puget (2016), Arthur Samuel se tornou pioneiro nesse tipo de tecnologia ao desenvolver um algoritmo capaz de aprender a jogar dama de forma autônoma. Mas, foi somente por volta das décadas de 1990 e 2000 que esse método começou a ser aperfeiçoado e a possibilidade de atingir a IA se tornou mais

⁶ <http://www.cyc.com>

factível. O *machine learning* está intimamente relacionado com a mineração de dados e o *big data*.

Para Sugomori (2016), este método é uma ferramenta potente em comparação às abordagens anteriores que simplesmente tinham como base o conhecimento fornecido previamente por um ser humano. Para esse autor, o *deep learning*, que é um subcampo do *machine Learning*, se tornou uma evolução não só do *machine learning* como também do terceiro *boom* da IA. A Figura 3 mostra de forma sintetizada os *booms* da IA no decorrer do tempo.

Figura 3 - *Booms* da IA



Fonte: Sugomori (2016)

O *deep learning* é apenas o primeiro passo para uma máquina obter conhecimento de forma semelhante aos humanos. É difícil prever qual a próxima inovação. De acordo com a lei de Moore, o número de transistores dobra a cada 18 meses, ou seja, um ano e meio (SUGOMORI, 2016). Segundo Sugomori, se essa evolução continuar nesse ritmo, em 2045, o número de transistores ultrapassará dez bilhões, número esse que corresponde à quantidade de células no cérebro humano. Considerando que a lei de Moore manterá esse crescimento, em 2045, acredita-se que chegaremos a um ponto crítico chamado Singularidade Técnica (SUGOMORI, 2016).

O termo Singularidade Técnica foi criado pelo cientista Vernor Vinge, que discute sobre o assunto no artigo "The Coming Technological Singularity: How to

Survive in the Post-Human Era”, publicado em 1993. Resumidamente, Vinge descreve quatro itens tecnológicos que podem levar à Singularidade Técnica e, conseqüentemente, à superação dos homens pelas máquinas: 1) a velocidade com que os computadores são aperfeiçoados e a evolução da Inteligência Artificial; 2) as redes de computadores se tornarem autoconscientes; 3) as interfaces homem-máquina se tornarem tão complexas que produziriam um estágio evolutivo do homem, e 4) a ampliação da inteligência humana natural através de melhores técnicas da ciência biológica (VINGE, 1993).

Do lado contrário, o filósofo John Searle, em seu artigo “Minds, brains and programs” publicado em 1980, critica a Inteligência Artificial a qual ele dividiu em Inteligência Artificial Forte e Inteligência Artificial Fraca. De acordo com Searle (1980), Inteligência Artificial Forte é uma IA capaz de simular totalmente a inteligência humana. Já a Inteligência Artificial Fraca está associada a métodos específicos, que dão conta de funções especializadas da inteligência humana. Halper (2017) explica que alguns pesquisadores consideram que existe uma IA aplicada, enquanto outros, como Searle (1980), defendem uma IA Fraca que executa tarefas específicas versus uma IA Forte, em que a inteligência da máquina corresponde à inteligência humana.

De acordo com Amorim (2014), o pensamento do filósofo John Searle teve por objetivo mostrar que não é possível simular a mente através de processos meramente formais por meio da manipulação dos números 0 e 1 em um programa de computador. Amorim (2014) explica que para comprovar essa concepção, John Searle apresentou um argumento denominado Quarto Chinês⁷ como crítica à Inteligência Artificial Forte e ao Teste de Turing. O argumento do Quarto Chinês teve

⁷ O argumento do Quarto Chinês supõe uma pessoa trancada em um quarto onde há um lote de textos em chinês. Esta pessoa não fala e nem escreve nada na língua chinesa. Posteriormente ela recebe um segundo lote de textos em chinês e um conjunto de regras para correlacionar o segundo lote com o primeiro. As regras estão em um idioma conhecido de forma que a pessoa possa compreendê-las e identificar os símbolos por meio de suas formas, ou seja, sem nenhum conhecimento do sentido e significado desses símbolos. Esta pessoa recebe um terceiro lote de símbolos chineses junto com algumas instruções no idioma conhecido que permitem correlacionar elementos desse terceiro lote com os dois primeiros. Estas regras instruem a pessoa como devolver certos símbolos chineses com base em determinadas perguntas. O primeiro lote é denominado código, o segundo história e o terceiro as perguntas. Além disso, os símbolos que a pessoa retorna são as respostas às perguntas e o conjunto de regras John Searle definiu como programa (SEARLE, 1980).

por finalidade elucidar a impossibilidade de duplicação da inteligência humana por meio de algoritmos e para isso Searle fez uma analogia entre uma pessoa dentro de um quarto chinês, onde o quarto é o computador e a pessoa o programa. Nesse quarto (computador) a pessoa (programa) não compreende o significado dos símbolos em chinês e mesmo assim consegue responder às perguntas em chinês com base em regras definidas (AMORIM, 2014). Para ajudar a elucidar esse tema, o próximo item explica o conceito de IA e apresenta exemplos de seus subcampos.

2.1 Inteligência Artificial e seus Subcampos

Carolina Bigonha define Inteligência Artificial da seguinte forma:

É um campo de estudo que surgiu na década de 50, cujo objetivo principal é o estudo e a construção de sistemas capazes de exibir comportamentos normalmente associados às pessoas, como aprendizado e resolução de problemas. Algumas linhas de estudo de Inteligência Artificial são mais focadas em reproduzir a maneira como pensamos e raciocinamos, ao passo que outras se concentram no entendimento e na simulação de comportamento (BIGONHA, 2018, p. 2).

Halper (2017) explica que embora os termos *machine learning* e Inteligência Artificial tenham sido criados na década de 1950, ainda há muita confusão, principalmente no mercado, em relação ao significado dos mesmos. Alguns utilizam esses termos de forma relacionada, mas essas tecnologias são diferentes. Para ela, essa confusão se agrava quando alguns fornecedores de *software* passam a rotular tudo como IA. Para esclarecer esse equívoco, Halper (2017) define que aprendizado de máquina (*machine learning*), processamento de linguagem natural (*natural language process*), robótica e reconhecimento de imagens são subcampos da IA.

Diante de tantas definições e jargões técnicos é importante esclarecer o que realmente é Inteligência Artificial. Há muitos anúncios de produtos, como os de sistemas de recomendação, assistentes virtuais dizendo que utilizam Inteligência Artificial. Para Sugomori (2016), esse tipo de anúncio não é correto, pois a palavra Inteligência Artificial é usada como um conceito mais amplo. No entendimento dele, as pesquisas e técnicas acumuladas no passado alcançaram apenas algumas partes da Inteligência Artificial e, erroneamente, as pessoas utilizam este termo para designar essas partes. Não é de se estranhar que muitas vezes *machine learning* e *deep learning* sejam definidos como Inteligência Artificial.

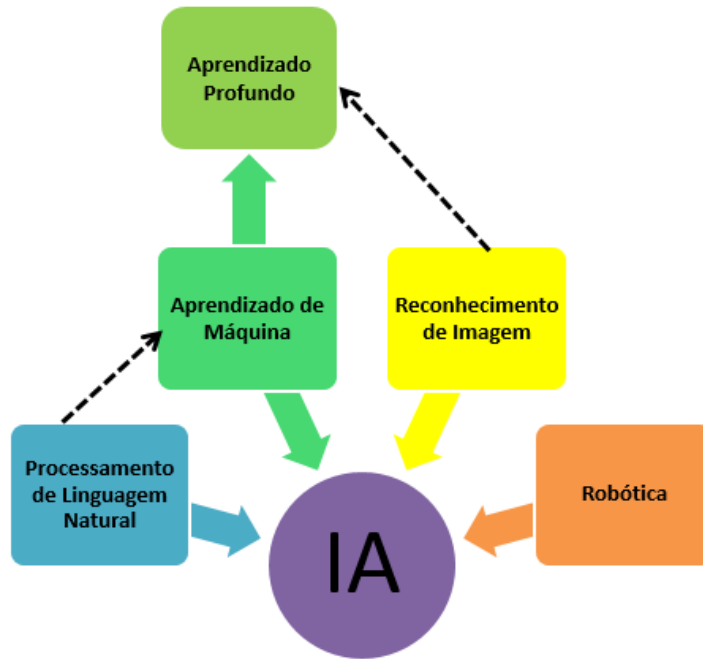
Para Biddle (2018), o termo Inteligência Artificial foi banalizado como tantas outras *buzzwords* do mundo digital. Ele considera que IA é um conceito amplo e possui subcampos como o aprendizado de máquina que permite aos computadores aprenderem sozinhos e se tornarem mais eficientes numa enorme variedade de aplicações, como a classificação de *spam*, a detecção de fraudes financeiras, entre outras.

De acordo com Domingos (2017), o aprendizado de máquina (*machine learning*) é conhecido por vários termos: reconhecimento de padrões, modelagem estatística, mineração de dados, descoberta de conhecimento, análise preditiva, ciência de dados, sistemas adaptativos, entre outros. Para ele, Inteligência Artificial deve ser considerada como um grande guarda-chuva para diversos métodos.

Tal amplitude de terminologias gera confusão. Além disso, outra questão que pode gerar dúvidas é que alguns subcampos da IA utilizam algoritmos e técnicas de outros subcampos. Por exemplo, algoritmos de *machine learning* podem ser utilizados no processamento de linguagem natural. Assim como, o reconhecimento de imagem pode utilizar *deep learning* que é um subcampo do *machine learning*.

A Figura 4 ilustra a Inteligência Artificial com exemplos de seus subcampos: aprendizado de máquina (*machine learning*), reconhecimento de imagem, robótica e processamento de linguagem natural. O aprendizado profundo (*deep learning*) deve ser considerado como um subcampo do *machine learning*. As setas tracejadas indicam os subcampos da IA que podem utilizar algoritmos de outros subcampos.

Figura 4 - Exemplos de subcampos da IA



Fonte: Autora

A Figura 5 ilustra a IA, seus subcampos e aplicações. Uma das aplicações mais conhecidas da IA foi, sem dúvida, uma máquina jogar xadrez. A classificação de *e-mails* como *spam*, é um dos exemplos mais clássicos de *machine learning*, subcampo da IA. A identificação da imagem de um gato, feita por um algoritmo do *Google*, é um dos exemplos mais célebres de *deep learning*, que é um subcampo do *machine learning*.

Figura 5 - Inteligência Artificial, Machine Learning e Deep Learning



Fonte: *Medium*⁸

⁸ <https://medium.com/data-science-brigade/a-diferen%C3%A7a-entre-intelig%C3%A2ncia-artificial-machine-learning-e-deep-learning-930b5cc2aa42>

De acordo com Halper (2017), à medida que os subcampos da IA aumentam e se popularizam, os questionamentos referentes à privacidade e como essa tecnologia vai influenciar a economia e a vida das pessoas começam a emergir. Para Halper (2017), não há dúvida de que a automação traz melhorias significativas para processos operacionais, compreensão de comportamentos, diagnóstico e previsão. No entanto, isso precisa ser analisado para levantar as consequências que toda essa automação poderá acarretar.

Halper (2017) investigou os riscos que a IA poderá acarretar e o resultado indica que a maior preocupação, no momento, é a privacidade. Esse temor engloba cenários como o acesso de entidades externas às informações privadas e o uso dessas informações de forma “inescrupulosa”, como uma “companhia de seguros determinar um seguro com base nos dados capturados sem consentimento” (2017, p. 23). Questões éticas envolvendo subcampos da IA assumem outra posição de destaque e estão relacionadas com o futuro do trabalho e o poder que esse tipo de tecnologia terá para fazer avaliações e tomar decisões de forma autônoma como, por exemplo, usar reconhecimento de imagem para diagnóstico de doenças e a complexidade dos carros autônomos que “podem acarretar perda de vidas” (2017, p. 23). O quanto antes essas preocupações forem abordadas e compreendidas, melhor. Para promover uma IA mais segura e responsável foram formadas várias organizações, como o *OpenAI*⁹, o *Machine Intelligence Research Institute (MIRI)*¹⁰, *Future of Humanity Institute*¹¹ e *AI Fund*¹². A fim de ajudar a entender melhor todo esse contexto, o próximo item detalha o conceito de aprendizado de máquina.

2.2 Aprendizado de Máquina (*Machine Learning*)

Antes do aprendizado de máquina (*machine learning*), uma máquina só fornecia resposta com base nas regras que já haviam sido inseridas. Nesse cenário, a máquina fornecia rapidamente respostas para uma pergunta previamente conhecida, mas literalmente travava quando surgiam questões desconhecidas.

Máquinas algorítmicas são projetadas para executarem tarefas precisamente definidas, com velocidade e exatidão, e fazem isso

⁹ <https://openai.com/>

¹⁰ <https://intelligence.org/>

¹¹ <https://www.fhi.ox.ac.uk/>

¹² <https://aifund.ai/>

extremamente bem. Todavia, máquinas são desprovidas de bom senso. Quando diante de uma situação não prevista pelo programador, é provável que o seu desempenho caia sensivelmente (BROOKSHEAR, 2001, p. 373).

Halper (2017) define aprendizado de máquina como um método de análise de dados que teve origem na Ciência Computação. Nesse método de aprendizado automático, os algoritmos aprendem com os dados e visam identificar padrões com a mínima intervenção humana. A máquina aprende por meio de exemplos, basicamente usando os tipos de aprendizagem supervisionada e não supervisionada. Um maior detalhamento sobre esses tipos de aprendizado é feito no item 2.3.

Ao contrário do que a maioria pensa, aprendizado de máquina não difere da análise preditiva. Essa confusão acontece devido às abordagens estatísticas como é o caso da regressão linear, ser frequentemente referenciada como análise preditiva. Mas, para desmistificar isso Halper (2017) diz que algoritmos de aprendizado de máquina que são utilizados para análise de fraude, também utilizam modelos estatísticos. Logo o aprendizado de máquina e a análise preditiva são sinônimos, como já observou Pedro Domingos (2017).

Outra característica do *machine learning* é que a máquina pode lidar com questões desconhecidas com base no conhecimento aprendido anteriormente. Este aprendizado ocorre quando é possível dividir um problema em sim ou não (0 ou 1). Neste caso, pode-se afirmar que o *machine learning* é um método de reconhecimento de padrões (DOMINGOS, 2017).

Sugomori (2016) explica que, no *machine learning*, a máquina utiliza uma quantidade gigantesca de dados de treinamento, substituindo perguntas complexas por perguntas do tipo sim ou não (0 ou 1) e descobre a regularidade com que os dados são marcados com sim e com não. Em suma, esse sistema pode dar uma resposta reconhecendo e classificando os padrões a partir dos dados fornecidos e, em seguida, classificando esses dados no padrão apropriado (previsão). Dessa forma, quando a máquina se depara com dados desconhecidos numa pergunta é capaz de fazer uma previsão e fornecer uma resposta.

O *machine learning* é um método em que a máquina pode processar esse reconhecimento de padrões de forma autônoma sem nenhuma interferência humana. Segundo Sugomori (2016), esse padrão de classificação é baseado em

uma fórmula numérica chamada de modelo estatístico probabilístico. Esta abordagem tem sido estudada com base em vários modelos matemáticos. No processo de treinamento, os parâmetros de um modelo são ajustados conforme o processo de aprendizagem vai sendo executado. Dessa forma, após a conclusão do aprendizado a máquina categoriza dados desconhecidos no padrão que se ajusta melhor aos parâmetros aprendidos. Com o *machine learning*, as máquinas tornaram-se capazes de processar dados e fornecer respostas que os humanos não são capazes de fazer (SUGOMORI, 2016).

De acordo com Sugomori (2016), o conceito de *machine learning* em si, vem de longa data, contudo pesquisadores não tiveram condições de provar a utilidade desse método anteriormente, devido à tecnologia da época não suportar o processamento de um alto volume de dados e também pela própria falta de dados.

Halper (2017) reforça esse argumento quando diz que Inteligência Artificial, aprendizado de máquina e processamento de linguagem natural são tecnologias antigas, e que só conseguiram aplicabilidade e popularidade atualmente por conta dos avanços computacionais e do desenvolvimento de novos algoritmos. A consequência disso é que essas tecnologias se tornaram mais acessíveis para as empresas que passaram a adotá-las para melhoria da eficiência operacional, entendimento de comportamentos e obtenção de vantagem competitiva.

Atualmente, com a enorme quantidade de dados disponíveis no universo *online*, códigos fontes abertos com várias opções de algoritmos e a alta capacidade de processamento dos computadores, os pesquisadores conseguiram provar a eficiência desse método e o *machine learning* foi cada vez mais aperfeiçoado e com isso surgiu o terceiro *boom* da IA.

New advancements in compute power along with some new algorithmic developments are making machine learning, NLP, and AI attractive to many companies (HALPER, 2017, p. 5).¹³

Ainda com toda essa otimização, o *machine learning* apresenta um ponto fraco que se encontra justamente no aprendizado supervisionado. É importante destacar que esse método estabelece uma técnica de classificação e predição de padrões baseada nos dados de entrada. Para a máquina aprender corretamente ela precisa

¹³ Tradução livre: Novos avanços no poder de computação, juntamente com alguns novos desenvolvimentos algorítmicos, estão tornando o aprendizado de máquina, o processamento de linguagem natural e a IA atrativos para muitas empresas.

ter dados apropriados. Porém, o método de aprendizado supervisionado não é capaz de decidir quais são os dados apropriados e quais não são. É necessária a intervenção humana no tratamento e seleção dos dados corretos, e somente de posse dos dados tratados é que o método de aprendizado supervisionado poderá encontrar um padrão. Sugomori (2016) explica que essa deficiência ocorre devido à falta de engenharia de recursos (*feature engineering*). De acordo com a *Microsoft*, engenharia de recursos é o processo que tenta criar outros recursos relevantes com base nos recursos brutos existentes nos dados para aumentar a capacidade de previsão do algoritmo de aprendizado (*MICROSOFT, 2017, online*¹⁴).

Em cada uma das fases da IA, Sugomori (2016) identificou problemas que distinguem o aprendizado de uma máquina do aprendizado humano. Na primeira fase ele cita o problema do frame, ou seja, quando uma máquina se depara com uma situação não prevista nas regras definidas para execução de uma determinada tarefa e trava. Na segunda fase, destaca o problema da fundamentação simbólica, mesmo que o conhecimento seja estruturado e sistematizado, a máquina entenderá esse conhecimento apenas como um rótulo e não como um conceito, pois o conhecimento é inserido por um humano em uma base de dados e o que a máquina faz é simplesmente comparar os dados e assumir significados com base em um dicionário. Na terceira fase ocorre o problema da engenharia de recursos no método de aprendizado supervisionado, quando uma máquina não possui a capacidade de separar o joio do trigo e selecionar os dados corretos de forma autônoma.

Esses problemas só podem ser resolvidos se uma máquina tiver condições de focar e definir quais informações ela deve utilizar. Esta é uma das grandes diferenças entre uma máquina e um humano. Os seres humanos conhecem características, possuem instintos e experiências e com base nisso compreendem as coisas por meio do conceito.

Mas, o que é conceito? De acordo com Sugomori (2016), cada coisa é constituída de um conjunto de representações simbólicas e do conteúdo dos símbolos. A palavra gato é a representação do símbolo e o conceito que você reconhece como um gato é o conteúdo desse símbolo. Quando as pessoas ouvem a palavra gato imaginam algo muito parecido. O conceito, portanto, é o conteúdo

¹⁴ <https://docs.microsoft.com/pt-br/azure/machine-learning/team-data-science-process/create-features>

simbólico. A representação simbólica é chamada de significante e o conteúdo simbólico (conceito) de significado. É possível dizer que o significante é a parte perceptível e o significado a parte inteligível. Em síntese, o significado pode ser entendido como o sentido, o conceito, a ideia de alguma coisa, ou seja, a representação mental de algo. Já o significante pode ser entendido como a imagem acústica (XAVIER, 2014, p. 89).

O signo linguístico não é uma coisa e uma palavra, mas um conceito e uma imagem acústica. Esta não é o som material, coisa puramente física, mas a impressão (empreinte) psíquica desse som, a representação que dele nos dá o testemunho de nossos sentidos (SAUSSURE, 2006, p. 80).

Machine learning, sem dúvida, é um poderoso método quando se trata da mineração de dados e foi o pontapé inicial para o terceiro *boom* da IA. Contudo, não foi suficiente para alcançar definitivamente uma IA forte. Mas, isso não significa que essa odisséia chegou ao fim. O terceiro boom está apenas começando e uma nova onda está a caminho com o *deep learning*. Com o advento desse novo método de aprendizado, pelo menos nas áreas de reconhecimento de imagem e voz, uma máquina tornou-se capaz de decidir a partir dos dados disponibilizados e escolher os melhores dados sem necessidade de manipulação humana. Antes uma máquina apenas poderia manipular símbolos e agora tornou-se capaz de obter conceitos.

A grande diferença entre as técnicas anteriores e o *deep learning*, segundo Sugomori (2016), é que o *deep learning* é uma adaptação do algoritmo de rede neurais do *machine learning*. O algoritmo de redes neurais visa imitar a estrutura do cérebro humano. Para este autor, *deep learning* é muito mais poderoso e significativo do que qualquer outro algoritmo de *machine learning*, incluindo as próprias redes neurais. O próximo item mostra mais detalhes sobre o *deep learning*.

2.2.1 Aprendizado Profundo (*Deep Learning*)

De acordo com Sugomori (2016), o aprendizado profundo (*deep learning*) surgiu em 2006 quando o professor Geoffrey Hinton, da Universidade de Toronto, e outros colegas, publicaram um artigo chamado “Deep Belief Nets (DBN)”. O DBN, uma expansão das redes neurais, foi testado usando o MNIST que é o banco de dados padrão para comparar a precisão de cada método de reconhecimento de imagem. Este banco de dados inclui 70.000 imagens de caracteres escritos à mão dos números de 0 a 9 (60.000 são para treinamento e 10.000 são para testes). Em

seguida, eles construíram um modelo de previsão com base nos dados de treinamento e mediram sua precisão baseada no fato de uma máquina poder responder corretamente qual número de 0 a 9 foi escrito.

Em 2012, no *Imagenet Large Scale Visual Recognition Challenge* (ILSVRC), competição mundial por reconhecimento de imagem, um método que usava *deep learning* chamado *SuperVision*, desenvolvido pela equipe do professor Geoffrey Hinton, da Universidade de Toronto, venceu a competição. Ele superou em muito os outros concorrentes com uma excelente precisão. Nessa competição o objetivo era fazer a máquina distinguir automaticamente se uma imagem era um gato, um cachorro, um pássaro, um carro, um barco e assim por diante. Para a realização do experimento foram fornecidas 10 milhões de imagens como dados de treinamento e 150.000 imagens para o teste. No teste, cada método compete para retornar a menor taxa de erro e, neste caso, quanto menor a taxa de erro maior é a precisão do método (SUGOMORI, 2016).

Tabela 1 - Resultados do teste do ILSVRC

Posição	Nome do Método	Taxa de Erro
1	<i>SuperVision</i>	0.15315
2	<i>SuperVision</i>	0.16422
3	ISI	0.26172
4	ISI	0.26602
5	ISI	0.26646
6	ISI	0.26952
7	OXFORD_VGG	0.26979

Fonte: Imagenet¹⁵

Antes do *SuperVision* (*deep learning*) a abordagem utilizada no campo do reconhecimento de imagens era o *machine learning*. Relembrando, o valor da engenharia de recurso no *machine learning* tinha que ser definido ou projetado por humanos. A grande evolução do *deep learning* foi justamente conseguir que a máquina fizesse isso de forma automática diminuindo a taxa de erro e, conseqüentemente, aumentando a precisão.

¹⁵ <http://image-net.org/challenges/LSVRC/2012/results.html#t1>

Outro grande evento que ajudou na divulgação do *deep learning* foi o *International Conference on Machine Learning* (ICML) que também ocorreu em 2012. Nesse evento, de acordo com Sugomori (2016), o *Google* anunciou que uma máquina poderia detectar automaticamente um gato usando vídeos do *YouTube*. O objetivo do *Google* foi utilizar o *deep learning* para identificar imagens aleatórias e não rotuladas de gatos, em vídeos do *YouTube*.

Segundo Dean e Ng (2012), o algoritmo utilizado foi o de redes neurais artificiais. Esses autores explicaram que para conseguir uma capacidade de processamento suficiente para simular uma rede neural profunda, a equipe do *Google* desenvolveu uma infraestrutura de computação distribuída para o treinamento de redes neurais em grande escala. Distribuíram essa rede entre 16.000 núcleos de CPU de seus *datacenters* e treinaram os modelos com mais de 1 bilhão de conexões. A meta era simular essa rede neural como se fosse o cérebro de um recém-nascido. Para isso, exibiram vídeos do *YouTube* durante uma semana e ao final desse período verificaram o que a rede neural tinha conseguido aprender. A hipótese do *Google* era de que a rede aprenderia a reconhecer objetos comuns nos vídeos (DEAN e NG, 2012).

O resultado foi que um dos neurônios conseguiu aprender a reconhecer imagens de gatos. O mais surpreendente é que nesta rede nunca foi informado anteriormente o que era um gato, pois nenhuma imagem rotulada com gato foi exibida. Esse neurônio descobriu sozinho como era um gato de forma completamente autônoma por meio das imagens do *YouTube* (DEAN e NG, 2012).

Figura 6 - Imagem de um gato identificada pelo algoritmo do *Google*

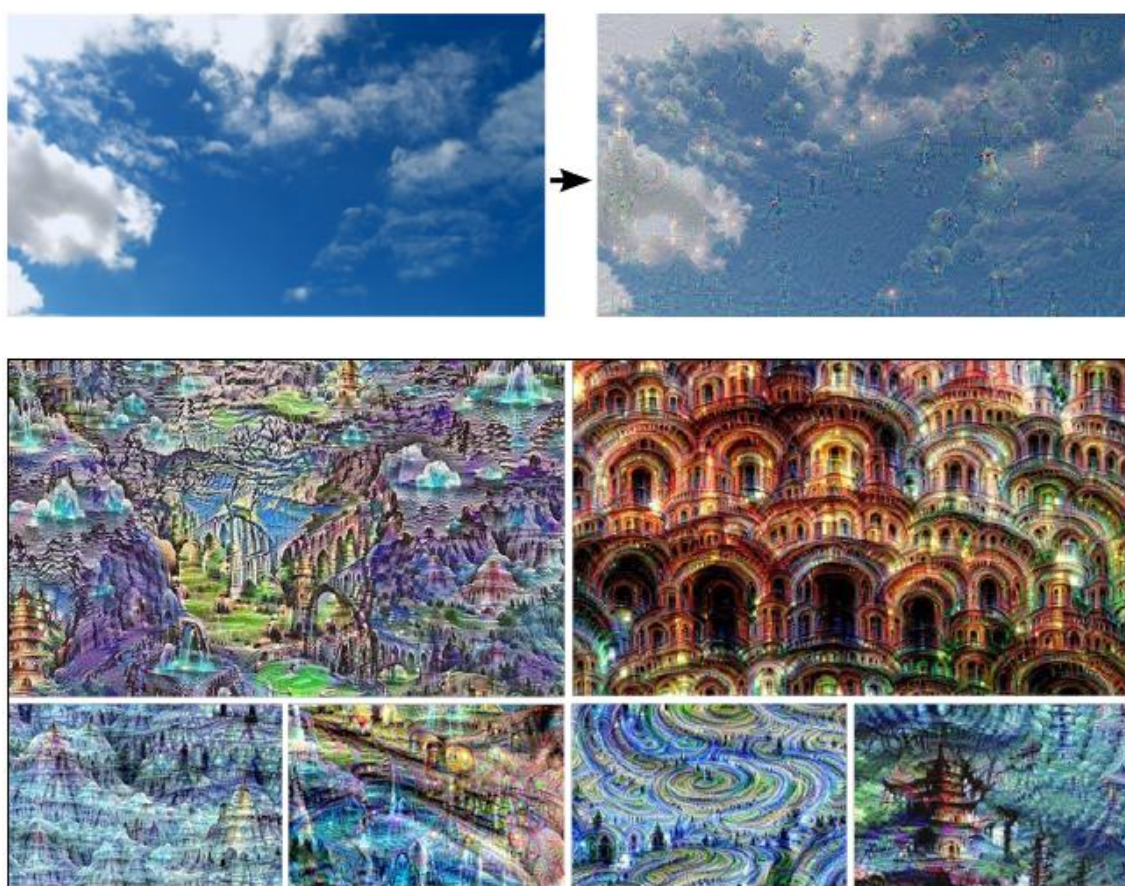


Fonte: Dean e Ng (2012)

Sugomori (2016) também cita que após o desenvolvimento do método de reconhecimento automático de um gato, o *Google* realizou outra experiência para uma máquina desenhar imagens utilizando o aprendizado profundo. De acordo com ele esse método foi denominado *Inceptionism*:

Whatever you see there, I want more of it! This creates a feedback loop: if a cloud looks a little bit like a bird, the network will make it look more like a bird. This in turn will make the network recognize the bird even more strongly on the next pass and so forth, until a highly detailed bird appears, seemingly out of nowhere (MORDVINTSEV, OLAH, e TYKA, 2015, *online*)¹⁶.

Figura 7 - Imagens geradas pelo *Inceptionism*



FONTE: Mordvintsev, Olah, e Tyka (2015)

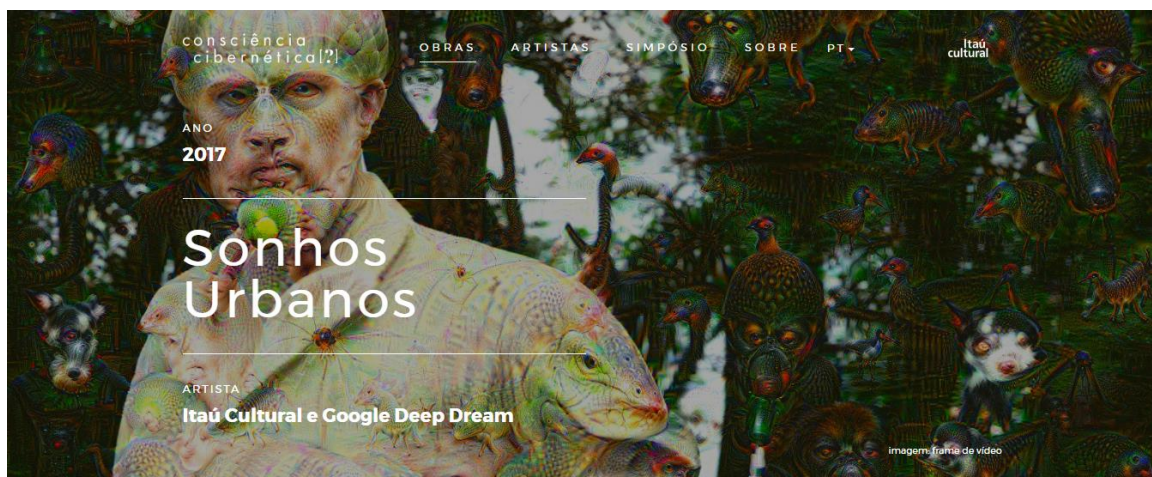
O método das redes neurais na aprendizagem de máquina é geralmente usado para detectar padrões e foi o que o algoritmo do *Google* fez no reconhecimento da

¹⁶ Tradução livre: O que quer que você veja lá, eu quero mais! Isso cria um loop de *feedback*: se uma nuvem parece um pouco com um pássaro, a rede fará com que pareça mais um pássaro. Isso, por sua vez, fará com que a rede reconheça a ave ainda mais fortemente na próxima passagem e assim por diante, até que um pássaro altamente detalhado apareça, aparentemente do nada.

imagem do gato. Já o *Inceptionism* faz justamente o contrário, ou seja, cria imagens estranhas que não possuem nenhum padrão. As imagens criadas são totalmente criativas, melhor dizendo, artísticas. A ferramenta *Deep Dream* que permite experimentar o *Inceptionism* é de código fonte aberto e está disponível no *GitHub*¹⁷ com implementação de exemplos. Para utilizar esta ferramenta é necessário conhecer a linguagem de programação *Python* (SUGOMORI, 2016).

Em 2017, a exposição do Itaú Cultural, intitulada “Consciência Cibernética [?]”, contou com dez trabalhos de artistas brasileiros e estrangeiros, que utilizaram a tecnologia para propor um olhar artístico sobre a evolução das máquinas e questionar se elas terão capacidade de ultrapassar funções executadas apenas pelo cérebro humano. Os artistas utilizaram tecnologias como realidade virtual, *big data*, redes neurais, processamento de linguagem natural e a ferramenta *Deep Dream*, do *Google*, com cenas urbanas da Avenida Paulista. Essas cenas foram definidas como “Sonhos Urbanos”.

Figura 8 - Imagem dos Sonhos Urbanos, da exposição “Consciência Cibernética [?]”



FONTE: Itaú Cultural¹⁸

Sugomori (2016) destaca que o cérebro humano é composto por inúmeros circuitos amontoados em várias camadas. Quando o algoritmo de rede neural surgiu, sua estrutura com apenas poucas camadas foi uma réplica bem simples do cérebro

¹⁷ <https://github.com/google/deepdream>

¹⁸ <http://www.itaucultural.org.br/conscienciaticibernetica/obra/sonhos-urbanos/>

humano. Por este motivo os padrões que ele poderia reconhecer eram bem limitados.

Por outro lado, Sugomori (2016) explica que com o *deep learning* foi possível transformar uma rede em várias camadas tornando essa rede profunda. A chave do sucesso foi fazer com que cada camada aprendesse por etapas. Com isso, a máquina torna-se capaz de aprender uma característica na primeira camada, outra característica na segunda e dessa forma o aprendizado vai acontecendo gradualmente. Ao aprender o que é um gato, a primeira camada verifica um contorno, a segunda a forma dos olhos e nariz, a terceira a imagem da face e assim por diante. Como cada camada aprende em etapas, o *feedback* de um erro de aprendizado também pode ser feito corretamente em cada camada melhorando a precisão (SUGOMORI, 2016).

Os algoritmos de *deep learning* permitiram que uma máquina pudesse classificar padrões complexos como imagens. Atualmente, o grande interesse do mercado pelo aprendizado de máquina (*machine learning*) se justifica principalmente por conta do *deep learning*, que é um subcampo dessa tecnologia. Segundo Halper (2017), enquanto as antigas redes neurais não conseguiam reconhecer padrões complexos, os avanços algorítmicos da última década somados ao poder computacional, tornaram o *deep learning* viável e bem mais preciso do que se imaginava. Isso estimulou o uso do *deep learning* para a classificação de áudios, imagens e textos. Para Halper (2017), grande parte do interesse atual em torno da IA vem do aprendizado profundo (*deep learning*). Sugomori (2016) reforça tal afirmação quando diz que o terceiro *boom* da IA evoluiu com o *deep learning*. A Tabela 2 apresenta, de forma resumida, os principais marcos na evolução das redes neurais artificiais até a chegada do *deep learning*.

Tabela 2 - Evolução das Redes Neurais Artificiais até o Deep Learning

Ano	Marco
1943	Warren McCulloch e Walter Pitts escreveram um artigo com o objetivo de estabelecer uma analogia entre as células nervosas e o processo eletrônico. Este artigo pode ser considerado como o primeiro modelo de redes neurais artificiais.
1958	Frank Rosenblatt criou o <i>Perceptron</i> , um algoritmo para o reconhecimento de padrões baseado em uma rede neural computacional de duas camadas. Ele também propôs camadas adicionais com notações matemáticas, mas isso foi realizado somente em 1975. Segundo

Ano	Marco
	Domingos (2017), o nome <i>Perceptron</i> deriva do interesse pela aplicação desse modelo em tarefas perceptivas, como o reconhecimento de voz e de caracteres.
1980	Kunihiko Fukushima propôs a <i>Neocognitron</i> , uma rede neural de hierarquia, multicamada, que foi utilizada para o reconhecimento de caligrafia e outros problemas de reconhecimento de padrões.
1989	Cientistas criaram algoritmos que usavam redes neurais profundas, mas o tempo de treinamento para os sistemas levaram dias, tornando-os impraticáveis para o uso no mundo real.
1992	Juyang Weng publicou o <i>Cresceptron</i> , um método capaz de realizar o reconhecimento de objetos 3-D automaticamente a partir de cenas desordenadas.
2006	O termo aprendizado profundo (<i>deep learning</i>) começou a ganhar popularidade após artigo de Geoffrey Hinton e seus colegas demonstrar como uma rede neural de várias camadas poderia ser pré-treinada uma camada por vez.
2009	Aconteceu o NIPS Workshop sobre aprendizagem profunda para reconhecimento de voz e descobriu-se que com um conjunto de dados suficientemente grande, as redes neurais não precisavam de pré-treinamento e as taxas de erro caíram significativamente.
2012	O algoritmo <i>SuperVision</i> , desenvolvido pelo professor Geoffrey Hinton e demais colegas, obteve um excelente desempenho na identificação de imagens. Nesse mesmo ano, também surgiu o algoritmo de aprendizado profundo (<i>deep learning</i>) do <i>Google</i> que foi capaz de identificar gatos por meio de imagens do <i>YouTube</i> .
2014	O <i>Google</i> adquiriu a <i>startup</i> de Inteligência Artificial, <i>Deep Mind</i> .
2015	O <i>Facebook</i> colocou em operação a tecnologia de aprendizado profundo, chamada <i>Deep Face</i> , para marcar e identificar automaticamente usuários do <i>Facebook</i> em fotografias. Algoritmos executaram tarefas de reconhecimento facial usando redes profundas que levaram em conta 120 milhões de parâmetros.
2016	O algoritmo do <i>Google Deep Mind</i> , <i>AlphaGo</i> , mapeou a arte do complexo jogo de tabuleiro Go e venceu o campeão mundial de Go, Lee Sedol, em um torneio altamente divulgado em Seul.
2017	Adoção em massa do <i>deep learning</i> em diversas aplicações corporativas e <i>mobile</i> , além do avanço em pesquisas. Todos os eventos de tecnologia ligados à <i>data science</i> , IA e <i>big data</i> , apontam o <i>deep learning</i> como a principal tecnologia para criação de sistemas inteligentes.

Fonte: *Deep Learning Book (online)*¹⁹

¹⁹ <http://deeplearningbook.com.br/uma-breve-historia-das-redes-neurais-artificiais/>

Halper (2017) explica que algoritmos de aprendizagem profunda, como as redes neurais artificiais, possuem nós de entrada e um número de camadas ocultas que funcionam como uma “caixa-preta” na modelagem dos nós de saída (2017, p. 6). Por conta disso, as etapas demoram muito tempo para serem concluídas e exigem uma enorme quantidade de cálculo e processamento computacional. Não é à toa que esse tipo de aprendizado é definido como profundo. Por isso, mesmo com todo o sucesso que o *deep learning* vem obtendo, é importante ressaltar que também possui pontos fracos, como a enorme quantidade de cálculo, necessidade de mais poder computacional e elevado tempo para conclusão.

2.2.2 Processamento de Linguagem Natural (*Natural Language Process - NLP*)

Embora o foco desta pesquisa seja aprendizado de máquina (*machine learning*), é fato que o subcampo da IA, denominado processamento de linguagem natural (*natural language process - NLP*), utiliza algoritmos do *machine learning* e, por este motivo, acaba sendo amplamente utilizado, conforme detalhamento das aplicações dessas tecnologias descrito no item 2.5. Diante disso, é necessário apresentar algumas definições importantes sobre este subcampo, com o intuito de facilitar a compreensão dos demais capítulos.

Segundo Halper (2017), o processamento de linguagem natural ou *natural language processing* (NLP) é uma técnica que visa analisar, compreender e gerar respostas e/ou resultados com base na linguagem humana.

De acordo com Liu (2010), as informações textuais podem ser categorizadas basicamente em dois tipos: fatos e opiniões. Fatos são expressões objetivas sobre entidades, eventos e suas propriedades. As opiniões, geralmente, são expressões subjetivas que descrevem avaliações ou sentimentos das pessoas em relação a entidades, eventos e suas propriedades.

Segundo Liu (2010), a *Web* mudou a forma como as pessoas expressam seus pontos de vista e opiniões. Atualmente, os internautas podem expressar suas opiniões sobre quase tudo na *Web* e nas redes sociais. Este comportamento “*online de boca a boca*” (2019, p.1) representa novas fontes de mensuração de informação com muitas aplicações práticas. Para uma empresa, pode não ser mais necessário realizar pesquisas, organizar grupos focais ou contratar consultores externos para encontrar opiniões de consumidores sobre seus produtos e sobre seus

concorrentes, pois o conteúdo gerado por usuários na *Web* já disponibiliza tais informações.

Entretanto, Liu (2010) alerta que encontrar fontes de opinião e monitorá-las na *Web* é tarefa difícil, porque há uma diversidade de fontes; e cada fonte pode também ter um volume enorme de texto, tanto de fatos quanto de sentimentos. Em muitos casos, as opiniões estão escondidas em longas postagens. Nesse cenário é difícil para um ser humano encontrar fontes relevantes, extrair frases relacionadas com opiniões, resumí-las e organizá-las. Assim, são necessários sistemas automatizados de descoberta e compactação de opinião. A análise de sentimento, também conhecida como mineração de opinião, cresceu a partir dessa necessidade e se transformou em um problema desafiador de processamento de linguagem natural. Devido ao seu grande valor para aplicações práticas, houve um crescimento explosivo desse segmento.

Liu (2010) explica que a classificação de sentimentos pode ser formulada tanto com algoritmos de aprendizado de máquina supervisionado quanto com aprendizado de máquina não supervisionado. O próximo item aborda com detalhes esses tipos de aprendizado de máquina.

2.3 Tipos de Aprendizado de Máquina e seus Algoritmos

Halper (2017) cita dois tipos de aprendizado de máquina, o supervisionado e o não supervisionado. No aprendizado supervisionado o algoritmo é treinado com recursos (dados) previamente rotulados; ao contrário do aprendizado não supervisionado em que os dados não são previamente rotulados. Já Theobald (2017) cita três tipos de aprendizado de máquina: aprendizado supervisionado, aprendizado não supervisionado e aprendizado por reforço.

De acordo com ele, o aprendizado supervisionado é guiado por padrões pré-existentes e resultados conhecidos. Ou seja, a máquina utiliza dados rotulados tanto na entrada quanto na saída, e com a combinação desses dados define um padrão de classificação. Uma vez identificado o padrão existente nos dados, um modelo é desenvolvido de forma a reproduzir os resultados com novos dados. Por exemplo, a classificação de mensagens em *spam* e não *spam*. Em um ambiente de aprendizado supervisionado, as mensagens recebidas podem ser classificadas em *spam* e não *spam* de forma que a máquina entenda as características das mensagens e as

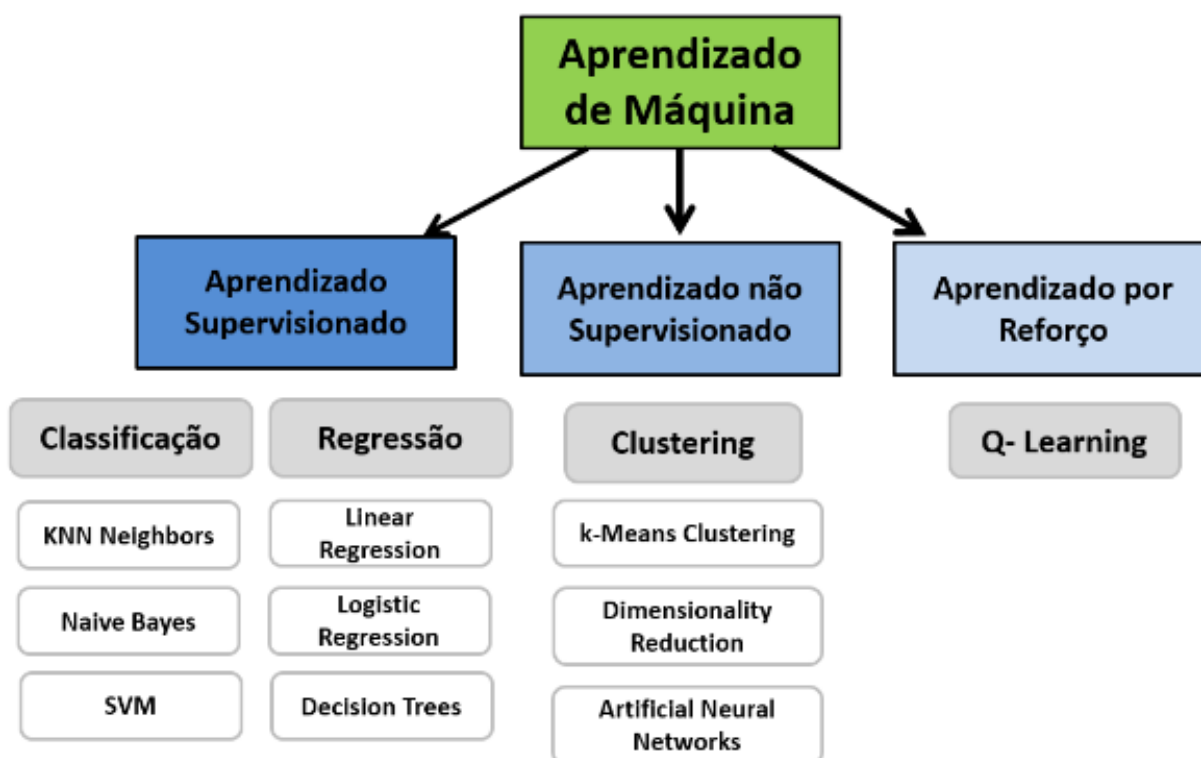
classifique nessas duas categorias com base nos resultados conhecidos (THEOBALD, 2017).

Seguem alguns exemplos de algoritmos de aprendizado supervisionado: a) algoritmos de classificação: *K-Nearest Neighbors* – KNN Neighbors (k-vizinhos mais próximos), *Naïve Bayes*, *Support Vector Machine* – SVM (Máquina de Vetores de Suporte) e b) algoritmos de regressão: *Linear Regression* (Regressão Linear), *Logistic Regression* (Regressão Logística), *Decision Trees* (Árvores de Decisão). Domingos (2017) cita alguns exemplos de aplicações para os algoritmos de aprendizagem supervisionada. O *Naïve Bayes*, além de ser um classificador de e-mails (*spam* e não *spam*) (2017, p. 175) também pode ser usado para leitura de notícias (2017, p. 177). Já o *Kinect* da *Microsoft* usa árvores de decisão para identificar partes do corpo a partir de sua câmera com sensor de profundidade (2017, p. 112).

Já em um ambiente de aprendizado não supervisionado, Theobald (2017) explica que não há padrões previamente conhecidos para servir de base na análise. Nesse caso, o modelo deve revelar padrões ocultos. Por exemplo, agrupar dados com base no peso e na altura de alunos de dezesseis anos, do ensino médio. Nesse exemplo, um grande conjunto será do sexo masculino e o outro do sexo feminino, isso porque meninas e meninos tendem a ter diferenças em relação ao peso e à altura. Uma das grandes vantagens do aprendizado não supervisionado é justamente a descoberta de padrões não conhecidos. Seguem exemplos de algoritmos de aprendizado não supervisionado: *K-Means Clustering*, *Dimensionality Reduction* (Redução da Dimensionalidade) e *Artificial Neural Networks* (Redes Neurais Artificiais).

De acordo com Theobald (2017), entre os tipos de aprendizagem de máquina apresentados, o aprendizado por reforço é o mais avançado, pois com base em *feedbacks* de interações anteriores, esse tipo de aprendizagem melhora continuamente seu modelo. Sendo que os veículos autônomos são ótimos exemplos desse tipo de aprendizagem. O *Q-learning* é um exemplo de algoritmo de aprendizado por reforço. A Figura 9 mostra os tipos de aprendizado de máquina e exemplos de seus algoritmos.

Figura 9 - Tipos de Aprendizado de Máquina e seus algoritmos



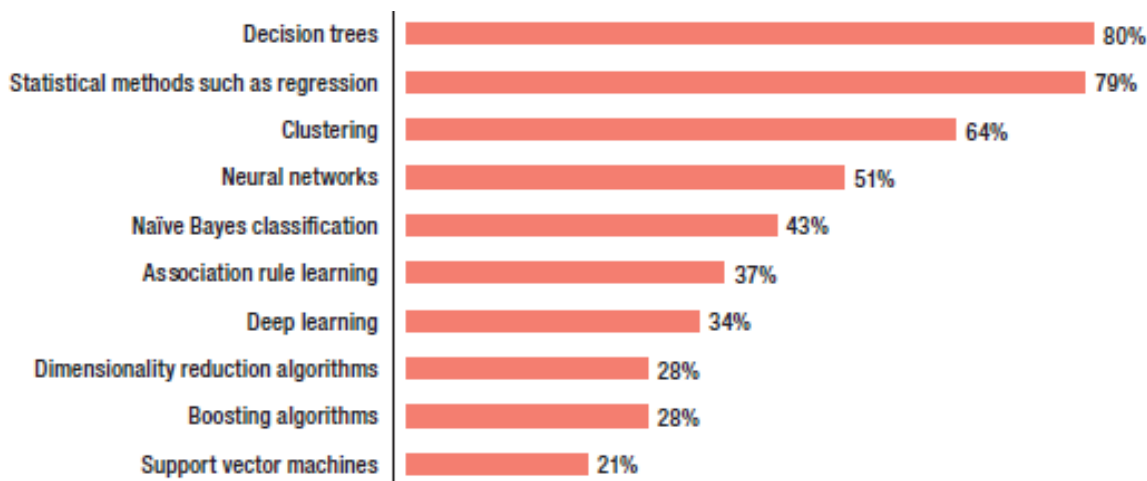
Fonte: Theobald (2017)

Devido à existência de inúmeros algoritmos que podem ser aplicados no aprendizado de máquina, Halper (2017) fez uma pesquisa, com 90 entrevistados, para levantar quais são os algoritmos mais utilizados.

A Figura 10 mostra o gráfico com o percentual dos algoritmos de aprendizado de máquina mais utilizados que foram levantados na pesquisa de Halper (2017)²⁰.

Figura 10 - Algoritmos de Aprendizado de Máquina mais utilizados

²⁰ Essa pesquisa abrange muitos outros dados que foram utilizados no decorrer deste documento. Seguem detalhes da amostra utilizada para realização da pesquisa de Halper (2017): as pessoas entrevistadas são da área corporativa, de TI, executivos, patrocinadores e consultores. A maioria reside nos EUA e na Europa e são de empresas de todos os tamanhos e segmentos como empresas de consultoria, software/Internet, serviços financeiros, indústrias de saúde, manufatura e outras indústrias, além do governo.



Fonte: Halper (2017)

De acordo com Halper (2017), as árvores de decisão (*Decision Trees*) e a regressão linear (*Statistical methods such as regression*) foram apontadas como os dois algoritmos mais utilizados. Ainda segundo Halper (2017), a regressão linear procura modelar a relação entre variáveis ajustando uma linha aos dados observados. Esse modelo é amplamente utilizado em estatística. As árvores de decisão também são populares, especialmente entre profissionais que não são estatísticos. Uma árvore de decisão é do tipo de aprendizado supervisionado que requer um conjunto de dados de treinamento com resultados conhecidos. Esse algoritmo usa uma abordagem de ramificação ou de árvore para modelar variáveis de um interesse específico. Por exemplo, um resultado pode ser comprar ou não comprar. Para Halper (2017), a popularidade das árvores de decisão ocorre porque elas são fáceis de entender e não são “caixa preta” (2017, p. 20), como acontece com os algoritmos de aprendizado profundo (*deep learning*).

De acordo com a Figura 10, algoritmos de *Clustering* também foram considerados como uma abordagem popular de aprendizado de máquina. Halper (2017) observa que o *Clustering* é uma técnica de aprendizado não supervisionado em que os agrupamentos (*clusters*) são feitos com base em semelhanças e a saída não requer ser previamente conhecida.

A pesquisa de Halper (2017) também mostrou que os algoritmos de redes neurais (*Neural Networks*) e *Naïve Bayes* (*Naïve Bayes classification*) também estão ganhando espaço. Halper (2017) explica que os classificadores *Naïve Bayes* são algoritmos baseados no teorema de Bayes, que examina a probabilidade de uma instância fazer parte de uma determinada classe, em que cada característica

classificada é independente do valor de qualquer outro atributo. Esses classificadores costumam ser rápidos e isso é um dos pontos fortes.

2.4 Tecnologias (*Hardware, Frameworks* e Linguagens de Programação)

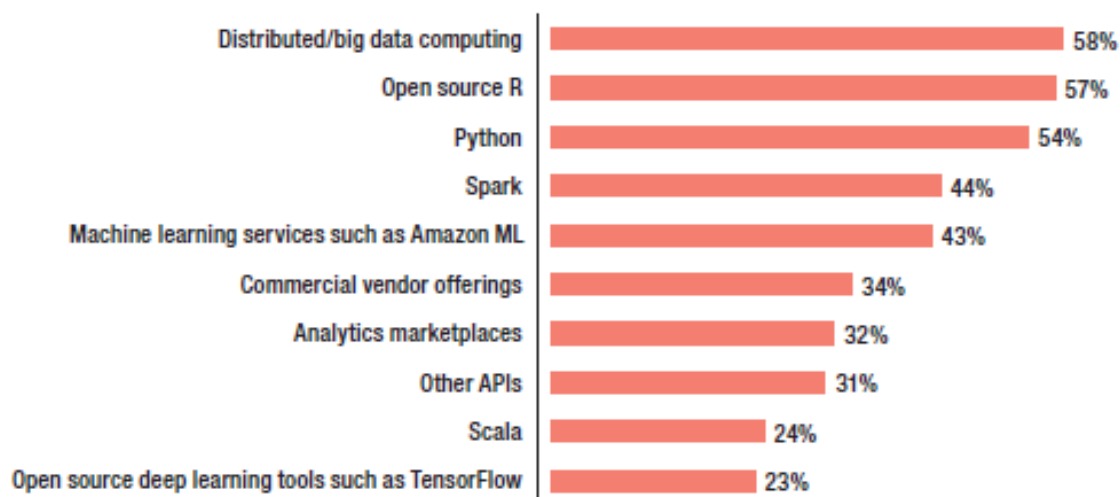
Este item visa tratar sobre as tecnologias (*hardware, frameworks*²¹ e linguagens de programação) que estão diretamente relacionadas ao aprendizado de máquina, de forma que sem esses recursos o aprendizado de máquina não teria se consolidado.

A primeira tecnologia que merece destaque é o *hardware*. Foi graças à Unidade de Processamento Gráfico (*Graphical Processing Unit - GPU*) utilizada nos jogos de vídeo *game* que o aprendizado de máquina se tornou realidade. Theobald (2017) explica que os algoritmos de aprendizagem de máquina executam mais rapidamente o processamento de um enorme volume de dados quando utilizam uma GPU. Isso ocorre porque a GPU é capaz de realizar muito mais operações matemáticas do que a Unidade Central de Processamento (*Central Processing Unit - CPU*), permitindo a resolução de problemas de álgebra linear e estatística de forma mais rápida e eficiente. Como o aprendizado de máquina requer um alto volume de cálculo isso explica o motivo da GPU ter papel tão importante.

Para levantar quais tecnologias são mais utilizadas no aprendizado de máquina e processamento de linguagem natural, Halper (2017) realizou uma pesquisa com 119 pessoas. O resultado desse levantamento foi sintetizado no gráfico da Figura 11.

Figura 11 - Tecnologias de Aprendizado de Máquina mais utilizadas

²¹ De acordo com Jaques (2016), o *framework* pode ser considerado como “uma caixa de ferramentas” com funcionalidades prontas para serem utilizadas (2016, *online*).

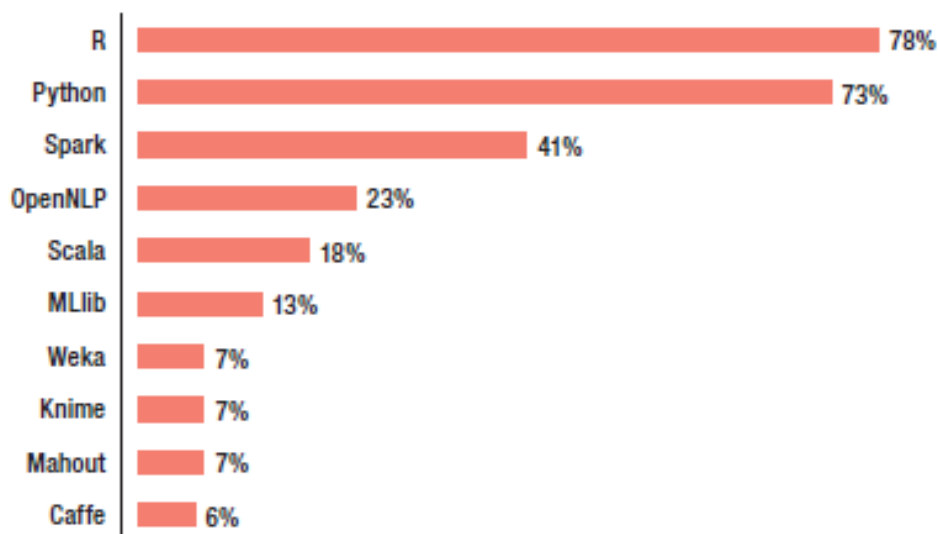


Fonte: Halper (2017)

A análise do gráfico mostra que computação distribuída e *big data* (*Distributed/big data computing*) estão no topo da lista seguida da linguagem de programação R (*Open source R*). *Python*, que ocupa o terceiro lugar no ranking também é uma linguagem de programação bastante popular. Outra informação importante segundo o gráfico, é que ferramentas comerciais (*Machine learning services such as Amazon ML*, *Commercial vendors offerings* e *Analytics marketplaces*) pontuaram abaixo das ferramentas de *software* livre (*R*, *Python*, *Spark*). No entanto, Halper (2017) considera que algumas organizações utilizam ferramentas de *software* livre para criar os modelos e depois fazem uso de ferramentas comerciais para colocá-los em produção.

Halper (2017) ainda levantou quais são as tecnologias *open source* de aprendizado de máquina e processamento de linguagem natural mais utilizadas. Nesse levantamento, ela entrevistou 130 pessoas. O gráfico da Figura 12 mostra o resultado dessa pesquisa.

Figura 12 - Tecnologias *open source* mais utilizadas no Aprendizado de Máquina



Fonte: Halper (2017)

De acordo com o gráfico, R está em primeiro lugar seguido do *Python* e *Spark*. A quarta posição é ocupada pelo *OpenNLP*. O levantamento de Halper (2017) também aponta que as ferramentas de *software* livre são as mais populares, embora seja comum o uso de ferramentas de *software* livre em conjunto com ferramentas comerciais. Segue abaixo a descrição das quatro ferramentas mais utilizadas.

Criada nos anos 1990, R é uma linguagem de programação amplamente utilizada pelas universidades. Possui recursos para armazenamento, manuseio e análise gráfica de dados, além de ferramentas de aprendizado de máquina e processamento de linguagem natural (HALPER, 2017). De acordo com Theobald (2017), R é baseada em *software* livre e tem como principal foco as operações matemáticas e o desenvolvimento de matrizes e funções estatísticas. Embora o ponto forte do R seja análise e mineração de dados, operações de aprendizado de máquina são bem suportadas.

A linguagem *Python* foi desenvolvida em 1990, pela *Python Foundation*, tendo como principal intuito ser simples e fácil. É uma linguagem de programação interpretada, interativa e orientada a objetos. Atualmente é utilizada em conjunto com várias ferramentas como *Hadoop* e *Spark*. Entre as bibliotecas, tem o *NumPy* utilizada para manipular dados e o *Scikit-learn* para algoritmos de aprendizado de máquina. Também possui o *NLTK* que é um conjunto de bibliotecas para processamento de linguagem natural (HALPER, 2017).

Spark, segundo Penchikala (2015), foi desenvolvido em 2009 pelo AMPLab da Universidade da Califórnia. Em 2010, seu código foi aberto como projeto da

Fundação *Apache*. Se trata de um *framework* para processamento de *big data* fornecendo recursos de processamento para vários tipos de dados (texto, gráfico, *streaming*), além de possuir bibliotecas como o *MLlib* focados no aprendizado de máquina (HALPER, 2017).

OpenNLP é um *kit* de ferramentas da *Apache* para processamento de linguagem natural. Halper (2017) esclarece que o *OpenNLP* suporta tarefas como tokenização (divisão das sentenças em partes para utilização em outros processos), segmentação de sentenças (marcação de substantivos, verbos, além de extração de entidades como pessoas, organizações, etc) e análise para detecção de idioma entre outras funcionalidades.

A ferramenta *Caffe*, embora tenha ficado na última posição do gráfico, merece destaque porque de todas as ferramentas apontadas é a única voltada ao aprendizado profundo (*deep learning*). De acordo com o site²², *Caffe* se trata de um *framework* para *deep learning*, desenvolvido na Universidade da Califórnia. É *open source* e está escrito em C ++, com uma interface *Python*.

Para Theobald (2017), a utilização do *Python* cada vez mais vem crescendo e essa popularidade ocorre por ser uma linguagem fácil e abrangente. Destaca que o *Python* pode ser utilizado na coleta de dados (*web scraping*), desenvolvimento de aplicações distribuídas, cuja principal ferramenta é o *Hadoop*; e também em *frameworks* de processamento de *big data*, como o *Spark*, conforme já mencionado por Halper (2017). Outra grande contribuição para a popularidade do *Python*, segundo Theobald, foi o fato de o *TensorFlow*, algoritmo do *Google*, possuir uma interface *Python* para a linguagem C. Isso permite que o *Python* seja executado em uma GPU. Essa interface é necessária porque a GPU usa linguagens de programação como C e C++ para execução das operações matemáticas. Além disso, o *Python* oferece uma variedade de bibliotecas de aprendizado de máquina, entre elas *NumPy*, *Pandas* e *Scikitlearn*. Theobald (2017) fez uma análise dessas bibliotecas, conforme o resumo a seguir:

O *NumPy* permite gerenciar matrizes, além de trabalhar com grandes conjuntos de dados. Essa ferramenta é utilizada na fase inicial do tratamento dos recursos (dados) em conjunto com o *Pandas* que permite que os dados sejam representados

²² <https://caffe.berkeleyvision.org/>

em uma planilha virtual. Já o *Scikitlearn* possui vários algoritmos de aprendizado de máquina, como regressão linear e máquinas de vetores de suporte (SVM). Em suma, os usuários podem utilizar o *NumPy*, o *Pandas* e *Scikitlearn* simultaneamente, carregando os dados via *NumPy*, limpando e fazendo cálculos com *Pandas* e, por fim, executando algoritmos de aprendizado de máquina com o *Scikitlearn*.

A Figura 13 apresenta um resumo das tecnologias relacionadas com o aprendizado de máquina tomando por base as informações de Halper (2017) e Theobald (2017).

Figura 13 - Tecnologias do Aprendizado de Máquina

Hardware	Linguagens de Programação	Frameworks	Bibliotecas
<ul style="list-style-type: none"> • GPU. 	<ul style="list-style-type: none"> • R; • Python; • C; • C++. 	<ul style="list-style-type: none"> • Hadoop; • Spark; • TensorFlow; • Amazon ML; • OpenNLP; • Scala; • Weka; • Knime; • Mahout; • Caffe. 	<ul style="list-style-type: none"> • NumPy; • Pandas; • Scikit-learn; • NLTK; • Mlib.

Fonte: Autora

O estudo de Halper (2017) revelou uma tendência do mercado: a de introduzir ferramentas totalmente automatizadas para auxiliar analistas de negócios e usuários de outras áreas a criarem modelos de aprendizado de máquina de forma ágil. A justificativa é a escassez de cientistas de dados e estatísticos. Com esse tipo de ferramenta tudo o que o usuário precisa fazer é fornecer as variáveis necessárias e os atributos que considerar preditivos. A ferramenta de forma autônoma escolherá o melhor modelo e espera-se que também forneça detalhes sobre a estatística e a matemática utilizadas.

Para Halper (2017), essas ferramentas podem impulsionar a produtividade das organizações que as utilizam. Mas, é importante que elas sejam transparentes e que os usuários tenham alguma compreensão do que acontece nos bastidores. Ou

seja, ter conhecimento básico das técnicas de aprendizado de máquina que ajude na interpretação e na defesa da análise.

A maioria dos entrevistados por Halper (2017) considera que o principal benefício do uso de aprendizado de máquina é entender o comportamento das pessoas (clientes, pacientes, etc) utilizando diversos tipos de dados. Outros benefícios identificados são *insights* para negócios, diferenciação competitiva e análise de dados para monetização. Embora, haja benefícios e valor nessas tecnologias, também existem vários desafios, como a necessidade de habilidades específicas, suporte e custo. Além desses obstáculos operacionais, questões técnicas como tratamento e integração dos dados e criação das massas de dados para a fase de treinamento demandam tempo. O próximo item mostra quais são as áreas em que o aprendizado de máquina está sendo utilizado.

2.5 Aplicações do Aprendizado de Máquina

De acordo com Halper (2017), as organizações estão utilizando aprendizagem de máquina de várias formas, como na análise de risco, análise de comportamento do cliente e manutenção preventiva. O processamento de linguagem natural, segundo Halper (2017), vem sendo aplicado na criação de *chatbots* interativos e aplicativos B2B. A aprendizagem profunda está sendo empregada com bastante ênfase para classificar imagens e diagnosticar doenças.




Para Halper (2017), o aprendizado de máquina, o processamento de linguagem natural e o aprendizado profundo estão sendo usados em uma variedade de segmentos de negócios. Halper (2017) realizou uma pesquisa com 130 pessoas e o resultado mostrou que 58% dos entrevistados aplicam esse tipo de tecnologia na área de operações, enquanto 53% estão usando no marketing. As demais áreas que empregam essas tecnologias são: TI; Serviço ao cliente; Vendas; Finanças; *E-commerce* e RH.

Essa pesquisa também demonstrou que essas tecnologias são utilizadas para: fazer manutenção preventiva; otimização de equipamento e de processos; previsão de fraude e desperdício; análise de redes sociais e sentimentos; análise da taxa de cancelamento (*churn*) e outros comportamentos do cliente; intervenção proativa; previsão de falha de equipamentos de TI; detecção de ameaça de segurança; *helpbots* e *chatbots*; análise de *feedback* do produto com base no texto; análise de interação pós-compra; previsão de vendas; previsão de *mix* de produtos; *score* de

crédito; categorização de transações financeiras; *chatbots* para sugerir produtos em *websites*; classificação de produtos por imagem; categorização automatizada de URLs; rotatividade de funcionários; *bots* internos do RH e análise das comunicações internas. A Tabela 3 demonstra o percentual de uso e a forma de aplicação em função do segmento.

Tabela 3 - Porcentagem de uso do Aprendizado de Máquina por segmento

AREA	% DE USO DA TECNOLOGIA E EXEMPLOS
Operações 	 58% <ul style="list-style-type: none"> • Manutenção preventiva • Otimização de equipamento • Otimização de processos, incluindo gestão da cadeia logística (<i>supply chain</i>) • Previsão de fraude e desperdício
Marketing 	 53% <ul style="list-style-type: none"> • Análise de redes sociais e sentimentos • Análise da taxa de cancelamento (<i>churn</i>) e outros comportamentos do cliente • Intervenção proativa
TI 	 48% <ul style="list-style-type: none"> • Predição de falha em equipamentos • Detecção de ameaça de segurança
Serviço ao cliente 	 42% <ul style="list-style-type: none"> • <i>Helpbots</i>, <i>chatbot</i> • Análise de <i>feedback</i> do produto com base no texto • Análise de interação pós-compra
Vendas 	 31% <ul style="list-style-type: none"> • Previsão de vendas • Previsão de <i>mix</i> de produtos
Finanças 	 27% <ul style="list-style-type: none"> • <i>Score</i> de crédito • Categorização de transações financeiras
E-commerce 	 18% <ul style="list-style-type: none"> • <i>Chatbots</i> para sugerir produtos em <i>websites</i> • Classificando produtos por imagens

AREA	% DE USO DA TECNOLOGIA E EXEMPLOS
	<ul style="list-style-type: none"> • Categorização automatizada de URLs
RH 	 15% <ul style="list-style-type: none"> • Rotatividade de funcionários • <i>Bots</i> internos do RH • Análise das comunicações internas

Fonte: Halper (2017)

Nos segmentos voltados para áreas como Operações, Halper (2017) explica que a manutenção preventiva visa o gerenciamento de ativos. Por exemplo, usa os dados de temperatura, número de rotações por segundo de uma peça que são gerados por sensores da Internet das Coisas (IoT) para analisá-los usando o aprendizado de máquina. O objetivo é prevenir falha de uma peça ou algum desgaste indevido. Para isso, o algoritmo aprende padrões que constituem a necessidade de reparo. Uma vez treinado, o modelo poderá ser utilizado com novos dados a fim de melhorar a manutenção e eficiência operacional de ativos.

Em áreas como Serviços ao Cliente, Halper (2017) destaca que o uso de *chatbots* para interagir com o cliente já está disponível no mercado há vários anos. Porém, *chatbots* mais recentes possuem maior capacidade para compreensão da linguagem humana e por isso são mais interativos. Halper (2017) cita como exemplo as empresas que estão usando *helpbots* para responder e/ou encaminhar perguntas de rotinas em *Help Desks*. Este tipo de aplicação utiliza processamento de linguagem natural e aprendizado de máquina.

Halper (2017) verificou que a análise dos comentários feitos nas mídias sociais e nos sites utiliza algoritmos de processamento de linguagem natural. O intuito desse tipo de análise é levantar problemas e opiniões. E cita como exemplo o monitoramento de *tweets* e comentários sobre produtos e marcas para compreender o que as pessoas estão falando, seus sentimentos (positivo, negativo, neutro) assim como as emoções. Este tipo de análise é bastante utilizado em áreas como o Marketing.

Outra aplicação do processamento de linguagem natural, segundo Halper (2017), é servir como mecanismo de pesquisa e análise de uma base histórica de usuários para fornecer sugestões e *insights* do que pode ser mais relevante para esses usuários. Tal intervenção proativa é bastante utilizada pelo Marketing.

O processamento de linguagem natural em conjunto com o aprendizado de máquina e aprendizado profundo também pode ser usado para resumo e classificação de texto. Para Halper (2017), há vários casos de uso para essa técnica, desde a descrição dos principais itens do texto, até resumir o que as pessoas estão enviando por e-mail para uma determinada empresa. Halper (2017) explica que os algoritmos para este tipo de aplicação ainda estão em fase de estudo e testes, pois é um campo de aplicação bem avançado que pode ser utilizado no Marketing, Vendas, Serviços ao cliente, RH e *E-commerce*.

Com o aprendizado profundo, um algoritmo pode ser treinado para reconhecer imagens e sons. Quando se fala em reconhecimento de imagem, a primeira coisa que vem à mente é a marcação automática de fotos no *Facebook*, mas, de acordo com Halper (2017), esse tipo de tecnologia é comumente usado para classificar fotos em empresas para vendas de produtos ou a venda de um objeto em um leilão *online*. Este tipo de aplicação é bastante empregado em áreas como Marketing, Serviço ao cliente e *E-commerce*. Halper (2017) ainda aponta que o reconhecimento de imagem pode ser usado na medicina para classificar mamografias como potencialmente cancerosas e para entender melhor as doenças.

O aprendizado de máquina, como reforça Halper (2017), também está sendo usado para gerenciamento de dados e aplicativos de BI, que vai desde a integração até a preparação de dados para a análise.

Halper (2017) identifica que segmentos como Serviço ao Cliente, *E-commerce* e RH utilizam aplicativos B2B baseados em processamento de linguagem natural. Algumas empresas incorporam esses aplicativos em portais para que parceiros que não estejam analiticamente bem informados possam obter facilmente *insights*. Aplicativos inteligentes como *Siri* e *Alexa*, e no caso do Brasil, a Inteligência Artificial do Bradesco (BIA), utilizam interface de processamento de linguagem natural permitindo que os usuários façam perguntas por voz ou texto e obtenham respostas do aplicativo.

Com base no resultado da pesquisa realizada por Halper (2017), é possível fazer um agrupamento de aplicações por subcampos da IA (aprendizado de máquina, processamento de linguagem natural e aprendizado profundo) conforme mostra a Figura 14.

Figura 14 - Aplicações agrupadas por subcampos da IA



Fonte: Autora

O agrupamento das aplicações pelos subcampos da IA (Figura 14) mostra que o aprendizado de máquina é a tecnologia que possui o maior número de aplicações, seguida pelo processamento de linguagem natural. Um fato que pode ser observado é que aplicações muito complexas, como o resumo e classificação de texto, requer o uso conjunto dos três subcampos da IA (aprendizado de máquina, processamento de linguagem natural e aprendizado profundo). O monitoramento de redes sociais é outro tipo de aplicação que necessita do uso desses três subcampos da IA.

A pesquisa de Halper (2017) também mostra que os três principais motivos para o não uso o aprendizado de máquina e/ou processamento de linguagem natural pelas empresas, incluem ausência de suporte, orçamento e habilidades necessárias para usar esse tipo de tecnologia. Empresas que não usam esse tipo de tecnologia acreditam no valor dessas análises avançadas, porém não conseguem obter adesão ao processo. Segundo Halper (2017), muitas vezes, isso acontece porque falta uma cultura de análise.

Halper (2017) conclui com sua pesquisa que o mercado está bem animado com o aprendizado de máquina e processamento de linguagem natural,

principalmente em segmentos como o Marketing, pois essas tecnologias podem fornecer um valor significativo às organizações que utilizá-las. Porém, o orçamento, na maioria das vezes, continua sendo um obstáculo para qualquer tipo de projeto de análise.

O capítulo 2 apresentou as definições fundamentais do aprendizado de máquina. Conceitos estes que serão utilizados nos capítulos seguintes em que serão analisadas as concepções de sociedades de controle, marketing, TICs, aprendizado de máquina, marketing político, campanhas eleitorais e as implicações dessas tecnologias de controle nas disputas políticas.

3 SOCIEDADES DE CONTROLE, MARKETING E TECNOLOGIAS DA INFORMAÇÃO E COMUNICAÇÃO

Em seu *post-scriptum* sobre as sociedades de controle, Deleuze (1992) descreve que Foucault situou as sociedades disciplinares nos séculos XVIII e XIX e seu apogeu no século XX. A grande marca desse tipo de sociedade são os meios de confinamento, onde os indivíduos passam de um meio fechado para outro, de casa para a escola, da escola para a fábrica, além de hospitais e presídios. De acordo com Deleuze (1992), Foucault fez uma análise desses meios de confinamento, especialmente na fábrica, onde era preciso compor no espaço-tempo uma força produtiva. No entanto, Foucault sabia da brevidade desse modelo e do surgimento de um novo tipo de sociedade.

A era da disciplina deu lugar a uma nova era que Deleuze denominou era do controle. Lazzarato (2006) aponta algumas características das sociedades disciplinares para facilitar a compreensão dessa nova era. De acordo com ele, as sociedades disciplinares são caracterizadas pelo poder disciplinar e biopolítico. As técnicas disciplinares transformam os corpos, enquanto as técnicas biopolíticas são aplicadas na massa, em processos como o nascimento, a morte, a reprodução e a doença. A disciplina conhece o corpo e o indivíduo, já o biopoder foca na população e na gestão da vida.

Em relação às sociedades de controle, Lazzarato (2006) explica que elas criaram suas próprias tecnologias e processos de subjetivação que são bem diferentes das tecnologias e processos das sociedades disciplinares. Com o intuito de descrever a sociedade de controle, ele recorre a três características desse novo tipo de tecnologia de poder: 1) a emergência da cooperação entre cérebros e seu funcionamento por fluxos e redes; 2) dispositivos tecnológicos que agem à distância e amplificam a potência de ação, tais como a televisão e a Internet e 3) a formação dos públicos por meio dos processos de subjetivação e sujeição (LAZZARATO, 2006, p. 76).

Em relação à cooperação entre os cérebros, Lazzarato (2006) explica que ela se expressa sob a forma de opinião pública, mediada pela tecnologia. Para ele, a “Internet integra e distingue as diferentes transformações da opinião pública, da percepção e da inteligência coletiva” (LAZZARATO, 2006, p.77). Além disso, o autor

destaca que a divisão da sociedade em públicos se sobressai da divisão religiosa, econômica, estética, política, porém sem substituí-las, e complementa:

Tarde nos oferece uma síntese bastante clara e eficaz da passagem das sociedades disciplinares às sociedades de controle: Ao regime anárquico da cobiça sucedeu o governo autocrático da opinião, que se tornou onipotente. As funções políticas e econômicas da opinião não podem mais ser reconduzidas aos mecanismos de exploração e sujeição específicos das sociedades disciplinares. (LAZZARATO, 2006, p. 79).

Lazzarato (2006) considera que “a captura, o controle e a regulação da ação à distância das mentes entre si se faz por meio da modulação dos fluxos de desejos, crenças e das forças (memória e atenção) que circulam entre os cérebros” (2006, p. 84). Para ele, “a memória, a atenção e as relações que elas atualizam tornam-se forças sociais e econômicas que precisam ser capturadas para que possam ser controladas e exploradas” (2006, p. 84). O filósofo aponta isso como sendo outra diferença entre as sociedades disciplinares e as sociedades de controle. De acordo com ele, esta última investe na memória mental e não no corpo. As “disciplinas moldavam os corpos” (2006, p.86) construindo hábitos, já as sociedades de controle modulam os cérebros criando hábitos na memória.

Existe, portanto, uma moldagem dos corpos, garantida pelas disciplinas (prisões, escola, fábrica), a gestão da vida organizada pelo biopoder (Estado-providência, políticas de saúde) e a modulação da memória e suas potências virtuais reguladas pela noopolítica (redes hertzianas, audiovisuais, telemática e constituição da opinião pública, da percepção e da inteligência coletiva) (LAZZARATO, 2006, p. 86).

Para Deleuze (1992), os diferentes modos de controle, os “controlatos”, são variações inseparáveis, formando um sistema cuja linguagem é numérica. Os confinamentos são moldagens fixas; já os controles são uma modulação que muda constantemente. Nas sociedades de controle, “o essencial não é mais uma assinatura” e uma “matrícula” como ocorria nas sociedades disciplinares, mas sim, uma “cifra” ou senha. “Os indivíduos tornaram-se individuais e as massas tornaram-se amostras, dados, mercados, bancos” (DELEUZE, 1992, p. 226).

A sociedade de controle exerce seu poder utilizando tecnologias de ação e controle à distância que enviam imagens, som e informações por meio de máquinas de modular. Essas tecnologias são “formas ultrarrápidas de controle ao ar livre que substituem antigas disciplinas que operavam em um sistema fechado”. (DELEUZE, 1992, p. 224). Para Lazzarato (2006), neste tipo de sociedade “as relações de poder

são midiaticizadas e enriquecidas pela tecnologia” (2006, p. 76). Além disso, essas máquinas de modular são dispositivos com capacidade para “intervir no acontecimento e na cooperação entre os cérebros” modulando as “forças envolvidas nessa cooperação”. Por isso, se tornaram um requisito necessário em todo processo de criação de “subjetividade” (LAZZARATO, 2006, p. 85-86).

Assim como é preciso compreender as diferenças entre as sociedades disciplinares e as sociedades de controle, é necessário diferenciar os dispositivos de manipulação dos dispositivos de modulação. Para Silveira (2017), as empresas do ramo jornalístico podem ser consideradas como dispositivos para expressão da “prática dos jogos de verdade”, ou seja, como dispositivos de manipulação. “A verdade é em si poder. Uma verdade é construída manipulando elementos da realidade, unindo em determinado sentido os fatos, selecionando o que relatar e o que desconsiderar ou omitir” (SILVEIRA, 2017, p. 1229).

De acordo com Foucault (1997), as práticas discursivas não são apenas modos de fabricação de discursos. Elas ganham corpo em conjuntos técnicos, em instituições, em esquemas de comportamento, em tipos de transmissão e de difusão que ao mesmo tempo as impõem e as mantêm. Foucault (2015) descreve que em nossa sociedade:

A “economia da política” da verdade tem cinco características historicamente importantes: a verdade é centrada na forma do discurso científico e nas instituições que o produzem; está submetida a uma constante incitação econômica e política; é objeto de uma imensa difusão e de um imenso consumo; é produzida e transmitida sob o controle, não exclusivo, mas dominante, de alguns grandes aparelhos políticos ou econômicos; por fim é objeto de debate político e de confronto social (FOUCAULT, 2015, p. 52).

Por isso, na percepção de Silveira (2017), os dispositivos de modulação são diferentes dos dispositivos de manipulação ou de criação de verdade por meio do discurso. Para ele, os “moduladores” são “actantes”²³, “humanos” e “não humanos” que visam intermediar e facilitar o cotidiano das pessoas (2017, p. 1229). Mas, como identificar esses dispositivos de modulação e entender como o marketing está utilizando esse tipo de tecnologia de poder? O próximo item procura demonstrar como isso funciona.

²³ Actante é tudo aquilo que gera uma ação, que produz movimento e diferença, seja ele humano ou não humano. O actante é o mediador, ou seja, é aquele que transforma, traduz, distorce e modifica o significado que ele supostamente transporta (Praude, 2015 apud Latour, 2012; Latour, 2000).

3.1 Aprendizado de Máquina e Modulação do Comportamento Humano

Para Lazzarato (2006), a meta das sociedades de controle é a criação e oferta de mundos “de consumo, de informação, de trabalho, de lazer” entre outros (2006, p.101). Contudo, esses mundos são formatados e sem nenhuma singularidade. Bauman (2013) mostra que nos tornamos uma “sociedade de consumidores”, em que a cultura se “manifesta como arsenal de artigos destinados ao consumo”, competindo pela atenção dos possíveis clientes. Essa é uma característica de uma sociedade que substituiu a estrutura por rede, em que a “conexão e desconexão” sucederam a “determinação, a lealdade e o pertencimento” (BAUMAN, 2013 p.18-19).

Diante desses mundos normalizados, nossa liberdade é exercida exclusivamente para escolher dentre possíveis que outros instruíram e conceberam. Ficamos sem o direito de participar da construção dos mundos, de formular problemas e de inventar soluções, a não ser no interior de alternativas já estabelecidas (LAZZARATO, 2006, p. 101-102).

Para Lazzarato (2006), somos submetidos à força e ao papel estratégico desempenhado pelas máquinas de expressão que são representadas pela “opinião, comunicação e marketing”. Nosso mundo é formado pelos “agenciamentos de enunciação e pelos regimes de signos” que são expressos pela publicidade que dita como “se vestir, ter um corpo, comer, comunicar, morar” e assim por diante. Deleuze estava certo ao dizer que o “marketing tornou-se centro estratégico e que os publicitários são criativos” (LAZZARATO, 2006, p. 100-101).

Segundo Lazzarato (2006), a publicidade dissemina formas de sentir para provocar maneiras de viver. Por conseguinte, as agências de publicidade inventam o que deve ser encarnado nos corpos e espalham essa criação por meio do fluxo de informações e comunicação divulgados pelo rádio, televisão e pelas redes de computadores. Na realidade, isso não passa de uma “palavra autoritária expressa pela sedução” (2006, p. 103). Ele afirma que os mundos da publicidade são “fechados e totalitários”, pois impedem outros “mundos possíveis” de existir (2006, p. 105).

Com a evolução da propaganda, uma máquina muito bem estruturada de actantes não humanos se desenvolveu, da qual o *Google* e o *Facebook* se destacam: “imensos bancos de dados” (*big data*) e *softwares* de aprendizado de máquina que atuam como dispositivos de reconhecimento de padrões,

comportamentos e de marketing. Esses dispositivos “reúnem, selecionam e vendem milhões de dados sobre nossas aquisições, hábitos de leitura, filmes favoritos, gostos, roupas, bem como o modo como passamos nosso tempo livre” (LAZZARATO, 2014, p. 38). Todos esses conjuntos de informações configuram os “individuais”, “amostras” e “dados” definidos por Deleuze, cuja finalidade é a modulação (DELEUZE, 1992, p. 226).

É fácil fazer corresponder a cada sociedade certos tipos de máquina, não porque as máquinas sejam determinantes, mas porque elas exprimem as formas sociais capazes de lhes darem nascimento e utilizá-las. As antigas sociedades de soberania manejavam máquinas simples, alavancas, roldanas, relógios; mas as sociedades disciplinares recentes tinham por equipamento máquinas energéticas, com o perigo passivo da entropia e o perigo ativo da sabotagem; as sociedades de controle operam por máquinas de uma terceira espécie, máquinas de informática e computadores, cujo perigo passivo é a interferência, e o ativo a pirataria e a introdução de vírus. (DELEUZE, 1992, p. 223).

Essa “megamáquina” aos poucos foi diminuindo o “número de operadores humanos, não confiáveis”, e aumentando os “agentes eletrônicos confiáveis”. “As novas máquinas sociais e técnicas” e seus dispositivos vão além da fábrica, elas se apoderaram do comportamento e das ações das pessoas, não apenas no trabalho, mas também na vida cotidiana (LAZZARATO, 2014, p. 34-35).

Para Silveira (2017), a modulação utiliza ferramentas que usam dados pessoais que são indispensáveis para o marketing. Depois de coletar e analisar esses dados, as empresas de marketing definem perfis que servem aos dispositivos de modulação. A “partir dos gostos, do temperamento, das necessidades, das possibilidades financeiras, do nível educacional”, entre outras categorias, as empresas oferecem “soluções, produtos e serviços” para estes possíveis consumidores que tiveram seus dados tratados e analisados (SILVEIRA, 2017, p. 1285). O sucesso da modulação depende da análise rigorosa das pessoas que serão moduladas.

No entanto, “modular o que passará a ser imprescindível ou útil para o cotidiano de uma pessoa ou grupo social é muito complexo” (SILVEIRA, 2017, p. 1293). Está relacionado com a prática preditiva que depende de tecnologias como *big data*, aprendizado de máquina e estudos das ciências cognitivas. De acordo com Silveira, atualmente percebe-se o uso cada vez maior das neurociências e da psicologia na composição das equipes de marketing. A predição do que as pessoas

estão interessadas em saber pode gerar possibilidades de “modular caminhos, restringir escolhas e incentivar opções” (SILVEIRA, 2017, p. 1293). Mas, qual é a relação entre aprendizado de máquina, rastros digitais dos usuários (*big data*), modulação do comportamento humano e marketing?

Michal Kosinski e seus artigos são ótimas fontes de pesquisa para entender o uso dos dados pessoais para predição e modulação do comportamento humano. Ele é psicólogo, cientista de dados e professor assistente na *Stanford Graduate School of Business*. Possui doutorado em Psicologia, mestrado em Psicometria e Psicologia Social. Kosinski pesquisa os seres humanos por meio de seus rastros digitais em plataformas e dispositivos. Para isso, combina conceitos de Psicometria, Ciência Social, Psicologia, *big data* e aprendizado de máquina. Coordena o projeto *MyPersonality*, que envolve colaboração global entre mais de 200 pesquisadores, analisando os perfis psicodemográficos de milhões de usuários do *Facebook*.

O aplicativo *MyPersonality* foi criado durante o doutorado de Michael Kosinski, na Universidade de Cambridge. Em sua pesquisa foram disponibilizados vários questionários psicométricos, a partir de perguntas retiradas do *Big Five*. O objetivo era que alguns amigos preenchessem os questionários, mas surpreendentemente “milhões lhes revelaram suas crenças, sentimentos e comportamentos”. Como resultado, eles se viram diante do maior banco de dados de personalidade da história (GUARESCHI, 2017, p, 173).

Avaliações psicométricas fazem parte de um ramo da psicologia denominado psicometria que é uma área da Psicologia que procura analisar e medir características psicológicas e a personalidade individual. Um dos mecanismos utilizados é o *Big Five* (cinco dimensões centrais da personalidade dos seres humanos). Essas avaliações foram compostas por diferentes psicólogos, na década de 1980, por meio de uma análise fatorial dos dados fornecidos pelas pessoas pesquisadas a partir de “inventários de personalidade” e de “extensos questionários”. Todas as informações foram sintetizadas em cinco principais fatores: abertura ao novo, consciência, sociabilidade, extroversão e traços neuróticos. (GUARESCHI, 2017, p. 173).

De acordo com Kosinski, Stillwell e Graepel (2013), a predição de características e atributos individuais com base em testes psicométricos vem de longa data. Atualmente, as atividades humanas e suas interações sociais são mediadas por plataformas e dispositivos digitais. Essa crescente imersão em

ambientes digitais e a dependência dos dispositivos fizeram com que os traços de comportamentos, comunicação, interações sociais e os passos das pessoas fossem facilmente registrados, gerando imensos rastros digitais (KOSINSKI et al, 2016).

Para Kosinski et al (2016), a disponibilidade desses rastros digitais, combinado com poder de computação e modernas ferramentas estatísticas, oferece inúmeras oportunidades para a ciência social, pois grandes volumes de dados facilitam a descoberta de padrões que podem não estar disponíveis em amostras menores, além de contribuir para reduzir erros de amostragem. Modelos preditivos baseados em pegadas digitais também podem ser usados para desenvolver ferramentas de diagnóstico e medidas psicométricas úteis para pesquisa e prática psicológica. Além dessas contribuições, abriu caminho para o surgimento da ciência social computacional.

De acordo com Kosinski et al (2016), as pegadas digitais podem ser utilizadas para estudar importantes questões psicológicas e sociológicas. No entanto, o interesse por estudar esses rastros digitais e as habilidades necessárias são itens raros entre os cientistas sociais, restringindo esse tipo pesquisa a cientistas da computação e a engenheiros que, na maioria das vezes, são carentes em teorias sociais e padrões éticos referentes às pesquisas humanas. Essas amostras oferecem oportunidades para melhorar a compreensão dos indivíduos, grupos e sociedades, mas a análise das mesmas apresenta grandes desafios técnicos, éticos e metodológicos.

Com o intuito de diminuir os desafios técnicos e metodológicos, Kosinski et al (2016) elaboraram o tutorial “Mining Big Data to Extract Patterns and Predict Real-Life Outcomes²⁴”. Neste documento foi explicado o uso da linguagem R e dos algoritmos SVD (*Singular Value Decomposition*) e LDA (*Latent Dirichlet Allocation*) com vários exemplos práticos. Eles citaram algumas fontes de dados e relacionaram as etapas envolvidas no estudo de grandes conjuntos de dados: a) pré-processamento; b) redução de dimensionalidade, e c) construção de modelos preditivos. E também listaram a vantagem de grandes amostras de dados para a inclusão de indivíduos sub-representados nos estudos tradicionais, além de ressaltarem os riscos relacionados à privacidade e à má conduta científica

²⁴ <http://www.pnas.org/content/early/2017/11/07/1710966114>

ocasionada pela falta de diretrizes e conselhos éticos claros para esse tipo de atividade.

Kosinski, Stillwell e Graepel (2013) alertam que as pessoas podem até optar por não revelar certas informações sobre suas vidas, e mesmo assim essas informações serem previstas com base em outros aspectos que elas expõem por meio de seus rastros digitais. Este tipo de previsão pode ser utilizado para melhorar produtos e serviços, mas se for empregado de uma forma negativa, pode levar a invasões de privacidade.

Em 2014, Kosinski e seus colaboradores fizeram uma pesquisa para validar se os julgamentos de personalidade feitos por um computador são mais precisos do que os realizados por humanos. Esta pesquisa foi documentada no artigo “Computer-based personality judgments are more accurate than those made by humans”. De acordo com Youyoua, Kosinski e Stillwell (2014), os resultados publicados nesse artigo demonstraram que os modelos baseados em computador são significativamente mais precisos que os humanos na tarefa de julgamento da personalidade. Segundo Cadwalladr (2018), este artigo serviu de inspiração para que o cientista de computação Robert Mercer investisse na criação da *Cambridge Analytica*. Este assunto será tratado em mais detalhes no item 5.2.

Com base nesse estudo, Youyoua, Kosinski e Stillwell (2014) concluíram que as ferramentas de avaliação de personalidade automatizadas são precisas e baratas e podem afetar a sociedade de várias formas: a) mensagens de marketing podem ser adaptadas às personalidades dos usuários; b) recrutadores podem combinar melhor os candidatos com trabalhos tendo como base a personalidade dos mesmos; c) produtos e serviços podem se ajustar para melhor corresponder aos seus usuários e d) os cientistas podem coletar dados sobre personalidade sem sobrecarregar os participantes com longos questionários.

Entretanto, Youyoua, Kosinski e Stillwell (2014) destacam que o conhecimento a respeito da personalidade das pessoas também pode ser utilizado como ferramenta de manipulação e influência. Por este motivo, as pessoas podem não aceitar esse tipo tecnologia ao perceberem que governo, rede social ou mecanismos de busca podem prever suas características pessoais com mais precisão do que seus próprios familiares e amigos. Diante de tal circunstância, os autores esperam que os desenvolvedores de tecnologia e formuladores de políticas considerem esses desafios e se apoiem em leis e tecnologias de proteção da privacidade dos dados.

De acordo com Matz et al (2017), a comunicação de massa persuasiva é utilizada por governos, profissionais de marketing e partidos políticos. Ela tem por objetivo incentivar grupos de pessoas a acreditar e agir de acordo com o ponto de vista do comunicador. Estudos mostram que, quando adaptada às características e motivações psicológicas das pessoas, a comunicação de massa persuasiva se mostra muito eficaz. Contudo, quando esses estudos são realizados por meio de intensos questionários, em laboratório e com uma amostra reduzida, a eficácia é limitada devido aos vieses de resposta e de outras razões que levam o comportamento das pessoas a diferir do apresentado em laboratório. Diante desses fatos, é questionável se esses resultados podem ser generalizados para a massa no mundo real.

Pesquisas recentes no campo das ciências sociais computacionais sugerem que os perfis psicológicos das pessoas podem ser previstos com exatidão a partir dos rastros digitais que elas deixam no universo *online*, como sites pessoais, *blogs*, mensagens do *Twitter*, *Facebook*, fotos do *Instagram* entre outros. Para validar se a segmentação psicológica traçada com base nos rastros digitais é eficaz para persuasão de massa, Matz et al (2017) realizaram três experimentos com mais de 3,7 milhões de pessoas. O resultado dessa experiência comprovou a eficácia do direcionamento psicológico no contexto da persuasão digital em massa. A adaptação de apelos persuasivos aos perfis psicológicos permitiu influenciar o comportamento e escolhas dos participantes. O experimento foi medido com base na quantidade de cliques²⁵ e conversões²⁶ efetuados pelos participantes. Contudo, apesar dos três experimentos terem sido bem-sucedidos, Matz et al (2017) evidenciam pontos críticos que merecem atenção.

O primeiro deles é que: a eficácia da persuasão psicológica em larga escala no ambiente digital depende muito da alta precisão na previsão dos perfis psicológicos. A precisão é um item importante, pois quanto menor a taxa de erro mais assertiva é a previsão. Mesmo que a previsão utilize uma enorme quantidade de dados (*big*

²⁵ Taxas de cliques (CTRs) são uma medida comum de marketing digital que quantifica o número de cliques em relação ao número de vezes que o anúncio foi mostrando.

²⁶ Taxa de conversão é uma métrica de marketing que reflete o número de conversões, como downloads de aplicativos ou compras em lojas *online*, em relação ao número de vezes que o anúncio foi exibido.

data) e seja feita por aprendizagem de máquina, não está livre de limitações, como a necessidade de calibrar e atualizar constantemente o algoritmo para manter a alta precisão. Matza et al (2017) dão o seguinte exemplo: Gostar de *Game of Thrones*, quando a série de TV foi lançada em 2011, poderia ser um indicador de introspecção; contudo, a crescente popularidade dessa série pode ter tornado isso menos previsível. Outro ponto que pode afetar a precisão é que, apesar da avaliação psicológica dos rastros digitais permitir definir o perfil de um grande número de pessoas sem a necessidade de preenchimento de questionários, a maioria dos algoritmos é desenvolvida com base nesses questionários e pode absorver alguns problemas dos mesmos.

Por fim, e mais importante, Matz et al (2017) chamam a atenção para o fato de que a implementação da persuasão psicológica em massa, embora ofereça várias oportunidades, acarreta enormes riscos e desafios éticos. Por um lado, poderia ser utilizada para ajudar as pessoas a tomar melhores decisões e aliviar problemas sociais. Por exemplo, identificar indivíduos que apresentam sinais precoces de depressão e direcionar materiais, aconselhamento profissional e ajuda. Por outro lado, a persuasão psicológica pode ser usada para explorar pontos fracos ou para modular as pessoas a se comportarem de maneira contrária aos seus interesses. Os autores inclusive citam relatos da mídia alegando que a campanha presidencial dos EUA, de 2016, usou perfis psicológicos de milhões de cidadãos americanos para mantê-los longe das cédulas no dia das eleições (mais detalhes sobre isso serão descritos no item 5.2). Eles também levantaram a questão da falta de legislação para proteção da privacidade dos dados no ambiente digital. Embora, na pesquisa eles tenham tido o cuidado de realizar uma segmentação indireta mantendo o anonimato e privacidade individual dos participantes, os experimentos também poderiam ser usados para revelar traços íntimos dos indivíduos sem o consentimento dos mesmos.

Lendo os artigos de Michal Kosinski percebe-se que no início, o interesse era apenas colocar os pesquisados em algumas das dimensões do *Big Five*. No entanto, aos poucos Kosinski e seus colaboradores começaram a comparar os resultados com o que os participantes curtiam e postavam no *Facebook*. Posteriormente passaram a cruzar esses dados com informações de outras plataformas *online*. Dessa forma surgiu a possibilidade de traçar perfis segmentados e comparações entre esses perfis (GUARESCHI, 2017).

Sem dúvida essa foi uma descoberta muito importante para áreas como a “psicologia da persuasão, publicidade, política e economia”, pois “dependendo da quantidade de pessoas investigadas nas redes sociais” e dos rastros digitais desses indivíduos em outras plataformas, é possível criar “perfis minuciosos e detalhados” que podem ser utilizados para estudos de todo o tipo, como para fins políticos, de publicidade, entre outros (GUARESCHI, 2017, p. 173).

Isso mostra o poder que o uso de enormes quantidades de dados (*big data*) pode ter na nossa sociedade. Não há nenhuma restrição tecnológica que impeça os computadores de analisar milhões de dados sobre o comportamento das pessoas e extrair tendências suficientes para que se possa conhecer o que elas pensam, acreditam, gostam, quais valores defendem e as suas motivações. Não há dúvidas de que “sempre haverá pessoas que discordam de determinadas opiniões ou valores”, “mas isso não tem importância para quem sabe que seus anúncios vão atingir milhões influenciados por essas mensagens” (Guareschi, 2017, p. 175). É claro que é difícil antecipar, com um alto grau de precisão, qual é a opção política de uma determinada pessoa. Mas, para quem deseja saber qual o político preferido, ou quais as qualidades exigidas pelas pessoas para votarem em determinado candidato, lança-se mão da lei das proporcionalidades, em que tudo depende do tamanho da amostragem. E tamanho não é problema para o grande volume de dados que temos hoje e ferramentas de aprendizado de máquina (GUARESCHI, 2017).

Atualmente, podemos considerar a postagem e a curtida como sendo semelhantes ao ato de responder ao “questionário ou uma entrevista dirigida”. Outras opções de “oferecimento gratuito de informação” são os compartilhamentos e comentários. À medida que os usuários postam mensagens no *Facebook*, ou em outras redes sociais, sem que tenham consciência, um algoritmo de aprendizado de máquina vai sendo constantemente construído e atualizado com base nesse imenso banco de dados “livremente oferecido”. (GUARESCHI, 2017, p. 176).

O *The Intercept*, site lançado em 2014, pelo jornalista Glenn Greenwald, pela cineasta Laura Poitras e pelo pesquisador Jeremy Scahill, teve acesso a um documento confidencial do *Facebook*. De acordo com este documento, o *Facebook* desenvolveu um *software* de *machine learning*, denominado *FB Learner Flow* que permite ofertar às empresas segmentação de público-alvo, com base no

comportamento, hábitos de consumo e tendências de posicionamentos futuros dos usuários (Biddle, 2018, *online*).

Segundo Biddle (2018), o “documento alardeia” que “a previsão de comportamentos futuros” permite a criação de campanhas publicitárias baseadas em decisões que o público-alvo ainda não tomou permitindo a terceiros alterar essas decisões. No texto, também é descrito que o *Facebook* pode varrer toda sua base de dados para levantar milhões de indivíduos que estariam “sob risco” de mudar uma “determinada marca pela concorrência”. Com base nessa informação, uma empresa poderia prever isso e direcionar uma campanha publicitária para esses usuários com a finalidade de fazê-los mudar de ideia. O *Facebook* denominou isso de *improved marketing efficiency* (eficiência de marketing aprimorada), cuja finalidade é utilizar informações da vida dos usuários para predizer se eles estão se cansando de um determinado produto ou marca. Com isso, o Facebook ofertaria o serviço de *loyalty prediction* ou previsão de fidelidade (BIDDLE, 2018, *online*).

Esse novo tipo de serviço de marketing do *Facebook* está intimamente relacionado com a prática de traçar perfis psicográficos utilizando os rastros digitais dos usuários. No caso específico do *Facebook*, isso dispara um sinal de alerta devido ao acesso irrestrito que essa empresa tem aos imensos bancos de dados sobre preferências e comportamento de seus usuários.

Embora o documento oficial do *Facebook* afirme que todas as informações são agregadas e anonimizadas, preservando desta forma a privacidade dos usuários, a empresa está monetizando informações extremamente íntimas de seus usuários utilizando *machine learning*. Para Biddle, o *Facebook*:

Tem muito mais a ver com empresas como Equifax e Experian do que qualquer outra empresa que se diz voltada para o cliente. O *Facebook* é basicamente um atacadista de dados (BIDDLE, 2018, *online*).

No documento são mencionadas algumas informações utilizadas pelo algoritmo *FB Learner Flow*, como localização, informações do dispositivo usado para acessar o site, redes de telefone, *Wi-Fi*, visualizações de vídeos, afinidades e certos detalhes sobre amizades, como as similaridades entre o usuário e seus contatos. Essa prática é denominada pela empresa como “a expertise do *Facebook* em *machine learning*” (BIDDLE, 2018, *online*). De acordo com Biddle:

Especialistas consultados pelo *The Intercept* afirmam que os sistemas descritos no documento levantam uma série de questões

éticas, como o uso dessa tecnologia para manipular usuários, influenciar eleições ou favorecer empresas (BIDDLE, 2018, *online*).

Ainda segundo Biddle (2018), o *FB Learner Flow* foi anunciado como parte das ferramentas internas, cujo objetivo é ajudar o *Facebook* a se ajustar às preferências do usuário. Ao fazer uma pesquisa sobre o *FB Learner Flow* foi encontrada a seguinte publicação (do próprio *Facebook*): “*Applied Machine Learning at Facebook: A Datacenter Infrastructure Perspective*”, escrita por Hazelwood et al (2018). Nessa publicação é descrita a infraestrutura de *hardware* e *softwares* que suporta o aprendizado de máquina do *Facebook* e entre a lista de *softwares* consta o *FB Learner Flow*.

De acordo com Hazelwood et al (2018), no *Facebook*, o aprendizado de máquina fornece recursos essenciais para impulsionar quase todos os aspectos da experiência do usuário. O *Facebook* utiliza uma ampla variedade de algoritmos de *machine learning* nos seus serviços. Por exemplo, a classificação do *feed* de notícias utiliza o algoritmo “*Multi Layer Perceptron – MLP*”. A finalidade é fazer com que as pessoas visualizem as histórias mais importantes para elas. Os modelos dos *feeds* de notícias são treinados para identificar vários fatores do ambiente e dos usuários e com base neles definir a ordem e classificação do conteúdo. O serviço de anúncios também se utiliza do algoritmo “*Multi Layer Perceptron – MLP*” para determinar quais anúncios devem ser exibidos para os usuários. Os modelos de anúncios são treinados para aprender com as características do usuário, as interações anteriores e quais atributos do anúncio podem ser mais preditivos para a probabilidade dos usuários clicarem, visitar um site e/ou comprar um produto. O serviço de reconhecimento facial do *Facebook* utiliza o algoritmo “*Support Vector Machine – SVM*”. Dada uma imagem, primeiro são encontrados todos os rostos nessa imagem. Em seguida, o algoritmo executa um reconhecimento facial específico dos usuários oferecendo uma sugestão para marcação dos amigos nas fotos. A tradução de Idiomas é o serviço que gerencia a internacionalização de conteúdo do *Facebook*. Este serviço suporta traduções para mais de 45 idiomas e utiliza o algoritmo “*Recurrent Neural Networks – RNN*” (HAZELWOOD et al, 2018, *online*).

Uma análise desses serviços mostra o quanto os algoritmos de *machine learning* (aprendizado de máquina) influenciam na escolha das notícias, postagens e anúncios que são exibidos na *timeline* do *Facebook*. O usuário tem a ilusão de que

escolhe o que lê, visualiza, curte, comenta e compartilha, mas isso é uma falsa sensação de liberdade. Na verdade, quem classifica, exclui e decide o que aparece na *timeline* é um algoritmo de aprendizado de máquina e é com base nessa classificação que as interações dos usuários do *Facebook* são realizadas.

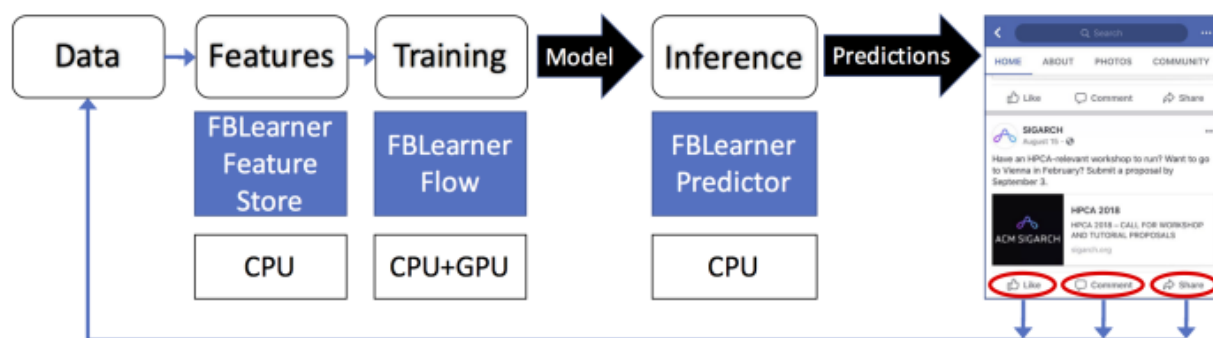
Como alertou Lazzarato (2006), somos submetidos à força e ao papel estratégico desempenhado pelas “máquinas de expressão” (2006, p. 81) e essas máquinas tornam as opções que são apresentadas “fechadas e totalitárias”. Elas destroem nossa capacidade para “escolher outros mundos possíveis ou que poderiam vir a existir” (2006, p. 105). Nossa liberdade se resume a escolher entre opções previamente formatadas. De acordo com Domingos (2017), quando os algoritmos de aprendizado de máquina “se tornam o intermediário, o poder se concentra neles”. Segundo esse autor, a última etapa nesse processo de escolha é sempre nossa, contudo as opções apresentadas são “99,9%” selecionadas por algoritmos de aprendizado de máquina (Domingos, 2017, p. 35).

De acordo com Hazelwood et al (2018), o *FBLearner Flow* faz parte do kit de ferramentas que visam simplificar a tarefa de alavancar o aprendizado de máquina nos produtos do *Facebook*. O *FBLearner* consiste num conjunto de três ferramentas, e cada uma concentra-se em diferentes partes do aprendizado de máquina. A maior parte do treinamento de *machine learning* no *Facebook* é executada na plataforma *FBLearner*.

A seguir, uma descrição dessas ferramentas: 1) o *FBLearner Feature Store* é um catálogo de vários geradores de recursos (dados) que pode ser usado para treinamento e previsão em tempo real. Ter essa lista de recursos é o ponto de partida para que as equipes comecem a usar o *machine learning*, além de ajudar a melhorar os modelos existentes com novos recursos. Os dados são obtidos por meio de interações (curtidas, comentários, compartilhamentos, fotos) que os usuários disponibilizam no *Facebook*; 2) o *FBLearner Flow* funciona como um sistema de gerenciamento que executa um fluxo de trabalho descrevendo as etapas para treinar e/ou avaliar um modelo e os recursos (dados) necessários para o seu uso. Possui ferramentas para o gerenciamento de experimentos e uma interface com o usuário que controla todos os artefatos e métricas gerados a cada execução ou experimento de fluxo de trabalho e 3) o *FBLearner Predictor* é o mecanismo de inferência interna que usa os modelos treinados no *FBLearner Flow* para fornecer previsões em tempo real. O *FBLearner Predictor* é usado por várias equipes de produtos no *Facebook*

(Hazelwood et al, 2018, *online*). A Figura 15 mostra as ferramentas de aprendizado de máquina utilizadas pelo *Facebook*.

Figura 15 - Ferramentas de Aprendizado de Máquina do Facebook



Fonte: Hazelwood et al (2018)

Biddle (2018) relata que o *Facebook* foi questionado a respeito do *FBLearner Flow* e sobre quais dados dos usuários são usados para prever comportamentos, e se essa tecnologia poderia ser aplicada a outros contextos mais sensíveis, como campanhas políticas ou assistência médica. Em vez de responder aos questionamentos, a equipe de relações públicas da rede social afirmou apenas que o *FBLearner Flow* é usado para gerenciar vários tipos de fluxos de trabalho, afirmando essa que está alinhada com o discurso da publicação sobre *FBLearner Flow*. Segundo Biddle (2018), o *Facebook* negou que o *FBLearner Flow* seja usado como aplicação para o marketing.

É difícil comprovar se o *Facebook* utiliza o *FBLearner Flow* como aplicação de marketing para realizar previsões futuras dos usuários ou que tenha desenvolvido alguma outra ferramenta com esta finalidade. No entanto, ao analisar a patente “Predicting Life Changes of Members of a Social Networking System²⁷”, registrada em 2010, fica claro o intuito dessa empresa em utilizar aprendizado de máquina e *big data* para prever ou inferir mudanças futuras na vida dos usuários, como alteração do estado civil, novos empregos, nascimentos de filhos, entre outras mudanças significativas.

Na visão de Frank Pasquale, professor de Direito da Universidade de Maryland e pesquisador sobre ética dos algoritmos da Universidade de Yale, esse projeto de

²⁷ <https://patents.google.com/patent/US20120016817A1/en>

previsão comportamental do *Facebook* é “assustador”. Ele se mostra preocupado com a possibilidade de que as previsões do algoritmo sejam transformadas pela empresa em “profecias autorrealizáveis”. “Porque, uma vez feita a previsão, a empresa tem um interesse financeiro na sua realização” (PASQUALE apud BIDDLE, 2018, *online*).

E no campo da política, quais seriam as implicações do aprendizado de máquina para predição do comportamento dos eleitores? De acordo com Jonathan Albright, diretor de Pesquisa do Tow Center for Digital Journalism, da Universidade de Columbia, a segmentação de anúncios via Inteligência Artificial “sempre pode virar uma arma” para influenciar, por exemplo, a política. Ele se preocupa com os possíveis usos dessas técnicas nas eleições (apud Biddle, 2018, *online*). Para Biddle (2018), os estímulos gerados pela Inteligência Artificial já são problemáticos o suficiente quando se trata de incentivar uma compra, imagina utilizar essa tecnologia para conseguir votos.

Para Guareschi (2017), a coleta e análise dos dados disponibilizados na Internet e nas redes sociais tornaram-se importantes recursos preditivos no campo político. O tratamento dessas massas de dados, com apoio de algoritmos de aprendizagem de máquina, garante uma previsão cada vez mais confiável. Diante disso, é possível afirmar que todas essas táticas, estratégias e ferramentas tecnológicas podem ser aplicadas no marketing político? Antes de responder a esta pergunta é importante explicar um pouco como o marketing político funciona e qual sua relação com o marketing comercial.

4 MARKETING POLÍTICO

Para Antoniutti (2015), “marketing é uma das mais antigas estratégias utilizadas no mundo” para levantar interesses, divulgação e venda de produtos (2015, p. 123). Devido à grande concorrência vivenciada pelas empresas no século XX, o marketing foi sistematizado e dessa forma os empresários tiveram que realizar análises para criação de novos mercados e desenvolvimento de novos produtos, visando ampliar as ofertas e não apenas abastecer o mercado com os itens já produzidos.

O que tornou o marketing amplamente estudado e aplicado não só nas empresas, mas em outros segmentos, foi a obra “Prática de Administração de Empresas”, de Peter Drucker. Segundo Antoniutti (2015), as técnicas de marketing de produtos aplicadas na política é uma prática contemporânea que se tornou dominante apenas nas últimas décadas.

Para Kotler (1975), conceitos e práticas utilizados pelo marketing comercial estão inteiramente relacionados com o marketing político, uma vez que a campanha de um candidato requer o mesmo tipo de planejamento, pesquisa e ações que são tomadas nas campanhas de mercado. Gomes reforça esse entendimento com o seguinte argumento:

Na empresa, o marketing é uma filosofia prática, uma forma de direção que se adianta ao futuro; faz sua previsão e consegue o desenvolvimento industrial, cumprindo a função técnica de pesquisar, planejar, executar e controlar empiricamente. Como na atividade empresarial, o marketing político e o eleitoral surgem como consequência da concorrência: saber que o adversário pode nos desbancar e que é necessário que se desbanque o adversário (GOMES, 2000, p. 30).

Contudo, é importante salientar que há algumas diferenças entre o marketing comercial e o marketing político. Tomazelli (1988) justifica essas distinções com o argumento de que um candidato é um produto vivo, que se comunica, toma ações por conta própria e interage com o ambiente, diferente de um produto comercial tradicional que é inanimado, não fala, pensa e muito menos interage. Ele também alerta que a campanha de uma pessoa não pode ser planejada da mesma forma que a campanha de um produto, pois a pessoa possui personalidade, histórico, e é compreendida e avaliada por meio de suas ações e comportamentos.

Gomes (2000) defende a existência do marketing político e do marketing eleitoral. Para ela, o marketing político trata estratégias permanentes, cujas ações devem ser sistemáticas e organizadas para uma longa duração, visando formar futuros eleitores. O marketing eleitoral, por outro lado, tem por objetivo fazer o partido vencer uma eleição, dentro de um período delimitado, em que “o objeto é a relação entre o eleitor e o candidato” (2000, p. 28). Ela resume marketing político e marketing eleitoral da seguinte forma:

O marketing político e o eleitoral são conjuntos de técnicas que trazem os meios de reflexão suficientes para fixar as estratégias mais oportunas, e que permitam ao candidato ou partido conseguir o objetivo político ou eleitoral pré-fixado (GOMES, 2000, p. 30).

De acordo com Manhanelli (1988), a essência do marketing político é destacar as qualidades do candidato e direcionar as estratégias para atingir os eleitores que se identificam com essas qualidades. No entanto, ele enfatiza que “o marketing político não ganha eleições; ajuda a quem tem condições de fazê-lo a chegar lá” (1998, p.70). Para isso, Gomes (2000) pontua que o marketing político e marketing eleitoral precisam ter como pontos de partida, a análise, a pesquisa e o conhecimento do eleitor, e que esses pontos se desdobram nas seguintes etapas: análise, previsão, objetivos, estratégias, táticas e controle.

Para Penteado (2011), o marketing político está se “reconfigurando” e se adequando “às novas possibilidades que o desenvolvimento das tecnologias digitais introduz”, especialmente com a utilização da Internet (2011, p. 10). Todas essas transformações que ocorreram na comunicação digital permitiram a criação de novas estratégias que estão sendo adotadas pelo marketing político contemporâneo.

Para compreender o desenvolvimento do marketing político no Brasil, Penteado (2011) destaca dois importantes elementos:

A formação de uma nova visibilidade política pelos canais da Internet e uma nova estrutura de comunicação que encontra no meio digital um espaço de convergência das novas tecnologias e formas de comunicação interativa (PENTEADO, 2011, p.11).

Na visão de Penteado (2011), a comunicação política vivencia uma nova fase com a introdução das TICs em suas atividades:

Essas inovações oferecem novas possibilidades de interação entre os partidos e candidatos com a população, por meio de canais diretos da rede de comunicação sem a necessidade da intermediação das práticas eleitorais consolidadas (uso dos cabos eleitorais) e dos meios de comunicação tradicionais (PENTEADO, 2011, p.12).

Penteado (2011) identifica três fases do marketing político, sendo que a terceira ele divide em dois momentos. Considerou o Brasil como referência para efetuar esta análise. A primeira fase pré-moderna (1945-84) foi definida como um estágio de pré-criação do marketing político:

Não existia a profissionalização dos participantes das campanhas, as ações eram caracterizadas pelo uso do instinto dos candidatos na condução de sua comunicação com o eleitorado que centralizavam as ações em torno de sua rede de relações pessoais. Nesse primeiro estágio, a comunicação com o eleitorado era intermediada pelos cabos eleitorais, pelo contato corpo a corpo entre eleitor e candidato, pelas ferramentas de comunicação do partido e pelo uso de discursos no rádio, principal meio de comunicação desta época (PENTEADO, 2011, p. 12).

A segunda fase, considerada moderna (1985-2002) possui as seguintes características:

Foi marcada pela profissionalização da atividade do marketing político, com a entrada de profissionais da área de comunicação nos núcleos de campanha, antes dominados por partidários políticos. Nesse período, a comunicação ganha maior destaque dentro das campanhas, principalmente pela consolidação dos meios de comunicação de massa, em especial a televisão. A grande penetração desse meio fez com que o Horário Gratuito de Propaganda Eleitoral (HGPE) se tornasse o centro das ações da campanha. As peças divulgadas nesse horário passaram a ser produzidas por especialistas, muitos deles da área de propaganda comercial, encarecendo as campanhas. A consolidação da atividade do marketing profissional gerou inúmeras críticas quanto à alienação do eleitorado e a desqualificação do debate político (PENTEADO, 2011, p. 12).

Na terceira fase, denominada como pós-moderna passou-se a considerar o espaço digital como um canal opcional de comunicação para divulgação das mensagens da campanha eleitoral:

As primeiras experiências são marcadas pela criação dos *websites* oficiais das campanhas dos candidatos, onde os usuários podiam encontrar material de campanha. Com o desenvolvimento da atividade, sob influência do modelo norte-americano de campanha, cria-se um mercado de informações sobre o eleitorado que permite a elaboração de mensagens personalizadas distribuídas via *e-mail* (PENTEADO, 2011, p. 13-14).

Penteado (2011) destaca um desdobramento desse terceiro estágio com a *Web 2.0*, que apresenta outras formas de utilização dessas tecnologias nas estratégias das campanhas eleitorais e impõe desafios para utilização mais eficiente desses novos dispositivos. Penteado (2011) chama a atenção para as características

interativas e colaborativas da *Web 2.0* que permitem “ações descentralizadas de militantes e simpatizantes” que utilizam a Internet e as redes sociais “para mobilizar os eleitores e, principalmente, atacar os adversários” (2011, p. 14).

Considerando o estágio atual que estamos vivendo, vale acrescentar mais um desdobramento que é a utilização do *big data*, aprendizado de máquina e *microtargeting* comportamental nas campanhas eleitorais. Esse estágio ainda está em desenvolvimento e são poucos os estudos que abordam esse tema. Para Penteadó (2011), essas novas práticas não substituirão técnicas antigas, pelo contrário, elas tendem a ser utilizadas em conjunto com as tradicionais. Nesse contexto de constantes transformações, o marketing político vem se configurando como uma prática dentro das campanhas eleitorais, despertando diversas questões que precisam ser analisadas e compreendidas.

Tabela 4 - Fases do Marketing Político no Brasil

Nome	Características	Tipo de comunicação	Meio de comunicação
Fase 1. Pré-moderna	Marketing político intuitivo, ligação próxima entre candidato e eleitor; ênfase na mobilização de grupos sociais.	Corpo a corpo (pessoal)	Jornal do partido, panfletos e rádio
Fase 2. Moderna	Marketing político de massa; profissionalização das campanhas; ênfase na persuasão dos eleitores.	Comercial de televisão, <i>spots</i>	Rádio e televisão
Fase 3.1. Pós-moderna	Marketing político segmentado; uso de banco de dados; ênfase em mensagens personalizadas.	<i>Website</i> de campanha	Internet: portais e e-mails
Fase 3.2. Colaborativa	Marketing político colaborativo; descentralização das ações de campanha; ênfase na mobilização virtual pelas redes sociais.	Redes sociais	Internet: redes sociais e <i>blogs</i>
Fase 3.3. Aprendizado de máquina e <i>big data</i>	Identificação de padrões; monitoramento das redes sociais; reconhecimento de imagens e <i>microtargeting</i> comportamental.	Redes sociais, <i>WhatsApp</i>	Internet, redes sociais, <i>fakenews</i> e <i>bots</i> (robôs).

Fonte: Penteadó (2011)

4.1 Marketing Eleitoral e Marketing *One to One* (Individualizado)

O marketing eleitoral é um “conjunto de planos e ações” que visa influenciar os eleitores no período da eleição, porém “não é ele quem define uma eleição”. Nesse cenário, as “campanhas têm se mostrado cada vez mais científicas e mais baseadas em resultados de análises de dados” (ANTONIUTTI, 2015, p. 126).

As estratégias de marketing nas disputas eleitorais devem considerar o tempo como um fator determinante. Por esse motivo, os candidatos necessitam de uma plataforma bem elaborada para fornecer os subsídios necessários para criação de imagem e direcionamento do discurso. Nesse sentido, ferramentas tecnológicas podem ser utilizadas a fim de apoiar os profissionais de marketing na coleta de informações para elaboração de uma plataforma consistente, gestão de imagem do candidato e adequação do discurso (ANTONIUTTI, 2015).

Ao contrário dos meios de comunicação de massa, a Internet possibilita, além do material oficial, uma ampla variedade de informações, o que permite uma redução de custos e a participação de diferentes interlocutores. A Internet também proporciona uma forma alternativa ao Horário Eleitoral Gratuito e às diversas formas de propagandas impressas. Nesse novo espaço os eleitores e os políticos podem se comunicar e trocar informações de forma direta.

Antes de iniciar uma campanha eleitoral, é necessário um estudo aprofundado do eleitorado, um planejamento bem elaborado e um acompanhamento contínuo que permitam análises e previsões que auxiliem tomadas de decisão mais assertivas e adequadas a cada situação. Nesse contexto, o foco no comportamento do eleitor é imprescindível. Por isso, a análise das informações dos eleitores é tão importante para a realização das estratégias de marketing eleitoral. “Quanto mais precisas forem as informações usadas para se montar uma estratégia, menor chance de erro na estratégia de ação proposta” (MANHANELLI, 1988, p.16).

Outra observação importante, levantada por Antoniutti (2015), é que o marketing eleitoral considera cada vez mais outros tipos de marketing para fidelizar os eleitores. Conforme descrito anteriormente, a finalidade do marketing político é manter um relacionamento mais duradouro com os eleitores e cativar novos, afinal “nenhuma disputa eleitoral para o candidato termina no dia da votação” (2015, p. 127). Para conseguir tal propósito, o marketing eleitoral acaba se espelhando nas

estratégias do marketing de relacionamento e marketing *one to one* (individualizado) muito usados pelas empresas (ANTONIUTTI, 2015).

O processo de implantação de um programa de marketing *one to one* pode ser pensado como uma série de quatro passos básicos: identificar os clientes; diferenciar os clientes; interagir com os clientes e personalizar (PEPPERS e ROGERS, 2001, apud ANTONIUTTI, 2015).

Fazendo uma adaptação do marketing *one to one* para o marketing eleitoral tem-se a seguinte estratégia: 1) identificar os eleitores, por meio da coleta de dados e mensuração das ações nas diversas formas de contato e comunicação (*online* ou *offline*); 2) diferenciar os eleitores utilizando ferramentas que permitam realizar segmentações; 3) se aproximar dos eleitores, ou seja, iniciar um diálogo contínuo com os eleitores de maneira a incrementar constantemente o conhecimento sobre cada um deles, especialmente quanto às necessidades, desejos e comportamento; e 4) personalizar o conteúdo adaptando as ideias e projetos de governo de acordo com a individualidade dos eleitores (ANTONIUTTI, 2015).

Assim como ocorre no segmento comercial e empresarial, o marketing individualizado aplicado em campanhas eleitorais traz um diferencial competitivo, pois permite entender o ambiente eleitoral e antecipar-se às ações dos candidatos adversários. Atualmente, as redes sociais influenciam na maneira de se trabalhar com a política, uma vez que as mudanças são constantes. Diante desse ambiente em contínua transformação, a tarefa do marketing no segmento político torna-se complexa e cada vez mais necessária para diminuir os riscos. A informação não só está inserida nesse contexto, como também é uma arma importante a ser considerada no arsenal de uma campanha eleitoral.

4.2 O Papel da Informação e da Pesquisa na Campanha Política

O sucesso de qualquer empreendimento está baseado na qualidade e utilidade das informações utilizadas. A qualidade da informação “pode determinar o vencedor numa guerra ou o fracasso de uma campanha” (KUNTZ e LUYTEN, 1982, p. 43). Os autores explicam que a informação só pode ser considerada importante quando possibilita determinar uma ação em benefício de um indivíduo ou quando ajuda a conquistar uma posição privilegiada. Em relação ao candidato eles afirmam o seguinte:

É importantíssimo ao candidato selecionar, entre as milhares que lhe chegam às mãos sobre numerosos temas, quais as que lhe interessam, para assim determinar estratégias que serão usadas na sua campanha (KUNTZ e LUYTEN, 1982, p. 44).

De acordo com Kuntz e Luyten (1982), após decidir o que é interessante ao candidato, o passo seguinte é levantar quais os meios disponíveis para coletar essas informações que, segundo eles, “se encontram diluídas, em sua forma original no universo social” (1982, p. 46). Os autores definem a pesquisa como o meio para obtenção dessas informações. Considerar a pesquisa como parte fundamental do núcleo de estratégias de uma campanha eleitoral é uma decisão importante. Por exemplo, atualmente, nenhuma ação de comunicação é tomada sem que antes se verifique os índices de aceitação e rejeição dos candidatos (FERRAZ, 2010 apud ANTONIUTTI, 2015).

Chalegre, Almeida e Passos (2006, p. 3) afirmam que se os candidatos e partidos desejam se manter competitivos no “mercado político”, principalmente em um ambiente complexo como o de hoje, precisam dedicar tempo e esforço para obter um conhecimento mais aprofundado dos componentes envolvidos no processo eleitoral. Nesse contexto, é necessário conhecer os eleitores, ou seja, saber quem eles são, o que pensam e como tomam suas decisões, antes de querer influenciá-los. Somente de posse desse conhecimento prévio é que será possível se posicionar melhor e identificar os eleitores que precisam ser conquistados. “A decisão dos segmentos a serem conquistados pode ser tomada com mais segurança se estiver respaldada por pesquisas sérias e abrangentes do comportamento dos eleitores” (CHALEGRE, ALMEIDA e PASSOS, 2006, p. 3).

Therriault (2016) explica que para alcançar os eleitores, as campanhas desenvolvem estratégias de mensagens baseadas em pesquisas de opinião e em outros tipos de pesquisas. Na maioria das vezes essas mensagens são testadas e refinadas por meio de testes experimentais, e as selecionadas são transformadas em anúncios oficiais. Esses anúncios, denominados de mídia paga, são personalizados para uso em TV, rádio, impressos, *e-mail* e *online*.

É vital que qualquer estratégia de marketing político seja embasada por pesquisas qualitativas e quantitativas. E é com base nelas que se definem a postura do candidato, os temas da campanha, as propostas a serem apresentadas aos eleitores, a forma de apresentá-las e os tipos de segmentos do eleitorado que devem ser priorizados (FIGUEIREDO, 2000 apud ANTONIUTTI, 2015).

Chalegre, Almeida e Passos (2006) explicam que o tempo entre o lançamento de uma campanha e a data de uma eleição é curto e isso restringe a ação das pesquisas. Mesmo que as pesquisas fossem realizadas antes do início das campanhas, ainda haveria imprevistos por conta do dinamismo do ambiente político, em que um acontecimento qualquer pode alterar a decisão de voto. Isso ocorre devido à pesquisa ser estática e efêmera. Ou seja, quando uma pesquisa é efetuada, ela retrata situações ou informações que ocorreram no exato momento em que foi feita. Como o ambiente em uma disputa eleitoral é passível de transformações em curto espaço de tempo, uma pesquisa estática não consegue retratar todas essas mudanças (KUNTZ, 1982).

Antoniutti (2015) ressalta que os dados de uma pesquisa quantitativa devem ser analisados em conjunto com o quadro político atual e com as pesquisas qualitativas para entender a motivação. Principalmente agora com o “uso cada vez maior das redes sociais pelos cidadãos, é importante estar atento a tudo o que acontece nesses ambientes” (2015, p. 136).

4.2.1 A Internet e as Redes Sociais como Estratégias de Pesquisa Eleitoral

Antoniutti (2015) explica que antes da Internet e das redes sociais, as pesquisas usavam métodos empíricos baseados em amostras obtidas por meio de questionários, entrevistas e vários outros métodos que permitiam entender as características de um determinado universo. No entanto, ela enfatiza que nos dias atuais, qualquer pessoa *online* nas mídias digitais pode ser um formador de opinião e em poucos segundos, atingir uma audiência muito maior que os meios de comunicação tradicionais.

Foi por conta de toda essa influência da Internet e das redes sociais na vida das pessoas que as campanhas políticas passaram a considerar à tecnologia da informação e comunicação (TIC) e toda inovação promovida por ela (ANTONIUTTI, 2015). Para Gaudêncio Torquato, as redes sociais alteram o modo de participação dos brasileiros nas campanhas eleitorais:

Agora existe a opinião pública virtual, que é muito influenciada pelo que circula na internet. Nunca se viu tanta propagação de mensagens de interesse político na internet: se acontece um escândalo, uma votação polêmica em Brasília, imediatamente as pessoas começam a se manifestar nos *blogs* e *Twitter* (TORQUATO, apud FRANÇA, 2009, *online*).

Segundo Antoniutti (2015), os políticos não podem mais desconsiderar o que as pessoas estão falando nas redes sociais, uma vez que esse meio de comunicação se tornou importante para expressão de opiniões, formação da opinião pública e principalmente mobilizações políticas. Para Brandão Júnior (2008), a Internet tem condições de se tornar um “um espaço midiático mais amplo e democrático” para a participação política (2008, p. 17). Antoniutti (2015) considera que se a Internet for usada corretamente, pode ser uma excelente ferramenta política.

Coutinho (2010) também corrobora sobre o potencial da internet. Para ele, *blogs, Facebook, Twitter, Instagram*, enfim, as redes sociais, se transformaram em repositórios de opiniões e comportamentos que se forem bem monitorados podem se transformar em uma importante fonte de informação para as campanhas eleitorais:

É possível que os movimentos de opinião observáveis nas redes sociais antecipem tendências que somente serão captadas posteriormente (e com um custo muito mais elevado) através das pesquisas tradicionais de intenção de voto e temas de campanha. Se e como os partidos estarão atentos a isso é outra história, mas sem dúvida o surgimento de fontes “alternativas” de informação sobre intenção de voto e desempenho de candidaturas vai tornar mais difícil a vida dos institutos (COUTINHO, 2010, p.18).

Não resta dúvida de que se a Internet e, mais precisamente as redes sociais, com toda a informação gerada e distribuída em tempo real, for bem monitorada, pode gerar métricas que permitem aos estrategistas de marketing eleitoral realizar ajustes em tempo real, conforme a interação acontece nessas redes. Antoniutti (2015) cita que as eleições presidenciais de 2014 entraram para a história das campanhas eleitorais aqui no Brasil. Segundo ela, os ataques verbais dos debates entre candidatos e entre eleitores colocaram a política brasileira em evidência nas redes sociais:

No *Facebook*, por exemplo, os brasileiros promoveram o maior número de interações já visto em outras disputas. Segundo pesquisas, foram contabilizados 674 milhões de interações, incluindo postagens, compartilhamentos, *likes* e comentários, entre 6 de julho e 26 de outubro. No *Twitter*, no mesmo período, foram 40 milhões de mensagens. E pela primeira vez foi usado o *WhatsApp* para enviar mensagens privadas por onde se espalharam tantas conversas de eleitores quanto peças de propaganda de candidatos. Ao contrário do *Facebook* e *Twitter*, que divulgam dados sobre a atividade de seus usuários, o *WhatsApp* guarda o anonimato (ANTONIUTTI, 2015, p. 140).

Segundo Ruffini (2016), dificuldades relacionadas aos custos de entrevistas feitas por telefone (fixo e móvel) estão motivando a indústria de pesquisa a buscar alternativas na Internet, análise de sentimento via redes sociais e análise de *big data*. Para Ruffini (2016), a indústria de pesquisas será muito diferente daqui a dez anos, e já é possível ver essa transformação nas eleições americanas de 2016, em que quase metade das pesquisas realizadas foi com base na Internet.

What we can learn about public opinion and human behavior from data is growing daily, and what we think of as “polling” is only a tiny part of it. Voters exhibit behaviors we can learn from even when not in a voting booth or talking to an interviewer. Often, the outward signs of this are digital: typing their questions into the *Google* search box or sharing what they think on *Twitter* and *Facebook*. Social media and search trends data is not truly representative of the general population, sure, but it can help us go places traditional methods can’t (RUFFINI, 2016, p.17)²⁸.

De acordo com Ruffini (2016), esse novo paradigma da pesquisa *online* já mostrou grande sucesso e as vantagens para adoção dessa nova forma de pesquisa são claras: ela custa menos, possui amostras bem maiores e provou que pode ser precisa nos resultados. A pesquisa na Internet pode ser especialmente adequada para entender o que ressoa com diferentes populações, afirma Ruffini (2016).

Para Ruffini (2016), a análise da mídia e a medição da taxa de menção aos candidatos tanto em fontes tradicionais quanto nessas novas mídias, foi uma estratégia muito importante adotada na campanha de Donald Trump. Segundo Ruffini (2016), a indústria de pesquisas no futuro, será composta pelos métodos tradicionais, mas cada vez mais trará novas disciplinas, como a modelagem e estatística, para capturar e dar sentido a montanhas de dados digitais em tempo real.

Antoniutti (2015) também levantou uma tendência de uso do *WhatsApp* “como uma nova força nas estratégias de comunicação direta com os eleitores” (2015, p. 140). Tozetto (2014) reforça essa ideia ao escrever o texto “Eleições 2014 revelam

²⁸ Tradução livre: O que podemos aprender a respeito da opinião pública e o comportamento humano a partir de dados está aumentando diariamente, e o que pensamos sobre pesquisa é apenas uma pequena parte disso. Os eleitores exibem comportamentos com os quais se pode aprender mesmo quando não estão em uma cabine de votação ou conversando com um entrevistador. Muitas vezes, os sinais externos disso são digitais: como digitar suas perguntas na caixa de pesquisa do *Google* ou compartilhar o que eles pensam no *Twitter* e no *Facebook*. Os dados de tendências de pesquisa e mídia social não são realmente representativos da população em geral, mas podem ajudar a descobrir padrões que os métodos tradicionais não conseguem.

novo cabo eleitoral: o *WhatsApp*". Nessa matéria Tozetto (2014) consegue o depoimento da estrategista digital da campanha de Barack Obama em 2012, Laura Olin. Segundo a declaração de Olin: "Os políticos ainda usam pouco o *WhatsApp*. Mas ele tem uma grande base de usuários e, por isso, certamente será um fator a ser explorado nas próximas eleições" (apud TOZETTO, 2014, *online*).

4.3 Campanhas Eleitorais

O termo campanha é derivado da palavra francesa "campo aberto" (POPKIN, 1996, apud ANTONIUTTI, 2015). Não é por acaso que esse termo relaciona política com "a metáfora militar de que as campanhas são batalhas" travadas por meio de um "debate público de opiniões com vistas a conquistar o voto dos eleitores" (ANTONIUTTI, 2015, p.113). Para reforçar esse termo, Manhanelli (1992, p.13), diz que "eleição é guerra" e que no mundo inteiro, as campanhas eleitorais usam cada vez mais conceitos de guerra em suas estratégias. Observa-se isso facilmente nas nomenclaturas utilizadas nas campanhas, como "táticas, batalha de votos, penetração no segmento adversário, domínio geográfico entre outras terminologias" (MANHANELLI, 1992, p.13).

Antes de explicar o que é uma campanha eleitoral, Gomes (2000), define primeiro o que é eleição:

Forma de procedimento, reconhecido pelas regras de uma organização pelo qual todos, ou alguns membros da mesma, escolhem uma pessoa, ou um reduzido número de pessoas, para desempenhar um cargo com autoridade em tal organização (GOMES, 2000, p. 14).

Após definir eleição, ela descreve campanha eleitoral da seguinte forma:

Conjunto de atividades legais, organizadas ou desenvolvidas pelos partidos, coligações ou comitês de eleitores e candidatos, com o objetivo de arrecadar votos para que determinados políticos possam ocupar, por representação, os cargos políticos públicos (GOMES, 2000, p. 14).

Manhanelli (1992), por sua vez, define campanha eleitoral como uma batalha, em que cada candidato usa da "persuasão e cooptação" para obter um número de eleitores capaz de "subjugar" os outros adversários e deixá-los sem resistência (1992, p. 13).

A disputa eleitoral basicamente é dividida em duas grandes fases: 1) avaliação e definição da estratégia que deve ser realizada antes do "combate direto" e 2) o

“confronto direto”, propriamente dito; é nessa fase “tática” que os conceitos usados na guerra são postos em prática. Conceitos como “ataque, defesa, aliança” são tão conhecidos pelos generais quanto pelos estrategistas políticos (GRANDI et al., 1992, p. 117, apud ANTONIUTTI, 2015).

Gomes (2004) considera que uma campanha eleitoral é um “jogo de competição” onde os partidos e candidatos duelam e se comunicam com os eleitores de forma intensa. Para ela “as campanhas eleitorais devem ser lutas de ideias ou de programas, mais do que de imagens e de sensacionalismos baratos” (GOMES, 2000, p.13).

Segundo Therriault (2016), todo o trabalho de uma campanha eleitoral visa unicamente os votos de uma eleição. Para ele, a campanha influencia o resultado de duas formas: primeiro quando influencia quem vota e segundo quando influencia em como votam. Os demais números de uma campanha, como veiculação de anúncios, resultados de pesquisas, são importantes por conta da ligação com os eleitores que são os responsáveis pelo voto.

Em síntese, o objetivo da campanha eleitoral é “fidelizar” o eleitor para conquistar votos e minimizar a possibilidade de desistência. Para que isso aconteça é necessário ter uma estratégia de comunicação bem definida e reforçada por um planejamento detalhado para identificação do “eleitor-alvo” (*target*), além de monitoramento do adversário de forma a prover subsídios para elaboração do discurso do candidato e divulgação de mensagens alinhadas com o desejo dos eleitores (ANTONIUTTI, 2015, p. 114).

Após esse processo, para estabelecer a comunicação com os eleitores é necessário escolher os canais de comunicação a serem utilizados durante a campanha. Segundo Antoniutti (2015), esses canais podem ser diretos ou mediados. Os canais diretos são aqueles que permitem uma comunicação direta entre os candidatos e seus eleitores, como é o caso dos *websites* dos partidos, *Whatsapp*, *Facebook*, *Twitter*, *YouTube*, *e-mail* e o Horário Eleitoral Gratuito. Os canais mediados são aqueles em que as mensagens dos partidos e dos candidatos são transmitidas pelos “órgãos de informação de massa” e mediadas por jornalistas (ANTONIUTTI, 2015, p. 115). Vale destacar que a mídia desempenha um papel importante nas campanhas políticas.

Para muitos estudiosos as campanhas políticas se alteram conforme a evolução da sociedade e as inovações da tecnologia. Todas essas mudanças foram

refletidas nas campanhas eleitorais que passaram a ser descritas por alguns como um processo de americanização (BLUMER e KAVANAGH, 1999, apud ANTONIUTTI, 2015).

A americanização é um termo que faz referência à campanha presidencial norte-americana de 1960, entre Richard Nixon e John Kennedy que utilizaram pela primeira vez a televisão em uma disputa. Desde então, o conceito de americanização é utilizado pelos especialistas para indicar o “domínio absoluto do marketing político nas disputas eleitorais” (SWANSON; MANCINI, 1996, apud ANTONIUTTI, 2015, p. 115). A importação dessas técnicas de marketing político está sendo aplicada e adaptada em vários países, inclusive no Brasil.

De acordo com Antoniutti (2015), Mancini e Swanson propõem utilizar o termo americanização de forma mais restrita para indicar apenas a descrição da exportação de profissionais e técnicas norte-americanos de marketing político. Em relação às transformações mais gerais das campanhas eleitorais dos outros países, eles preferem utilizar o termo modernização (SWANSON; MANCINI, 1996, apud ANTONIUTTI, 2015, p. 116). Já para Norris (2000), as modificações nas estratégias de comunicação política e eleitoral ocorreram por conta da atualização da imprensa, evolução da televisão, da Internet e de outros meios. Manin (1995) usa o conceito de democracia de público para descrever o estado atual das campanhas eleitorais embasadas pelas pesquisas de opinião e as técnicas de marketing e centralizadas nos meios de comunicação.

O processo de modernização das campanhas eleitorais no Brasil teve início no final da década de 1980 com a volta da democracia e se solidificou em 1989 com as primeiras eleições presidenciais. Esse processo de modernização incorporou o uso das pesquisas de opinião, pesquisas quantitativas e qualitativas para definição das estratégias de campanha e adoção da mídia como meio de comunicação entre o candidato e o eleitor.

Outra grande evolução nas campanhas eleitorais originou-se no início da década de 1990 com o surgimento das Tecnologias de Informação e Comunicação (TICs). Para Iasulaitis (2012), o papel das mídias de massa em conectar os cidadãos não foi substituído pelas TICs, mas sim complementado por elas. A Internet e posteriormente as redes sociais são exemplos desse incremento e a cada ano ganham mais espaço nas campanhas eleitorais.

4.4 Mídias de Massa e *Web 2.0* nas Campanhas Eleitorais

O uso da mídia de massa, principalmente o rádio, para divulgação de campanhas políticas foi uma estratégia adotada por Goebbels, ministro da propaganda durante o governo de Hitler. Ele utilizou tão bem essa mídia que vários “historiadores afirmaram que Hitler seria inconcebível sem o rádio” (MANHANELLI, 1988, p. 72). No Brasil, Getúlio Vargas também seguiu essa estratégia e fez do rádio o principal meio para divulgação da sua política.

Antoniutti (2015) cita que em 1952, nos EUA, o general Dwight Eisenhower foi pioneiro ao contratar uma agência de publicidade, a BBDO, para auxiliar na sua campanha presidencial. A função da BBDO era ajustar a linguagem do general ao rádio e televisão, que eram as mídias eletrônicas da época.

Seguindo os passos do general Dwight Eisenhower, John Kennedy, em 1960, foi o primeiro político a usar a TV de forma racional e eficiente, como um “cabo eleitoral”. Mas, quem realmente aperfeiçoou o uso dessa mídia para fins eleitorais, foi Ronald Reagan, um “ex-ator”, na campanha de 1984 (MANHANELLI, 1988, p. 72).

No Brasil, o uso da TV para fins eleitorais começou em 1989, na campanha de Fernando Collor de Mello. Foi nessa época que as “ferramentas de marketing” ganharam destaque (ANTONIUTTI, 2015, p. 125). Segundo Manhanelli (1988), Fernando Collor de Mello utilizou a TV para transmitir “a imagem de homem forte, austero e capaz” (1988, p.72).

Para Manhanelli (1988), os meios de comunicação influenciam profundamente as pessoas, a política, e o “estado psicológico de todo um país” (1988, p. 67). Ele ainda destaca que a utilização dos meios de comunicação em campanhas eleitorais tem como principal objetivo “aumentar a quantidade de votos de um candidato” (1988, p. 68), uma vez que a mídia se tornou um forte canal de comunicação com o eleitorado e a principal responsável pela formação de ideias e comportamento político. Contudo, Manhanelli (1988) alerta que a televisão e o rádio não permitem segmentação dos ouvintes e isso dificulta a implantação da plataforma pelos estrategistas políticos quando o discurso precisa ser segmentado. Esse tema da segmentação será abordado com mais detalhes no item 4.5.

Segundo Miguel (2002), o desenvolvimento da comunicação de massa alterou todo o ambiente político, pois esses novos meios exigiram políticos capacitados para

usá-los. Por isso, ele julga necessário reconhecer o papel central da mídia na política contemporânea. Para ele, os meios de comunicação de massa permitiram uma maior divulgação dos discursos políticos para o grande público. Ele considera a comunicação de massa, nas sociedades contemporâneas, uma ferramenta para difusão de projetos políticos.

Já Norris (2000) concebe os meios de comunicação de massa como os principais mediadores entre o candidato e os eleitores. No período das campanhas eleitorais essa influência se torna bem visível. Para os pesquisadores da área de ciência política, também há um consenso sobre a grande importância da mídia nos regimes políticos, especialmente nos regimes democráticos (ANTONIUTTI, 2004).

Antoniutti (2015) cita, como prova de reconhecimento e consolidação da comunicação de massa no campo político, a criação do Código Eleitoral Brasileiro, em 1965, que instituiu o Horário Eleitoral Gratuito e regulamentou as atividades de marketing político e o acesso dos partidos à propaganda eleitoral. Mas, ela também chama a atenção para a Internet que atualmente se tornou uma forte aliada nas campanhas eleitorais, que passaram a incorporar “os dispositivos da *web* em suas estratégias de comunicação”, aumentando a atuação da equipe de comunicação e marketing político que passou a ser descrito como “marketing digital” ou “marketing 2.0” (ANTONIUTTI, 2015, p. 121).

Antoniutti (2015) explica que a estrutura em rede descentralizada da Internet permite uma troca de informação interativa e colaborativa entre seus usuários. Diferente dos meios de comunicação de massa tradicionais, cujo fluxo de informação é unidirecional, a *Web 2.0* permite criar mecanismo de comunicação nos quais o receptor de uma mensagem poderá “interagir com o conteúdo que vai circular na rede, de forma mais segmentada e personalizada” (2015, p. 121) e com base nisso gerar novos conteúdos.

Vale ressaltar que a *Web 2.0* foi uma terminologia criada para descrever a segunda geração da Internet. Segundo Primo (2017), essa segunda geração tem por características o aumento das formas de publicação, compartilhamento e organização das postagens, além de promover uma maior interação e colaboração entre os usuários. Para ele, a “*web 2.0* refere-se não apenas a uma combinação de técnicas informáticas, mas também a um determinado período tecnológico, a um conjunto de novas estratégias mercadológicas e a processos de comunicação mediada por computador” (PRIMO, 2007, p.2).

Após a eleição de Barack Obama nos Estados Unidos, em 2008, todos os candidatos a cargos políticos passaram a fazer uso da Internet em suas campanhas eleitorais. No Brasil, não foi diferente. Vários candidatos, principalmente os que concorriam à Presidência da República, passaram a utilizar a Internet e as redes sociais, almejando uma maior visibilidade para suas campanhas e mais proximidade e interação junto aos eleitores.

Para Antoniutti (2015), embora as mídias sociais tenham se tornado “um campo fértil para a divulgação de ações políticas” (2015, p. 123), também passou a chamar a atenção para a possibilidade de se coletar dados e informações dos eleitores, medida que poder ser extremamente útil na definição de estratégias de comunicação e condução das campanhas eleitorais. Lundry (2016) pondera que, desde as eleições de 2012, a otimização de mídia orientada por dados tornou-se cada vez mais comum.

4.5 Segmentação

Para influenciar os eleitores, saber quais conquistar e se posicionar melhor perante esse público-alvo é necessário ter um conhecimento prévio do comportamento desses votantes. Dessa forma, é necessário o emprego de métodos mais sistemáticos e científicos para conhecer o mercado eleitoral e um desses métodos é a segmentação.

A prática da segmentação no contexto eleitoral é tão antiga quanto às eleições. A segmentação eleitoral é proveniente do marketing comercial. O conceito de segmentação evoluiu de um conceito econômico até se tornar, atualmente, uma estratégia de conquista de mercado (ANTONIUTTI, 2015).

A segmentação de eleitores segue as mesmas premissas adotadas no marketing comercial. Segundo os especialistas, a segmentação do eleitorado, além de ser uma estratégia necessária, torna a campanha mais barata. A importância da segmentação é mais evidente em cidades de médio e grande porte, como São Paulo e Rio de Janeiro, uma vez que determinados bairros são maiores que muitas cidades brasileiras (ANTONIUTTI, 2015).

De acordo com Castleman (2016), por meio da segmentação, o eleitor pode ser dividido em três grupos: 1) os eleitores que certamente votarão no candidato. Este segmento requer menos custos, ou seja, tempo e recursos. São os eleitores que já conhecem o candidato e não o rejeitam. Nesse caso, a chance do voto é bem maior;

2) os eleitores que estão determinados a não votar no candidato. O tempo e o custo para persuadir esse eleitor são inviáveis, levando em conta a pequena probabilidade de sucesso. Eleitores que rejeitam determinada candidatura dificilmente mudarão de opinião. Nesse caso, é melhor a campanha não ser direcionada a esses eleitores; e 3) os eleitores que podem vir a votar no candidato. É justamente nesse grupo que deve se concentrar o maior esforço da campanha. O *target*, ou eleitor alvo, é aquele que quantitativamente pode eleger o candidato, já que o primeiro segmento pode não ser muito expressivo. É nesse público que a campanha deve direcionar suas atenções e investimentos.

Segundo Nickerson e Rogers (2014), para executar uma análise de custo-benefício, as campanhas necessitam de previsões precisas sobre as preferências dos eleitores, seus comportamentos esperados e as suas respostas em relação à campanha. Por isso as campanhas políticas devem começar a repensar suas estratégias em relação às pesquisas de opinião, principalmente olhando as mídias sociais como potenciais espaços de coleta de dados e informações de eleitores.

No cerne dessa discussão está a microssegmentação do eleitorado, que coloca o eleitor de forma individualizada como centro das ações de marketing e comunicação numa campanha eleitoral. Diante da disponibilidade de grandes volumes de dados sobre os indivíduos, proporcionada pelas análises de *big data* e aprendizado de máquina, as campanhas passam a operar com estatística, lógica e matemática, buscando assim um custo/benefício nas suas ações de comunicação de forma muito mais assertiva.

4.6 Microssegmentação

É necessário segmentar ainda mais (*microtargeting*) para ter um melhor entendimento do eleitorado a ser alcançado em uma campanha. Nesse sentido, a microssegmentação se configura numa técnica mais indicada para dar esse nível de detalhamento. Por isso, atualmente os dados se tornaram essenciais para as eleições (THERRIault, 2016). Para Nickerson e Rogers (2014), as campanhas contemporâneas precisam de dados para construir modelos preditivos, que têm como objetivo fazer a segmentação da campanha com comunicações mais eficientes, a fim de apoiar estratégias mais amplas.

A modelagem e a microssegmentação são extensões de técnicas antigas de segmentação. Embora essas técnicas tenham atingido o ápice na campanha de

Obama em 2012, já haviam sido usadas desde 2004 pelos democratas. Mas, até que ponto vai a influência da microsegmentação? O uso de microsegmentação e de modelos preditivos oferecem grandes vantagens às campanhas eleitorais. Contudo, não podem fornecer aos políticos impopulares uma receita “mágica” para conquistar eleitores. O que essas técnicas podem e devem fazer é melhorar a eficiência de como as campanhas executam seus programas. A modelagem preditiva permite aplicar uma metodologia mais robusta para identificar alvos e proporciona ganhos de eficiência quando aplicados em caráter individual, permitindo usar recursos limitados da maneira mais sábia (CASTLEMAN, 2016, p. 2).

Além dos três tipos de segmentação (eleitores que votarão no candidato; eleitores que não votarão e eleitores que podem votar) descritos anteriormente, Castleman (2016) cita que há outros tipos que podem ser valiosos para as campanhas eleitorais, conforme pode ser observado na Tabela 5.

Tabela 5 - Tipos de segmentação

Tipo de Segmentação	Descrição
Partido político	Probabilidade de se identificar com um partido político ou outro.
Participação eleitoral	Probabilidade de votar em determinada eleição.
Suporte ao candidato	Probabilidade de apoiar um candidato específico.
Posição sobre uma questão específica	Probabilidade de manter uma determinada posição sobre uma questão específica (igualdade no casamento, controle de armas, etc).
Propensão a ser voluntário e/ou doador	Probabilidade de ser ativista de campanha ou doador (no caso dos Estados Unidos).
Demográfico e comportamental	Probabilidade de ser um graduado da faculdade, dono de arma, cliente da <i>Netflix</i> , etc.
Persuasão	Probabilidade de ser sensível à persuasão de campanha.

Fonte: (Castleman, 2016)

Castleman (2016) explica que a maioria dos modelos é construída com base nos dados de pesquisa. Um modelo de suporte ao candidato é gerado por meio de uma pesquisa pequena e uma amostra grande (entre 2.500 a 20.000 ou mais respostas), mas isso varia conforme a geografia e o tipo de modelo. Para casos mais complexos, como a persuasão do eleitor, os modelos podem usar várias perguntas ou estimar os efeitos do tratamento a partir de testes experimentais randomizados. O resultado final da maioria dos modelos é uma pontuação (*score*)

que pode ser usada pela campanha para segmentação. A maioria das pontuações representa probabilidades estimadas, ou seja, ajuda a identificar eleitores persuasivos, mas não pode tornar eleitores persuasíveis ou tornar uma mensagem ruim mais eficaz.

Para Castleman (2016), a etapa mais importante na modelagem é o processo de seleção dos dados. Para ele, os melhores algoritmos de aprendizado de máquina são aqueles que se adaptam ao grande volume e tipos de dados, conseguindo descobrir padrões não lineares. Castleman (2016) cita que os modelos mais comuns são os baseados em árvores de decisão (*decision trees*), particularmente aqueles que combinam árvores de decisão com o modelo floresta aleatória (*random forest*). Mas, cada campanha possui necessidades diferentes, e algoritmos mais simples, como regressão linear (*linear regression*) e regressão logística (*logistic regression*), também são usados. Ainda segundo Castleman (2016), assim como grande parte da comunidade de ciência de dados, o campo de modelagem política está se voltando cada vez mais para ferramentas de código aberto, como R e *Python*, embora alguns profissionais façam uso de pacotes comerciais como o *Stata*.

Segundo Nickerson e Rogers (2014), escores preditivos são usados para atingir todos os aspectos da campanha de divulgação como as ações porta-a-porta, maladireta, telefonemas, *e-mail*, divulgação de anúncio de televisão, mídias sociais (NICKERSON; ROGERS, 2014, p.19). Os autores estabelecem as seguintes categorias de escores preditivos: escores de comportamento, escores de apoio e escores de respostas que serão descritos a seguir.

Os escores de comportamento utilizam o comportamento passado do eleitor e informações demográficas para calcular a probabilidade dos eleitores irem às urnas ou fazer um trabalho voluntário. Já os escores de apoio visam prever as preferências políticas dos eleitores. Como não é possível entrar em contato com todos os eleitores, por conta das restrições de orçamento, as campanhas procuram levantar dados de um subconjunto dos eleitores e utilizar esses dados para elaborar modelos preditivos com o objetivo de tentar antecipar as preferências do restante do eleitorado. Os escores de respostas são aqueles que apresentam previsões de como os eleitores irão responder à campanha de divulgação. Porém, conseguir prever quais indivíduos são mais ou menos sensíveis a determinadas comunicações em um determinado contexto eleitoral é complicado (NICKERSON e ROGERS, 2014).

Um dos grandes desafios das campanhas eleitorais no Brasil é conseguir segmentar o eleitorado. Ao contrário dos Estados Unidos, no Brasil a legislação é um pouco mais rígida em relação ao uso de dados, apesar dos avanços na qualidade do trabalho do IBGE (Instituto Brasileiro de Geografia e Estatística), os dados gerados por este órgão não são suficientes para conseguir realizar uma microssegmentação do eleitor (MOURA, 2011, apud ANTONIUTTI, 2015).

Para Grassia (2016), o nível de granularidade e *insights* que se pode produzir usando os dados disponíveis é incomparável, e os modelos voltados para o campo político aumentam cada vez mais a compreensão sobre quem são os eleitores, como alcançá-los e como se comunicar com eles. Conforme a tecnologia para tratamento de dados e análise se torna mais sofisticada e fácil de usar, é otimizada a tomada de decisão e a capacidade para extrair percepções mais assertivas.

Contudo, Grassia (2016) aponta que o tratamento de dados requer conhecimento técnico e que ainda há falta de cientistas de dados qualificados trabalhando na política. Outro grande desafio sinalizado é que há fontes e estruturas de dados diferentes que “não falam” entre si. Embora esse não seja um desafio exclusivo das campanhas políticas, a velocidade e a eficiência na unificação de dados estão atreladas a um prazo curto (dia da eleição) que não pode ser alterado.

Lundry (2016) acrescenta que essa complexidade ressalta a necessidade de analistas no espaço político que possuam facilidade para processar, limpar e mesclar rapidamente dados. Isso implica ter familiaridade com soluções de armazenamento baseadas em nuvem, como *Amazon Web Services*, plataformas de processamento de dados distribuídas como *Spark* e linguagens de programação como *R* e *Python*. Ao mesmo tempo, esses analistas também precisam entender o processo de compra de mídia e como este se encaixa em uma operação de campanha mais ampla.

Para os especialistas, a técnica do *microtargeting* tem mostrado vantagem competitiva para os candidatos durante uma campanha eleitoral, principalmente naquelas bem acirradas. Essa técnica permite cruzar dados eleitorais com dados de consumo, informações públicas e mídias sociais. É uma ferramenta que permite responder às questões fundamentais durante uma campanha eleitoral como: Quem apoia meu candidato? Onde posso encontrá-lo? Como posso convencer os outros para apoiar o meu candidato? (ANTONIUTTI, 2015).

Quando bem feito, o *microtargeting* responde a cada uma dessas questões e faz isso de uma forma que a segmentação tradicional não consegue realizar. Para alguns especialistas, o *microtargeting* nada mais é do que a antiga campanha feita porta-a-porta, porém praticada de maneira muito mais ampla, dinâmica e automatizada por meio de tecnologias como *big data* e aprendizado de máquina. O que permite uma comunicação direta e totalmente personalizada. O *microtargeting* ficou mundialmente conhecido, quando em 2012, na campanha de Barack Obama desenvolveu uma estrutura tecnológica para saber exatamente quem procurar e o que dizer (ANTONIUTTI, 2015). Os detalhes dessa campanha, assim como a campanha de Donald Trump são tratados no capítulo 5.

5 APRENDIZADO DE MÁQUINA NAS CAMPANHAS POLÍTICAS

Pode-se dizer que três importantes eventos políticos ocorridos nos Estados Unidos foram marcados pela intensa utilização das TICs. O primeiro foi referente à vitoriosa campanha de Barack Obama para a presidência dos Estados Unidos em 2008. O segundo corresponde à reeleição de Barack Obama em 2012. Já o terceiro, foi a polêmica vitória de Donald Trump que utilizou os serviços de *big data*, *machine learning* e psicometria prestados pela empresa *Cambridge Analytica*.

5.1 Campanha Eleitoral de Barack Obama

De acordo com Antoniutti (2015), o modelo de campanha que tornou Obama vitorioso foi estudado por vários especialistas, entre eles Umair Haque, do Harvard Media Lab, que em artigo publicado pelo site Harvard Business Publishing, afirmou que: “A estrutura de campanha Obama está para as campanhas tradicionais como a empresa *Google* está para uma empresa convencional” (HAQUE apud ANTONIUTTI, 2015, p. 152).

Segundo Scarvalone (2016), as campanhas presidenciais de Barack Obama revolucionaram o modo como a tecnologia e os dados poderiam ser usados para identificar e falar com os interesses dos eleitores, refletindo a aplicação cada vez mais sofisticada do marketing digital no campo da política eleitoral. Para Scarvalone (2016), como os eleitores passam cada vez mais tempo *online*, a publicidade digital combina as capacidades de segmentação, persuasão e mensuração que movimentam a maioria dos votos com o menor custo. Campanhas modernas não podem mais vencer usando apenas anúncios de TV (SCARVALONE, 2016).

Andrew Therriault foi diretor de Ciência de Dados do Comitê Nacional Democrata de 2014 a 2016, liderando uma equipe que desenvolveu modelos de segmentação de eleitores e outras ferramentas analíticas usadas nas campanhas democratas. Para Therriault (2016), a cada eleição nos EUA as campanhas se tornam mais dependentes de tecnologias, dados e análises que são utilizados na estratégia geral, envio de mensagens personalizadas e na publicidade. Para ele, toda essa evolução foi bem documentada nas campanhas presidenciais de 2008 e 2012, e essas ferramentas tecnológicas se espalharam em 2016 para outras campanhas políticas, além da presidencial. Therriault (2016) explica que para tratar a imensidade de números, em alguns casos conflitantes, os agregadores de

pesquisa utilizam modelos de previsão com o objetivo de combinar resultados e gerar estimativas mais claras a respeito da campanha eleitoral.

De acordo com Domingos (2017), “foi o *machine learning* que elegeu” Barack Obama em 2012. Segundo ele, a campanha de Mitt Romney seguiu uma “abordagem convencional de consulta”, compilando os eleitores em categorias genéricas. Já Obama contratou Rayid Ghani, especialista em *machine learning*, para ser cientista-chefe de sua campanha (DOMINGOS, 2017, p. 40-41).

Segundo Domingos (2017) Rayid Ghani empreendeu uma das “maiores operações de análise da história política” (2017, p. 41). Ele reuniu todas as informações sobre eleitores em um único banco de dados; cruzou estes dados com o que conseguiu minerar em redes sociais, marketing e outras fontes; e com base na análise de todas essas informações conseguiu prever estes quatro itens para cada eleitor: “qual a probabilidade de apoiar Obama; comparecer às pesquisas; reagir aos lembretes da campanha e mudar de opinião a partir de uma troca de ideias sobre um assunto específico” (2017, p. 41). Com base nesse modelo, a campanha realizou 66.000 simulações da eleição e, de acordo com os resultados, direcionou as equipes com “as informações de quem deveriam chamar, em quais portas deveriam bater e o que dizer” (Domingos, 2017, p. 41).

Na política, como nos negócios e na guerra, não há nada pior que ver seu oponente realizar movimentos que você não entende e sobre os quais não sabe o que fazer até ser tarde demais. Foi isso que aconteceu na campanha de Romney. Eles podiam ver o adversário comprando anúncios em determinadas emissoras a cabo de cidades específicas, mas não sabiam o porquê; sua bola de cristal estava muito embaçada. No fim das contas, Obama ganhou a preferência de todos os estados decisivos, exceto Carolina do Norte, com margens maiores que o previsto até pelos mais confiáveis peritos em opinião pública (Domingos, 2017, p. 41).

5.2 Campanha Eleitoral de Donald Trump

Segundo Therriault (2016), o ciclo eleitoral de 2016 será lembrado como o ano em que os dados políticos atingiram a maturidade. Durante anos, grande parte da “velha guarda” (2016, p.45) dos estrategistas políticos resistiu à influência dos dados e análises, preferindo manter uma fórmula mais tradicional. Esse conflito persistiu mesmo quando as corridas presidenciais de 2008 e 2012 mostraram as vantagens que os dados podem proporcionar, particularmente quando uma campanha tem uma vantagem tecnológica distinta.

Para Therriault (2016), no entanto, está claro que essa luta acabou, e o lado dos dados venceu. Quando se trata de tecnologia, é provável que a ciência de dados políticos siga uma trajetória semelhante àquela que se desenrola no campo mais amplo da ciência de dados, pois o campo político é particularmente adequado para a rápida adoção de novas tecnologias, uma vez que o ciclo eleitoral permite que muitas organizações reformem sua tecnologia a cada 2 a 4 anos. Embora o *hardware* físico ainda seja usado por muitas organizações, a adoção de plataformas na nuvem, como *Amazon Web Services* e *Microsoft Azure*, rapidamente se tornou uma realidade. Em relação ao *software*, a maioria das organizações está utilizando ferramentas de código aberto como *Python* e R. E, à medida que a escala dos dados aumenta, bancos de dados relacionais como *MySQL* e *Microsoft SQL Server* estão sendo substituídos por plataformas distribuídas como o *Hadoop*.

Para Rosenberg, Confessore e Cadwalladr (2018), nas eleições americanas de 2016, o uso de *machine learning* e *big data* já não eram o que havia de mais moderno e eficiente. Steve Bannon, conselheiro da campanha de Donald Trump e membro do conselho da *Cambridge Analytica*, apresentou um novo arsenal tecnológico e estratégico muito mais potente. Para tal empreitada, a *Cambridge Analytica* recebeu um investimento de US\$ 15 milhões de Robert Mercer, um bilionário cientista da computação americano, com a promessa de criar ferramentas que pudessem identificar as personalidades dos eleitores americanos e influenciar seu comportamento. Contudo, eles ainda não tinham os dados necessários para tão grande missão (ROSENBERG, CONFESSORE e CADWALLADR, 2018).

De acordo com Rosenberg, Confessore e Cadwalladr (2018), essa história começou quando Alexander Nix, líder da divisão de eleições do *SCL Group*, contratou o consultor de dados Christopher Wylie com o objetivo de entrar no mundo de dados políticos. Wylie foi ligado aos veteranos das campanhas de Obama e estava interessado em usar traços psicológicos para influenciar o comportamento dos eleitores. O grupo fez experiências no Caribe e na África, onde as regras de privacidade eram negligentes ou inexistentes, e os políticos que empregavam o SCL forneciam dados de propriedade do governo. Nix entrou em contato com Steve Bannon, responsável pelo site de extrema direita, *Breitbart News*, e posteriormente com Robert Mercer.

Rosenberg, Confessore e Cadwalladr (2018) destacam que Mercer já havia financiado um projeto piloto de US\$ 1,5 milhão para entrevistar eleitores e testar

mensagens psicográficas na campanha para governador da Virgínia, em novembro de 2013. Nesta eleição o procurador-geral republicano Ken Cuccinelli concorreu contra o democrata Terry McAuliffe. Embora o Cuccinelli não tenha vencido, Mercer não desistiu de seus planos e novos negócios nesse sentido estavam despontando. No início de 2014, o investidor Toby Neugebauer e outros conservadores estavam se preparando para colocar dezenas de milhões de dólares na campanha do senador Ted Cruz, do Texas, trabalho que Nix estava ansioso para vencer.

Christopher Wylie encontrou a solução para criação da nova ferramenta, no Centro de Psicometria da Universidade de Cambridge. Os cientistas dessa universidade desenvolveram uma técnica para mapear traços de personalidade com base no que as pessoas curtiam no *Facebook*. Para realizar esse estudo, os pesquisadores criaram o aplicativo *MyPersonality*, e pagaram pequenas quantias para os usuários baixarem este aplicativo e realizarem o teste de personalidade. O aplicativo, por sua vez, coletava algumas informações particulares de seus perfis e de seus amigos, atividade que o *Facebook* permitia na época.

Segundo Rosenberg, Confessore e Cadwalladr (2018), o Centro de Psicometria não aceitou a proposta para trabalhar com a *Cambridge Analytica*. Wylie, então encontrou o Dr. Kogan, professor de psicologia e especialista em psicometria de mídias sociais, na Universidade de Cambridge, que construiu seu próprio aplicativo, chamado *This is your Digital Life*, e em junho de 2014 começou a coletar os dados para a *Cambridge Analytica*.

Os usuários receberam um pagamento para baixar o aplicativo e conceder suas informações do *Facebook* em troca do resultado do teste. Na época, foi informado que esses dados seriam utilizados para fins acadêmicos. O que não foi avisado é que, além dos dados dos usuários, o aplicativo também coletaria dados de toda a rede de amigos desses usuários e que essas informações seriam utilizadas para fins políticos.

De acordo com Rosenberg, Confessore e Cadwalladr (2018), o aplicativo forneceu mais de 50 milhões de perfis brutos para a empresa. Desses perfis, cerca de 30 milhões continham informações que a empresa poderia construir perfis psicográficos e apenas cerca de 270.000 usuários haviam consentido a coleta de dados, que foram aqueles que participaram da pesquisa.

Para entender melhor as consequências da polêmica envolvendo a *Cambridge Analytica*, a eleição de Donald Trump e o vazamento de dados *Facebook*, é importante conhecer um pouco sobre os atores envolvidos com a *Cambridge Analytica*.

Alexander Nix, CEO da *Cambridge Analytica*, é formado pela Universidade de Manchester, trabalhou como analista financeiro no México e no Reino Unido. Em 2003 ingressou na *Strategic Communication Laboratories - SCL*, uma empresa de comunicação estratégica. A partir de 2007 assumiu a divisão de eleições dessa empresa (*SCL Elections Ltd*). Declara ter trabalhado em 260 campanhas globais. Montou a *Cambridge Analytica* para trabalhar na América, com investimento de Robert Mercer (CADWALLADR e GRAHAM-HARRISON, 2018).

Steve Bannon é um ex-banqueiro. Era o responsável pelo site de extrema direita, *Breitbart News* quando conheceu Wylie, Nix e aconselhou Robert Mercer a investir em pesquisa na área de dados políticos por meio da criação da *Cambridge Analytica*. Em agosto de 2016, se tornou o CEO da campanha de Trump contribuindo para vitória do mesmo. Por conta disso, Bannon recebeu o cargo de estrategista-chefe do presidente. No entanto, em agosto de 2017 seu relacionamento com Trump não estava indo muito bem e ele foi afastado.

Robert Mercer é um bilionário cientista da computação que usou sua fortuna para se tornar um dos homens mais influentes na política dos Estados Unidos, além de ser um dos principais doadores republicanos. Atualmente é o CEO da *Renaissance Technologies*, mas trabalhou na IBM como especialista em Inteligência Artificial, na área de “reconhecimento de fala e tradução”, como explica Cadwalladr:

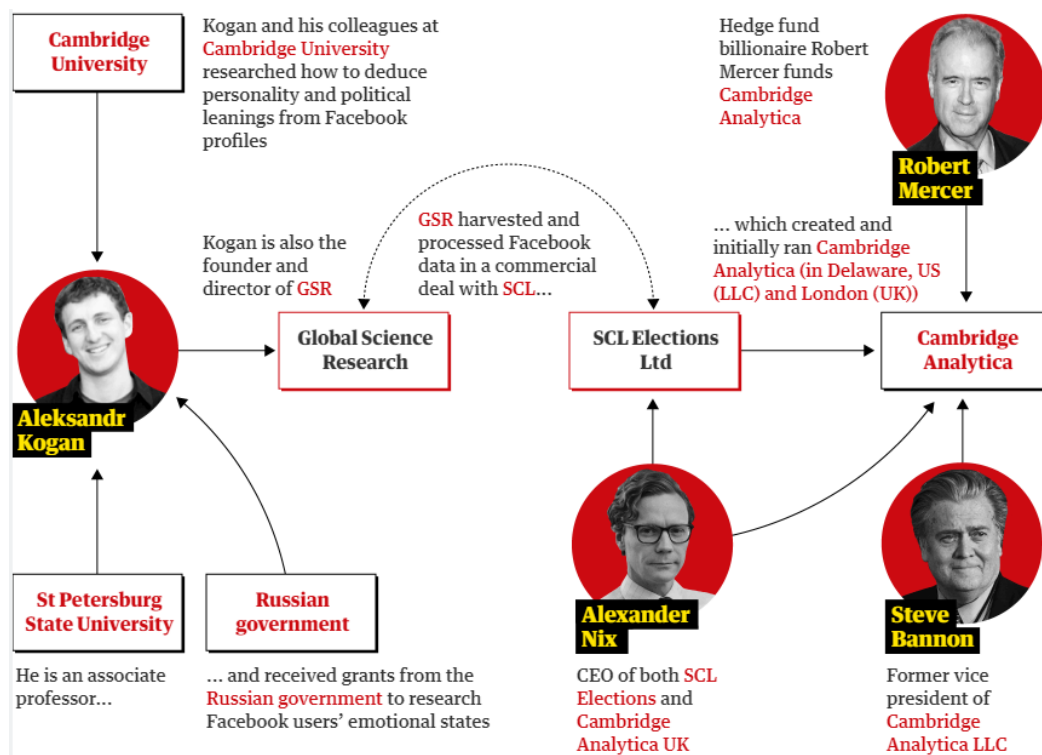
Robert Mercer was a pioneer in AI and machine translation. He helped invent algorithmic trading – which replaced hedge fund managers with computer programs – and he listened to Wylie’s pitch. It was for a new kind of political message-targeting based on an influential and groundbreaking 2014 paper researched at Cambridge’s Psychometrics Centre, called: “Computer-based personality judgments are more accurate than those made by humans” (CADWALLADR, 2018, *online*)²⁹.

Aleksandr Kogan viveu em Moscou, mas depois mudou-se para os Estados Unidos com a família, onde se naturalizou. Estudou em Berkeley e fez doutorado na Universidade de Hong Kong. Depois ingressou na Universidade de Cambridge como professor de psicologia e especialista em psicometria de mídias sociais. Criou a empresa *Global Science Research - GSR* para realizar pesquisas de dados para a *Cambridge Analytica*. Aceitou uma posição na Universidade Estadual de São Petersburgo, e também recebeu subsídios do governo russo para pesquisa (CADWALLADR e GRAHAM-HARRISON, 2018).

Em síntese, a *Global Science Research - GSR*, empresa de análise de dados, de Aleksandr Kogan, em parceria com a *Cambridge Analytica*, que foi financiada por um especialista em Inteligência Artificial, coletou dados dos perfis do *Facebook* de milhões de eleitores americanos para traçar o *microtargeting* comportamental para prever e influenciar os eleitores (CADWALLADR e GRAHAM-HARRISON, 2018). A Figura 16 mostra a relação entre os atores relacionados com a *Cambridge Analytica*.

Figura 16 - Relação entre os atores envolvidos com a *Cambridge Analytica*

²⁹ Tradução livre: Robert Mercer foi pioneiro em IA e tradução automática. Ele ajudou a inventar o algorítmico para atividades comerciais - que substituiu os gerentes de fundos de investimentos por programas de computador - e ele ouviu o discurso de Wylie. Que era a respeito de um novo tipo de direcionamento de mensagem política baseado em um artigo influente e inovador de 2014, elaborado pelo *Cambridge's Psychometrics Center*, denominado: “Julgamentos de personalidade baseados em computador são mais precisos do que aqueles feitos por humanos”.



Fonte: Cadwalladr e Graham-Harrison (2018)

Desde a vitória de Trump em 2016, alguns artigos vêm sendo escritos a respeito das estratégias de marketing em campanhas políticas envolvendo o uso de *big data*, aprendizado de máquina e *microtargeting* comportamental, porém o assunto alcançou enorme repercussão quando Christopher Wylie denunciou ao jornal britânico *Observer* como a *Cambridge Analytica* usou os dados para construir um sistema com o intuito de traçar o perfil dos eleitores americanos para direcionamento de anúncios políticos personalizados.

Segundo Cadwalladr e Graham-Harrison (2018), os documentos vistos pelo *Observer* e confirmados pelo próprio *Facebook* mostram que no final de 2015 a empresa descobriu que as informações haviam sido obtidas em uma escala sem precedentes. No entanto, o *Facebook* tomou apenas medidas limitadas para recuperar e proteger as informações privadas de mais de 50 milhões de pessoas.

O *Facebook* negou que a coleta de milhões de perfis pela GSR e pela *Cambridge Analytica* tenha sido uma violação de dados. De acordo com um comunicado do *Facebook*, Kogan obteve acesso aos dados de maneira legítima por meio de canais apropriados; no entanto, não cumpriu com as regras do *Facebook* quando passou a informação para terceiros. Além disso, o *Facebook* alegou que removeu o aplicativo em 2015 e exigiu a certificação de que todas as cópias dos

dados tinham sido destruídas. Contudo, agora o *Facebook* está investigando relatórios que mostram que nem todos os dados foram excluídos (CADWALLADR e GRAHAM-HARRISON, 2018).

Toda essa situação gerou grande indignação e culminou com a *Cambridge Analytica* e o *Facebook* sendo focos de um inquérito sobre dados e política feito pelo *British Information Commissioner's Office*, uma autoridade independente, no Reino Unido, que promove a proteção de dados pessoais. De acordo com Cadwalladr e Graham-Harrison (2018), o senador democrata Mark Warner disse que a coleta de dados em uma escala tão ampla como essa para a segmentação política requer controles mais rígidos e uma regulamentação da propaganda política *online* da mesma forma que a televisão, o rádio e a mídia impressa. Para o senador democrata Mark Warner essa situação é preocupante:

This story is more evidence that the online political advertising market is essentially the Wild West. Whether it's allowing Russians to purchase political ads, or extensive microtargeting based on ill-gotten user data, it's clear that, left unregulated, this market will continue to be prone to deception and lacking in transparency (apud CADWALLADR e GRAHAM-HARRISON, 2018, *online*)³⁰.

As eleições de Barack Obama em 2012 e, principalmente, a de Donald Trump em 2016, mostraram a consolidação do uso de aprendizado de máquina como ferramenta de marketing político nas campanhas eleitorais americanas e a influência que este tipo de ferramenta exerceu nas eleições. A predição do comportamento dos eleitores nas eleições americanas de 2016 para microssegmentação e direcionamento de discurso personalizado gerou vários questionamentos. Em 2018, quando o escândalo envolvendo a *Cambridge Analytica* e o vazamento de dados dos usuários do *Facebook* veio à tona, esses questionamentos ganharam força e acendeu um alerta para a privacidade dos dados e a necessidade de uma regulamentação da propaganda política *online*, entre outras medidas. O capítulo 6 visa direcionar essa análise para o cenário político brasileiro e levantar se as campanhas eleitorais no Brasil estão utilizando o aprendizado de máquina e as implicações do uso dessas tecnologias de controle nas disputas políticas.

³⁰ Tradução livre: Essa história é mais uma evidência de que o mercado de publicidade política *online* é essencialmente o “Velho Oeste”. Seja para permitir que os russos comprem anúncios políticos ou para uma ampla segmentação baseada em dados ilícitos de usuários, fica claro que, se não for regulamentado, esse mercado continuará propenso a enganos e falta de transparência.

6 APRENDIZADO DE MÁQUINA NAS CAMPANHAS POLÍTICAS NO BRASIL

Este capítulo é dedicado à pesquisa empírica, que tem como objeto de investigação o uso de aprendizado de máquina nas campanhas políticas no Brasil. Aqui são apresentados a metodologia da pesquisa, métodos e técnicas utilizados, assim como os dados levantados e os resultados obtidos. Mas antes, é feita uma síntese das últimas eleições no Brasil pontuando o papel da Internet e das TICs.

6.1 Uma Síntese das Últimas Campanhas Eleitorais no Brasil

Desde 1989 os brasileiros escolhem seus governantes por meio de eleições diretas. Mas, Santos e Varon (2018) acreditam que as eleições de 2018 foram marcadas como as primeiras, na história, a ter a Internet e ferramentas tecnológicas como um campo de atuação decisivo e controverso que superou a hegemonia das mídias de massa. Mas, toda essa mudança faz parte de um processo que teve início no primeiro mandato da Presidente Dilma Rousseff, em 2013, quando uma onda de protestos dominou o país. Esse acontecimento foi um divisor de águas no cenário político brasileiro, pois culminou em uma grande polarização política na eleições seguintes e destacou as redes sociais e a Internet como importantes atores no cenário político.

Nas eleições de 2014, Dilma Rousseff (PT) foi reeleita com 51,64% dos votos contra 48,36% de Aécio Neves (PSDB), com essa estreita diferença as eleições de 2014 foram as mais disputadas no Brasil desde o final da ditura.

Em um contexto de disputa tão feroz, a corrida presidencial foi cheia de brigas pessoais politicamente vazias entre os candidatos e uma série de reclamações jurídicas no Tribunal Superior Eleitoral sobre uso indevido do *Facebook* e publicidade *online* por coalizões políticas (SANTOS e VARON, 2018, p. 12).

Durante as eleições de 2014, pesquisadores do Laboratório de Estudo sobre Imagem e Cibercultura (Labic), da Universidade Federal do Espírito Santo, reportaram o uso de robôs no *Facebook* e *Twitter* pela campanha do candidato Aécio Neves (PSDB), durante um dos debates com Dilma. Analisando os perfis falsos que estavam publicando mensagens em favor do candidato, os pesquisadores identificaram que a maioria estavam ligados à conta de Eduardo Trevisan, do Comitê Financeiro do PSDB e da agência de propaganda responsável pela campanha.

Trevisan negou o gerenciamento dos perfis, mas admitiu que os serviços prestados incluíam monitoramento de redes sociais e análise de sentimentos. O comitê da campanha de Dilma fez uma denúncia sobre o uso de robôs no Ministério Público Eleitoral (MPE). Como não existia nenhuma lei sobre o uso de robôs, a denúncia exigiu uma multa alegando “propaganda política paga na internet atribuindo a prática a terceiros” (SANTOS e VARON, 2018, p. 36).

Segundo Santos e Varon (2018), o jornal *O Estado de São Paulo* também mostrou evidências sobre o uso de robôs pelo PSDB, durante e após as duas campanhas presidenciais. De acordo com o relatório, baseado em um documento interno da Secretaria de Comunicação Social da Presidência, a operação *online* da oposição após as eleições tinha por objetivo continuar alimentando as redes sociais com material contra Dilma, com denúncias sobre o envolvimento dela e do ex-presidente Lula na corrupção da Petrobrás e na operação Lava Jato, além da geração de conteúdo pró-impeachment. Também é mencionado que, além de robôs, a oposição também estava trabalhando de forma muito profissional com o *WhatsApp*.

A Diretoria de Análise de Políticas Públicas, da Fundação Getúlio Vargas (DAPP/FGV), publicou em agosto de 2017, o estudo “Robôs, Redes Sociais e Política no Brasil”. Nesse estudo foram analisados “casos de interferências ilegítimas no debate político” (SANTOS e VARON, 2018, p. 36). De acordo com o estudo:

Os robôs estão sendo usados tanto pela esquerda quanto pela direita, não apenas para aumentar o número de seguidores, mas também para conduzir ataques contra a oposição, promover *hashtags* e debates artificiais, criar e replicar mensagens em uma escala mais ampla e até disseminar “*fake news*” ou *links* maliciosos que roubam dados pessoais para criar novos perfis falsos e assim aumentar sua rede de robôs (SANTOS e VARON, 2018, p. 37).

Ainda em 2017, a *Cambridge Analytica* desembarcou no Brasil e firmou parceria com a consultoria Ponte Estratégia, de André Torreia, formando a CA-Ponte. De acordo com Monteiro (2018), o objetivo era atuar nas campanhas eleitorais de 2018 por meio de uma “versão tropicalizada” das metodologias aplicadas em outros países (2018, *online*). A *Cambridge Analytica* chegou a mencionar que seu interesse pelo Brasil surgiu em função das leis de proteção de dados não serem rígidas, no entanto isso demonstrou um total desconhecimento do Marco Civil da Internet.

Para Monteiro (2018), um ponto que acabou favorecendo o modelo de negócio da *Cambridge Analytica* foi a alteração na Lei Eleitoral que passou a permitir que candidatos e partidos utilizassem o impulsionamento de conteúdo pago (patrocinado). Monteiro (2018) chama a atenção para o fato de que a legalidade do impulsionamento de conteúdo eleitoral levanta diversas controvérsias, principalmente referentes às caixas-pretas dos algoritmos que governam e decidem os conteúdos que são exibidos a usuários nas plataformas *online*.

Tudo estava transcorrendo em harmonia e os ventos estavam favoráveis à parceria CA-Ponte. Mas, isso durou até o momento em que as denúncias de Christopher Wylie vieram à tona. A parceria não resistiu a uma onda de escândalos e a Ponte anunciou que estava desfazendo a sociedade, apesar de “declarar que continuaria a buscar novas formas de atingir o eleitorado nacional” (MONTEIRO, 2018, *online*).

De acordo com Santos e Varon (2018), em junho de 2018, pesquisadores de marketing levantaram que o candidato com maior influência nas redes sociais era Jair Bolsonaro, com mais de 10 milhões de seguidores no *Facebook*, *Twitter* e *Instagram*. Além disso, a estratégia dele focou em grupos do *WhatsApp* e páginas de seus defensores de extrema direita nas redes sociais.

No dia 06 de setembro, aconteceu outro fato muito importante, Bolsonaro foi esfaqueado em um evento de campanha, em Minas Gerais. Por conta desse incidente justificou sua ausência nos debates promovidos pelas mídias tradicionais. Passando então a utilizar cada vez mais as mídias sociais e o *WhatsApp* como meios de divulgação e comunicação com os eleitores.

Segundo Santos e Varon (2018), o jornal *Folha de São Paulo*, em 18 de outubro de 2018, publicou uma matéria sobre empresários que estavam pagando disparos pelo *Whatsapp* contra o PT, sem declará-los, infringindo com isso a lei eleitoral que proíbe doações corporativas. Ao passo que “a campanha de Bolsonaro registrou apenas cerca de US\$ 34.000 como gasto em mídias digitais em um contrato com a empresa AM4 Brasil Inteligência” (2018, p. 17).

Em maio de 2019, o INTERNETLAB fez um estudo exploratório sobre esse “*spam* político no *WhatsApp*”. De acordo com o estudo, não só Bolsonaro como também outros candidatos e partidos utilizaram esse tipo de tática como estratégia

de marketing político. Esse estudo está disponível na Internet³¹ e fornece muita informação sobre esse assunto.

Além de todos os cenários descritos anteriormente que contribuíram para toda essa mudança nas eleições de 2018, merece destaque o da Lei da Reforma Política (Lei 13.488 / 2017), além de permitir que partidos, candidatos e coalizões utilizem as plataformas digitais e anúncios patrocinados como formas de propaganda eleitoral, também reduziu o tempo obrigatório de propaganda eleitoral na TV e no rádio (SANTOS e VARON, 2018).

Diante desse cenário, o aprendizado de máquina está sendo utilizado em campanhas eleitorais no Brasil? Em caso positivo, a partir de quando e como o aprendizado de máquina está sendo utilizado? Para responder a estas questões serão analisados os actantes relacionados com o aprendizado de máquina que é o objeto desta pesquisa.

6.2 Metodologia da Pesquisa

A Metodologia de pesquisa utilizada neste trabalho é exploratória e qualitativa, além de contar com revisão bibliográfica e pesquisa empírica.

De acordo com Souza (2018), para investigar um ecossistema que envolve um conjunto de atores humanos e não humanos, denominados como actantes, a perspectiva da teoria ator-rede (TAR) é uma excelente bússola. Essa teoria nasceu no âmbito dos estudos de ciência e tecnologia e foi desenvolvida por Bruno Latour, John Law, Madeleine Akrich, Andy Barry, entre outros pesquisadores.

A teoria ator-rede consiste em retrair as ações que atores realizam, descrevendo as associações e redes que se formam na composição de um coletivo. Para Latour, o trabalho de descrever o social é, portanto, um trabalho de formiga, e não foi à toa que ele usou a abreviação ANT (formiga em inglês) para a teoria ator-rede. Para ele, é preciso seguir o caminho (trilha) das formigas para em seguida levantar e verificar seus movimentos (LATOUR, 2014).

A teoria ator-rede pode ser utilizada como uma base norteadora para os pesquisadores que investigam objetos contidos em um ambiente complexo onde humanos e não humanos atuam conjuntamente. Contudo, Souza (2018) pontua que

³¹ <http://www.internetlab.org.br/pt/informacao-e-politica/santinhos-memes-e-correntes-um-estudo-sobre-spams-nas-eleicoes/>

o “pensamento de Latour sobre rede, em nenhum momento faz referência às redes digitais” (2018, p. 70). Para ela, isso fica claro quando ele explica o termo ator da teoria ator-rede: “o ator não é a fonte de um ato e sim o alvo móvel de um amplo conjunto de entidades que enxameiam em sua direção” (LATOURE, 2014, p. 75).

Com a teoria ator-rede é possível seguir rastros e com eles levantar as ações que os actantes efetuam, assim como as associações que se formam com o uso de técnicas de aprendizado de máquina no marketing e como isso pode ser transportado para as campanhas políticas.

Por conta das características do segmento de campanhas eleitorais possuem um teor bem sensível, considerando principalmente o lado sigiloso das empresas envolvidas nesse segmento, é fundamental o uso da metodologia conhecida como *snowball* (bola de neve). Esta metodologia é uma técnica de amostragem que utiliza cadeias de referência, que seria como uma espécie de rede (SOUZA, 2018). Bernard (2005) explica que essa técnica é um método de amostragem de rede útil para investigar questões difíceis de serem analisadas ou que não possuam uma precisão quantitativa. Para reforçar esse entendimento, Biernacki e Waldorf (1981) consideram que a amostragem em bola de neve também pode ser utilizada quando a pergunta da pesquisa estiver relacionada com questões problemáticas ou de certa forma sigilosas para os entrevistados.

A amostragem no método bola de neve é feita da seguinte forma: considera-se documentos e/ou informações-chaves, como sementes, para identificar algumas pessoas dentro da “população geral” com o perfil relacionado com a pesquisa. Em seguida, procura-se com base nas pessoas selecionadas pelas sementes identificar novos contatos que estejam de acordo com as características desejadas e assim sucessivamente. “É uma forma de executar o preceito da teoria ator-rede de seguir os rastros” (SOUZA, 2018, p. 74).

Os actantes que podem contribuir para entender o uso aprendizado de máquina nas campanhas políticas aqui no Brasil são: consultores de marketing político, empresas que fornecem soluções para a área do marketing e que utilizam aprendizado de máquina e especialistas em aprendizado de máquina. Assim como as soluções e serviço disponibilizados por estas empresas.

6.2.1 Métodos de Pesquisa

Para o desenvolvimento da pesquisa empírica foram utilizados os seguintes métodos: entrevistas com os actantes relacionados com o objeto desta pesquisa e análise documental dos sites, *blogs*, *papers* das empresas selecionadas, assim como matérias veiculadas na mídia.

A escolha pela entrevista objetivou esclarecer mais sobre a temática e o uso de aprendizado de máquina em campanhas políticas no Brasil. Foram utilizadas entrevistas semiestruturadas e não estruturadas. A entrevista semiestruturada é guiada por um roteiro de questões, é uma ferramenta útil quando a entrevista é feita por telefone, *e-mail* ou *WhatsApp*. A entrevista não estruturada é aquela que permite mais liberdade na formulação de perguntas e na intervenção da fala do entrevistado, é válida quando a entrevista é feita pessoalmente (ANTONIUTTI, 2015).

A análise dos sites, *blogs*, *papers* e matérias veiculadas na mídia visa levantar informações complementares que ajudem a embasar as informações coletadas nas entrevistas.

Foram entrevistados os seguintes actantes humanos: consultor de marketing político, empresas que fornecem soluções para a área do marketing e que utilizam aprendizado de máquina e especialistas em aprendizado de máquina. Posteriormente, com base nos dados levantados junto às empresas, foram identificados e analisados os actantes não humanos (soluções das empresas).

A Figura 17 mostra de forma gráfica os actantes selecionados. Todos eles estão localizados na cidade de São Paulo.

Figura 17 - Actantes selecionados



Fonte: Autora

Por ser um ator muito importante no cenário político brasileiro e por conta de sua parceria com a *Cambridge Analytica*, Torreta foi o consultor de marketing político escolhido para contribuir com esta pesquisa.

Pesquisas referentes à campanha presidencial de Barack Obama levaram ao *software NetBase*, que foi trazido para o Brasil, pela Polis Consulting. Esse foi o motivo que justificou a escolha desta empresa.

A Elife é umas das empresas descritas na tese “Usos do Big Data em Campanhas Eleitorais”, de Antoniutti (2015), e foi por conta dessa referência que esta companhia foi selecionada.

O estudo “Liga *Insights IT Startups*”³², de julho de 2018, destaca as *startups* Tail e Stilingue como sendo de *Data Science* e Enriquecimento de Dados. Devido *Data Science* ser sinônimo de aprendizado de máquina estas *startups* foram consideradas na pesquisa.

A empresa MarketData foi selecionada por fazer parte da rede de atuação de Marcelo Sousa que atualmente é o Presidente da ABRADi³³ (Associação Brasileira dos Agentes Digitais). A ABRADi tem um papel importante no marketing digital, assim como Marcelo Sousa tem uma atuação relevante em palestras e eventos referentes a esse segmento. Foi por meio do IX Seminário de Proteção à

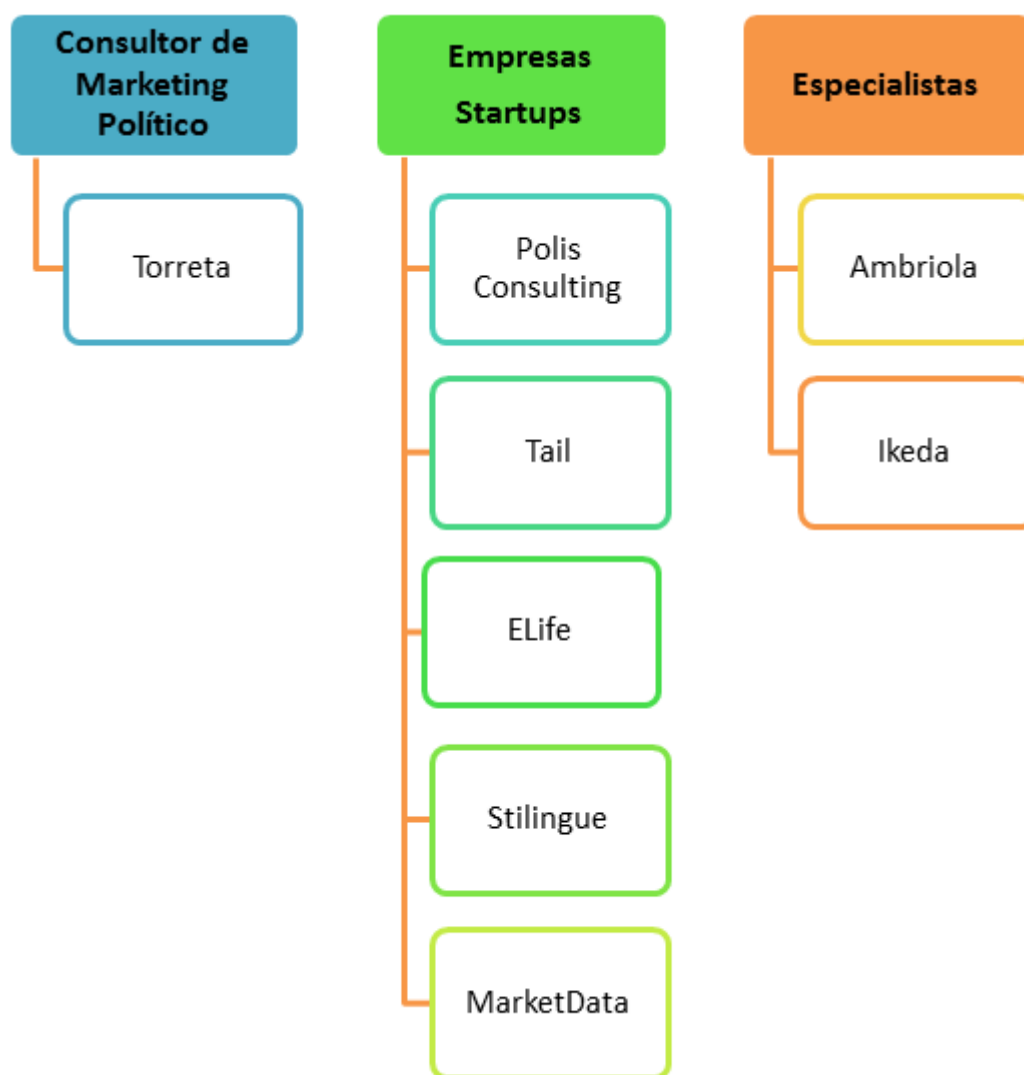
³² <https://insights.liga.ventures/itstartups/>

³³ <https://abradi.com.br/abradi/diretoria/>

Privacidade aos Dados, promovido pelo Comitê Gestor da Internet no Brasil (CGI.br), que esse actante foi identificado.

Os especialistas foram considerados nesta pesquisa com o intuito de trazer uma visão mais técnica referente ao objeto pesquisado.

Figura 18 - Actantes pesquisados



Fonte: Autora

Torreta é publicitário, empresário, escritor e palestrante. Trabalhou como redator e diretor de criação em empresas de comunicação, como Editora Abril, Editora Globo e Duda Mendonça. Além dos segmentos privados, também atua como estrategista em campanhas políticas. Em 2007 lançou a Ponte Estratégia, consultoria de marketing voltada para a classe C. No cenário do marketing político, Torreta é um ator muito importante, não só por conta de sua vasta experiência nesse segmento, mas também pela parceria com a *Cambridge Analytica*.

A Polis Consulting é um grupo brasileiro, fundado em 2014, com sedes em São Paulo e Hergiswil, na Suíça. Atua com tecnologias e serviços para BI, CRM e *Social Intelligence*. Trabalha também com as plataformas tecnológicas em *Social Media Management*. O escritório, em São Paulo, fica na Alameda Lorena, 800, cj 1904, Cerqueira César. Tem como CEO, Alexander Schmitz-Kohlitz, alemão, M.Sc em Engenharia Mecânica pela RWTH Aachen, da Alemanha. DEA, Engenharia Química pela *Ecole Centrale* e MBA pelo *Collège des Ingénieurs*, ambos em Paris.

A Tail (*Target Audience & Insights Lab*) é uma *startup* brasileira criada em 2012. Atua com plataforma de *Data Intelligence*. Tem filial no México, Colômbia, Argentina e Chile e sua sede fica em São Paulo, na Av. Pedroso de Moraes, 1553 - cj 51, Alto de Pinheiros. Tem como CEO e Cofundador Cristiano Nobrega que é Mestre em Administração de Empresas pela EAESP-FGV e pós-graduado em Economia e Marketing pela UC-Berkeley (EUA). Foi eleito um dos 50 Profissionais Mais Inovadores de 2013 pela Revista *ProXXima*. Além disso, foi sócio-fundador do portal *ObaOba* e diretor de Novos Negócios da RBS-Digital. Cristiano Nobrega também é Presidente do Conselho Executivo³⁴ da IAB Brasil (*Interactive Advertising Bureau*)³⁵.

A Elife foi criada em 2004, atua no Brasil, Portugal, Espanha e México. É uma empresa que oferece serviços e *software* especializados em inteligência de mercado e gestão de relacionamento com o consumidor em meios digitais e redes sociais. O escritório em São Paulo fica na Rua Pamplona, 518, Jardim Paulista. Tem como CEO, Alessandro Barbosa Lima que é especializado em inteligência e *big data*, Mestre em Comunicação pela Universidade de São Paulo, Bacharel em Jornalismo pela UFPE, autor e palestrante especializado em Comunicação e Marketing.

A Stilingue é uma *starturp* que tem Rodrigo Helcer como fundador e CEO. Helcer é formado em Administração pela FEA-USP, professor, palestrante e curador para instituições como FGV, ESPM, Festival Path, Youpix e eventos de inovação no país.

³⁴ <https://iabbrasil.com.br/sobre-iab/diretoria-iab/>

³⁵ A Associação de Mídia Interativa (AMI) é uma entidade sem fins lucrativos que foi fundada em 1998, com a missão de desenvolver o mercado de mídia interativa no Brasil. Em 2006 passou a fazer parte do *Interactive Advertising Bureau*, mudando sua denominação para IAB Brasil que tem Ana Moises como a atual Presidente.

A MarketData está localizada à Av. das Nações Unidas, 12495 - 7º Andar - Brooklin Novo, São Paulo. Marcelo Sousa, que atua nessa empresa, é um *Data Marketeer*, com experiência em *customer analytics* e *data driven marketing strategies*, design e implementação em projetos de CRM. Além disso, ele também é Diretor Executivo no grupo WPP e Presidente³⁶ da ABRADi (Associação Brasileira dos Agentes Digitais)³⁷.

Ambriola é formada em Matemática pela USP, com pós-graduação em Inteligência de Mercado. Atualmente é mestranda em Engenharia da Informação, na UFABC, onde atua com pesquisa utilizando *machine learning* aplicada à neurociência. Trabalhou em *Pricing* e Planejamento Estratégico e em *Bureau* de Crédito, além de prestar consultorias em empresas com modelagem matemática. É suplente na CGI Br e presidente do Garoa Hacker Clube. Participou do edital “Elas nas Exatas” para disseminar conhecimentos de exatas entre as meninas de comunidades carentes em São Paulo.

Ikeda é engenheiro de Computação, formado pela Unicamp. Há sete anos estuda *machine learning* e *data science*. Possui certificação em *Data Science* da EMC (hoje *Dell Technologies*). Em 2018 concluiu três especializações na plataforma Coursera, nos seguintes cursos: *Deep learning*, *machine learning with TensorFlow on Google Cloud Platform* e *Architecting with Google Cloud Platform*. Atua na CAS, em um projeto de P&D com a Light Distribuidora de Energia, para aplicar *deep learning* em detecção de fraudes. Além disso, há alguns projetos de sua empresa relacionados a reconhecimento de imagem.

6.3 Consultor de Marketing Político

Em relação a este actante foi feito um levantamento das matérias veiculadas na mídia referentes à parceria da Ponte Estratégia com a *Cambridge Analytica* (CA). Também foi analisada palestra apresentada por Torreta, em 2018, na Câmara dos Deputados. Além disso, foi realizada entrevista presencial para levantamento de informações adicionais.

³⁶ <https://abradi.com.br/abradi/diretoria/>

³⁷ A ABRADi é uma entidade sem fins lucrativos que visa apoiar empresas desenvolvedoras de serviços digitais no Brasil. Tem sede em São Paulo e reúne cerca de 600 empresas, presente em 14 estados brasileiros.

6.3.1 Matérias Veiculadas sobre Torreta e a *Cambridge Analytica*

Segundo Ribeiro (2018), a Ponte Estratégia, empresa de Torreta, foi procurada pela *Cambridge Analytics* (CA) no início de 2017, para formar uma parceria no Brasil. De acordo com este autor, a parceria foi aceita e nasceu a CA Ponte. Ribeiro (2018) explica que o principal critério da *Cambridge* para a escolha da Ponte foi a experiência de Torreta. Já o interesse da Ponte, foi complementar o conhecimento na área de dados.

De acordo com Passarinho (2018), após criação da CA Ponte, Torreta se reuniu com equipes de candidatos para apresentar a possibilidade de ter o *expertise* da *Cambridge Analytica* na segmentação e análise dos dados de eleitores. A meta era “direcionar mensagens políticas personalizadas, que apelassem para os medos, anseios e demandas de cada tipo de brasileiro” (PASSARINHO, 2018, *online*).

Nesse período, Torreta apresentou várias entrevistas como representante da *Cambridge Analytica* no Brasil. De acordo com Passarinho, ele “estava ansioso para ampliar o uso do direcionamento inteligente das mensagens políticas para o *Whatsapp*”, ideia que CA ainda não havia vislumbrado (PASSARINHO, 2018, *online*).

Segundo Passarinho (2018), Torreta mencionou que, embora a *Cambridge* promettesse uma metodologia eficaz na segmentação de eleitores em categorias psicológicas (patriota, progressista, emocional), ela não tinha um banco de dados com perfis dos brasileiros. O objetivo era que ele coletasse os dados de eleitores:

Mas eu não iria pedir isso no *Facebook*. Imagina eu indo lá e pedir. O pessoal do *Facebook* ia rir da minha cara. Ia usar o que já tem disponível na rede social e fazer pesquisas em bairros, entrevistas, usar dados do IBGE (TORRETA, apud PASSARINHO, 2018, *online*).

Conforme relata Passarinho, a CA seria responsável em dar apoio na análise dos dados recolhidos por Torreta, para classificar os eleitores brasileiros em perfis. A partir disso, “criariam em conjunto estratégias de comunicação direcionada para os diferentes segmentos e adequadas aos candidatos que contratassem o serviço” (PASSARINHO, 2018, *online*).

Passarinho (2018) ainda observa que a *Cambridge* tinha interesse nas eleições brasileiras de 2018, pois enxergava uma oportunidade de expandir seu negócio na América Latina. Tinham como meta atuar com candidatos “à Presidência, a governador, a deputado e senador”. As negociações com possíveis candidatos

estavam no início e nenhum contrato havia sido firmado. Torreta inclusive enfatizou que “nunca trabalharia para Jair Bolsonaro” (PASSARINHO, 2018, *online*).

Em 2018, quando estourou o escândalo envolvendo a *Cambridge Analytica* e o vazamento de dados do *Facebook*, a parceria entre a CA Ponte foi desfeita. Na época, Torreta disse ter ficado surpreso quando soube da investigação aberta para apurar a atuação da *Cambridge Analytica* (RIBEIRO, 2018, *online*).

6.3.2 Uso de Dados Pessoais como Instrumento de Campanha Eleitoral e a Persuasão da Opinião Pública

Em maio de 2018, Torreta fez uma apresentação sobre “Uso de dados pessoais como instrumento de campanha eleitoral e a persuasão da opinião pública”, durante o Seminário sobre Proteção de Dados Pessoais, promovido pela Câmara dos Deputados, em Brasília. O vídeo³⁸ e os slides³⁹ dessa apresentação estão disponíveis no site da Câmara dos Deputados.

Após fazer uma breve apresentação pessoal, Torreta (2018) conta que entrou para o marketing político durante a campanha eleitoral de Tancredo Neves e Maluf, e enfatizou a importância da informação numa campanha política: “Quando eu preciso persuadir alguém, eu preciso saber quem você é para eu poder dizer o que você vai fazer” (TORRETA, 2018, *online*).

De acordo com Torreta (2018), as informações estão nos grandes bancos de dados, “nos Estados Unidos muito mais, na Europa muito menos e no Brasil um pouco menos” (2018, *online*). De acordo com ele, o acesso a esses bancos de dados permite conhecer melhor e, conseqüentemente, persuadir melhor para vender o que se deseja.

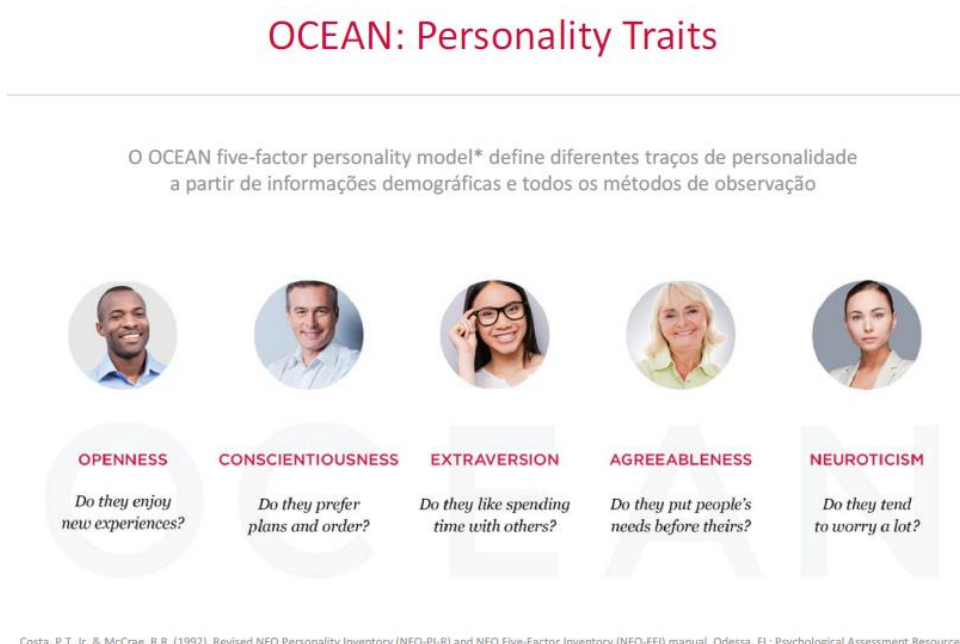
Em seguida Torreta (2018) fala sobre o *big data* e como essa tecnologia pode ser utilizada para conhecer as pessoas. De acordo com ele, as informações contidas nesses bancos de dados são necessárias para conhecer melhor o público-alvo de uma campanha.

³⁸ <http://www2.camara.leg.br/atividade-legislativa/webcamara/videoArquivo?codSessao=72720#videoTitulo>

³⁹ <https://www2.camara.leg.br/atividade-legislativa/comissoes/comissoes-permanentes/cctci/Eventos/2018/2018-05-22-seminario-dados-pessoais/painel-2/andre-torreta-a-ponte-estrategia>

Torreta (2018) comenta que em 1961 descobriu-se uma forma de segmentar por comportamento e que posteriormente o método de psicometria OCEAN ou *Big Five* foi aprimorado. Segundo ele, hoje é possível não apenas comandar a comunicação certa, mas também enviar a linguagem certa. É o que eles chamam de “*targeting* comportamental” (Torreta, 2018, *online*). A seguir, na Figura 19, confira o slide apresentado por Torreta com os traços de personalidade OCEAN ou as cinco dimensões do *Big Five*, como também é conhecido.

Figura 19 - Slide com os traços de personalidade OCEAN

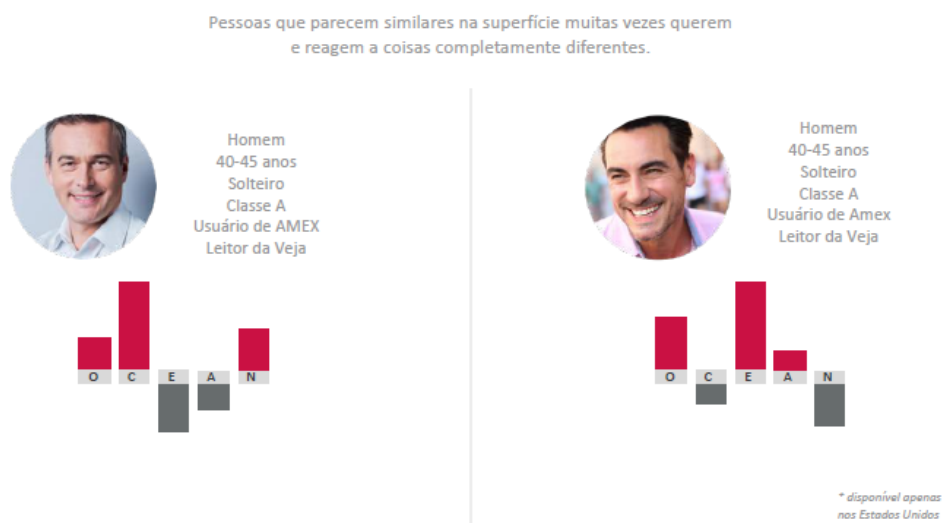


Fonte: Torreta (2018)

Para explicar melhor a questão da personalidade, Torreta (2018) usou um slide para demonstrar que pessoas da mesma faixa etária, classe social, com estilo de vida e gostos de leitura parecidos tiveram resultados diferentes no OCEAN, o que conseqüentemente denota que elas possuem traços de personalidades diferentes.

Figura 20 - Slide com pessoas parecidas com resultado do OCEAN diferente

Iguais mas Diferentes



Fonte: Torreta (2018)

Torreta (2018) explica que dentro das redes sociais é possível definir quem é quem. Qual é o grau de Abertura (*Openness*) as novas experiências; a Consciência (*Conscientiousness*) e o quanto as pessoas são perfeccionistas; a Extroversão (*Extraversion*) se uma pessoa é sociável ou não; a Amabilidade (*Agreeableness*) se uma pessoa é cooperativa e atenciosa e o Neuroticismo (*Neuroticism*) se a pessoa se irrita com facilidade. Torreta (2018) cita que os neuróticos que correspondem ao *Neuroticism*, no OCEAN, são pessoas mais nervosas e mais intranquilas:

Então se eu quero fazer amanhã uma campanha para o Bolsonaro eu vou fazer com quem? Aqui (os neuróticos). Se eu for fazer uma campanha onde o cara é mais extrovertido, é mais consciente e mais aberto eu vou fazer para Gabeira. Por quê? Porque o cara é assim (Torreta, 2018, *online*).

Para Torreta (2018), uma vez que seja entendido o comportamento das pessoas e feita a migrosegmentação dos traços de personalidade, é possível mandar mensagens diferentes para cada uma delas de forma personalizada, com imagens e narrativas diferentes e considera:

E aí eu começo a entender que o cara que mora em Brasília, que tem 45 anos, não é exatamente igual a outro cara que mora em Brasília com 45 anos. Eles são pessoas diferentes. Eles têm personalidades diferentes e merecem mensagens diferentes, e por isso mesmo eu comando para cada um deles uma comunicação diferente (Torreta, 2018, *online*).

Segundo Torreta (2018), para o homem de 45 anos que mora em Brasília e tem alta conscienciosidade deve ser enviado o anúncio de um carro focado na proteção do meio ambiente.

Figura 21 - Slide com anúncio para pessoa com alta conscienciosidade

Alta Conscienciosidade



Homem
40-45 anos
Solteiro
Classe A
Usuário de AMEX
Leitor da Veja

O C E A N

Focus Electric

Zero gas, zero oil changes and zero CO2 emissions

It delivers a 100 percent electrifying driving experience. What's more, you have the choice of charging options —

the standard 120-volt/10-amp convenience charge cord or the available 240-volt/30-amp home charging station.

Equipment Group 500A

- 107kW electric motor
- 1-speed automatic transmission
- SmartKey® with Start/Stop

Ford
Go Further

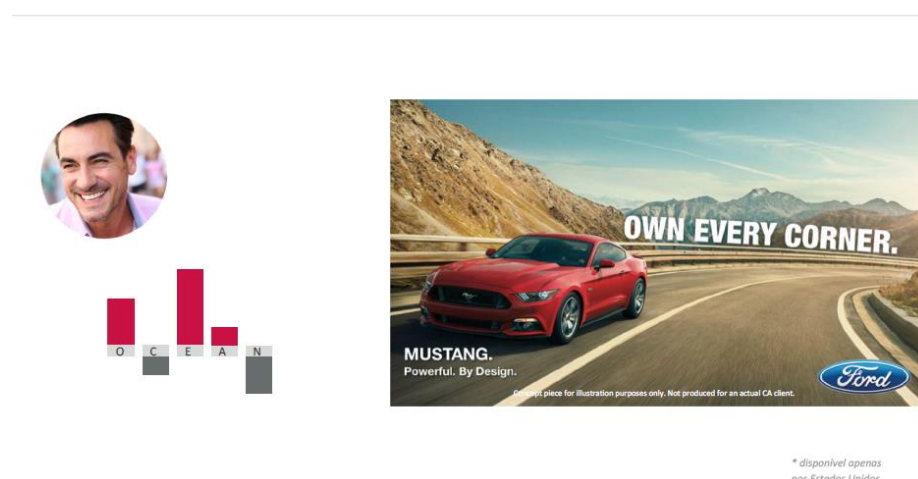
** disponível apenas nos Estados Unidos*

Fonte: Torreta (2018)

Para o homem de 45 anos que também mora em Brasília e tem alta extroversão, Torreta (2018) explica que deve ser enviado o anúncio de um carro com foco mais na esportividade.

Figura 22 - Slide com anúncio para pessoa com alta extroversão

Alta Extroversão



O C E A N

MUSTANG.
Powerful. By Design.

OWN EVERY CORNER.

Ford

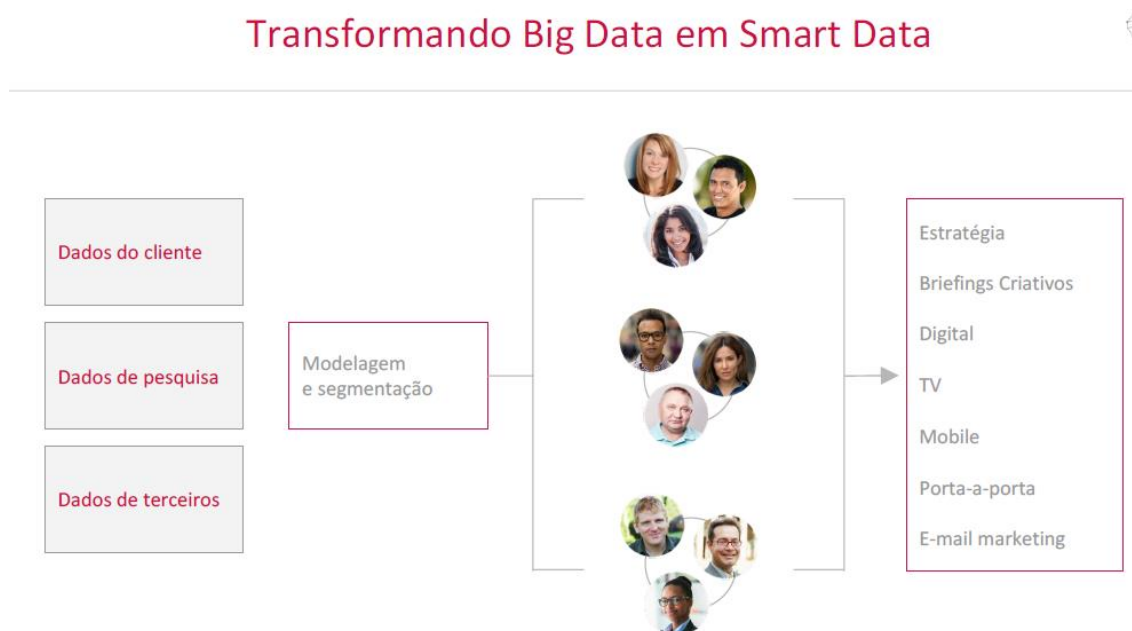
** disponível apenas nos Estados Unidos*

Fonte: Torreta (2018)

Torreta (2018) mostra que por meio de dados de cliente, dados de pesquisa e dados de terceiros é possível fazer uma modelagem e segmentação capaz de

mapear o *Big Five* de forma a viabilizar o envio de campanhas mais direcionadas e personalizadas. Ele esquematizou isso no slide “Transformando *Big Data* em *Smart Data*”.

Figura 23 - Slide Transformando *Big Data* em *Smart Data*



Fonte: Torreta (2018)

“E ai a gente” começa a descobrir quem é quem. Começa a distribuir a comunicação de uma maneira certa para o cara do *Whatsapp* e para o cara do *Facebook*, vou distribuindo a comunicação conforme o meu alvo (Torreta, 2018, *online*).

Segundo Torreta (2018), na campanha do Donald Trump foi aplicada essa metodologia. Para ilustrar, Torreta (2018) mostra um slide (Figura 24) com dois exemplos de mensagens. Para pessoas com alto neuroticismo e conscienciosidade foi enviada a primeira mensagem, cujo foco está no direito ao porte de arma e mais segurança. Já para pessoas fechadas e amáveis, foi enviada a segunda mensagem, cujo porte de arma foi relacionado com a tradição que vai passando de pai para filho. Torreta (2018) explica que, na verdade, o que mudou nas duas mensagens foi a argumentação, e, teoricamente, modificando a argumentação se consegue falar melhor com as pessoas, e com essa comunicação mais assertiva se conquista mais votos.

Figura 24 - Slide Mensagens segmentadas de acordo com os traços do *Big Five*

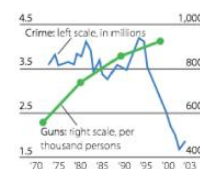
Mensagem com Segmentação Psicográfica



Alto neuroticismo e
conscienciosidade



Fechado e amável



Alta conscienciosidade
e amável

Fonte: Torreta (2018)

6.3.3 Principais Pontos da Entrevista

Para Torreta (2019), o que distingue o marketing comercial (que ele denomina como marketing da iniciativa privada) do marketing político são as características de cada um: “As características são diferentes da iniciativa privada. Na iniciativa privada ela vende serviços e produtos. Aqui (no marketing político) eu vendo ideias e causas” (TORRETA, 2019).

Torreta (2019) comentou que atua nesses dois segmentos. Na iniciativa privada ele fez um estudo de marketing para a população de baixa renda:

Com a ascensão do Lula, a população de baixa renda começou a ter recursos e as empresas não sabiam lidar com esse novo consumidor. Como eu venho da política, as pessoas viam que eu entendia de povo (TORRETA, 2019).

De acordo com Torreta (2019), o mercado publicitário brasileiro olha muito para Nova York e Europa. Para ele, esse comportamento é errôneo, conforme observa:

O mercado publicitário olha muito para Nova York e para a Europa. O que dá errado é que a gente não vive na Europa. Eles não notaram ainda. A gente não vive em Nova York. Então, quando se fala, por exemplo, em tecnologia, o que dá certo em Nova York não dá certo em Petrolina. E não é que Nova York e Petrolina sejam errados é que são cidades com características diferentes. Então, foi por isso que eu resolvi ir para a iniciativa privada. Ajudar a compreender esse Brasil de verdade (TORRETA, 2019).

Ao ser questionado se utilizou algum banco de dados e/ou alguma ferramenta tecnológica quando fez o estudo sobre a população de baixa renda Torreta (2019) disse o seguinte:

Naquela época não existia ferramenta tecnológica. Como eu tinha entrado no marketing político e o meu pai também, a gente viajou muito (TORRETA, 2019).

De acordo com Torreta (2019), com base no conhecimento adquirido sobre o Brasil por meio das viagens realizadas ele começou prestar assessoria para iniciativa privada:

Hoje na iniciativa privada eu faço muita palestra explicando o que é Brasil. Tanto para executivo de fora como executivos daqui. Meu *know-how* em relação a isso, em primeiro lugar, vem dessa minha vivência de coisas e pelo interesse na pesquisa. A gente faz muita pesquisa qualitativa (TORRETA, 2019).

Sobre o questionamento se o volume de dados que temos hoje impõe um novo olhar para as estratégias políticas e campanhas eleitorais, ele argumentou:

Esse mundo em que a gente está vivendo impõe um novo modelo para tudo. Você tem isso na indústria fonográfica; há 15 anos se vendia LP e, de lá para cá, olha a mudança que já teve. O mesmo acontece com a indústria de livros, olha por quanta transformação ela já passou. Se você vai para o serviço de táxi, olha a mudança que já teve (TORRETA, 2019).

Se as agências de marketing e as agências de publicidade estão tendo uma visão (sacada) desse novo modelo, Torreta aponta alguns entraves:

O problema não é a sacada. Você tem algumas variáveis. Primeiro é a legislação. A legislação americana é muito aberta e deixa você ter acesso a quase tudo de dado. A legislação Europeia já não deixa ter. Aqui no Brasil, mudou agora com a Lei de Proteção de Dados. Ela está se fechando, o que é bom (TORRETA, 2019).

Ao ser questionado se essa reação teve alguma relação com o ocorrido com a *Cambridge Analytica*, Torreta acredita que a criação da Lei de Proteção de dados não foi influenciada por conta do escândalo da *Cambridge*. Inclusive ele argumentou o seguinte: “Toda vez que tem uma campanha presidencial norte-americana há um escândalo” (TORRETA, 2019).

Na campanha presidencial de 2008, de Obama, Torreta (2019) explica que o *Facebook* foi usado, principalmente, para arrecadar dinheiro e falar com os líderes comunitários. Segundo ele, os Estados Unidos nessa época tinham muita mala direta e muita visita porta a porta.

Antes das redes sociais você tinha o quê? Tinha o telemarketing, e isso também era feito no Brasil. O telemarketing era muito poderoso no Brasil. Foi utilizado pela primeira vez na campanha do Quércia e do Paulo Maluf. Com Paulo Maluf na prefeitura, você tinha um programa chamado Bairro a Bairro, então dividia-se São Paulo em bairros, e com o telemarketing você conseguia falar com aquele bairro sobre um assunto e você endereçava o conteúdo correto (TORRETA, 2019).

Torreta foi indagado se nessa época do programa do Maluf já existia segmentação e a resposta foi:

A segmentação sempre existiu. A diferença é que agora existe o *microtargeting* comportamental que é você diferenciar, não apenas o conteúdo que você envia, mas a linguagem que você envia. No começo era só o conteúdo que era diferente, a forma não. Com o *microtargeting* comportamental passou a ser conteúdo e forma (TORRETA, 2019).

A segmentação na campanha presidencial dos Estados Unidos, de 2012, para Torreta foi um divisor de águas:

As pessoas acham que foi a campanha do Trump, mas foi a do Obama. E como o Obama era amigo do pessoal de informática, dos *hackers* e da mídia, “você não bate” (TORRETA, 2019).

Em relação à campanha de Trump, o que Torreta considera inovador é a psicometria e não necessariamente o *Big Five*. Para ele:

Foi o marketing agressivo da *Cambridge* que a prejudicou. Na verdade você tem a psicometria criada por aquele professor de Stanford, mas de qualquer jeito as segmentações comportamentais, existem há 40 anos. O que não dava para fazer era do jeito que existe atualmente por falta de dados. Então agora você tem os dados. Nos Estados Unidos os dados estão disponíveis. No Brasil eu não tenho, tenho o Serasa Experian, que é nosso grande fornecedor de dados (TORRETA, 2019).

Para Torreta (2019), o problema está em onde conseguir obter os dados comportamentais ou um dado que sirva para a psicometria.

Se eu for pegar o banco de dados do Corinthians, os corinthians dentro da psicometria têm perfil diferente dos são-paulinos, ou baianos têm perfil diferente dos cearenses. Isso é psicometria. É que isso nunca foi utilizado dessa maneira. O que a *Cambridge* fez de genial, pelo menos o que eu acho, é como utilizar o que estava na mesa (TORRETA, 2019).

Quanto à efetividade do uso da psicometria, Torreta (2019) considera que pode haver uma melhora na possibilidade de argumentação uma vez que essa técnica visa atingir o desejo e a emoção das pessoas. Torreta (2019) explica que o marketing explora tanto o emocional quanto o racional. Citou o exemplo da indústria

do tabaco, que nos anos 60 e 70 comprava estudos científicos dizendo que o tabaco era bom para a saúde. Segundo ele, “isso a *Cambridge* não fez” (TORRETA, 2019).

De acordo com Torreta (2019), o Brasil enfrenta dificuldades na aquisição de dados:

Não existem dados à venda como nos Estados Unidos, e montar uma base de dados custa caro. Quem atua nesse segmento de análise de dados é a inteligência de mercado. Nos Estados Unidos você consegue comprar dados na esquina. Aqui no Brasil ninguém tem os dados prontos. Ninguém tem porque custa trabalho (TORRETA, 2019).

Sobre o aprendizado de máquina no Brasil Torreta considera o seguinte:

O aprendizado de máquina não chegou aqui ainda. Principalmente nesse assunto específico. Pelo menos que eu saiba. Quem está usando isso são empresas privadas, mas essas empresas não costumam declarar o que fazem. É uma “caixa preta”. A grande pergunta na verdade é se isso é realmente utilizável ou é “um sonho de uma noite de verão”. Essa foi a diferença com a *Cambridge*. Eles disseram: “Vamos segmentar o sonho de todo mundo? Vamos segmentar” (TORRETA, 2019).

Ao ser questionado se isso é o sonho de todo publicitário, Torreta (2019) disse que “publicitário não entende disso”. Segundo ele, “agências de propaganda que trabalham com dados de verdade são poucas”. “Poucas pessoas dominam isso, porque é algo novo que acabou de ser criado” (TORRETA, 2019).

Quanto à possibilidade de a *Cambridge Analytica* ter influenciado as eleições nos Estados Unidos, em 2016, Torreta (2019) presume que sim. De acordo com ele:

O que todo mundo admite que a *Cambridge* fez na verdade, de forma muito inteligente, foi a estratégia e não foi a segmentação. Qual estado iria lutar pelos votos e qual estado não iria. Mas, todo mundo ficou pendurado no OCEAN (TORRETA, 2019).

Sobre a tecnologia ter um papel importante nesse cenário ou ser apenas uma coadjuvante Torreta (2019) considera o seguinte:

Toda vez que tem uma nova tecnologia no mundo, o povo enlouquece. Com a tipografia, o povo queimou livro e bruxa. Depois veio a Segunda Guerra Mundial que foi um grande aprendizado para todo mundo sobre marketing político, com Goebbels e a utilização do rádio. Você pega isso e está sendo replicado até hoje. Aqui no Brasil teve o Departamento de Imprensa e Propaganda do Getúlio Vargas (DIP) que também utilizou muito o rádio (TORRETA, 2019).

Quanto a ter tido experiência com aprendizado de máquina na parceria com a *Cambridge*, Torreta (2019) comentou:

Vamos usar o que dá para ser usado. Até ensinar uma máquina no Brasil como o Bradesco está fazendo com a BIA⁴⁰, iria demorar 3 anos para ficar pronto. A gente não tinha tempo para isso. O marketing político é mais sensível do que o marketing da iniciativa privada. No Bradesco, você topa ser atendido por uma máquina, numa campanha eleitoral você não topa. Você quer ser atendido por uma pessoa (TORRETA, 2019).

Em relação à estratégia de comunicação da campanha de Jair Bolsonaro nas eleições de 2018, Torreta (2019) expressou a seguinte opinião:

Quando você tem uma estratégia de comunicação como eles tiverem, você precisa de um volume de gente para trabalhar, porque muita coisa é feita na mão. O *WhatsApp* é fechado. O *WhatsApp* tem uma atuação pesada no Brasil e na Índia. Até apareceu lá fora o quanto essa campanha do Bolsonaro foi inovadora, pois foi a primeira campanha no *WhatsApp* no mundo. Quanto ao uso de dados não houve nenhum. Teve segmentação, mas não através de dados. O *WhatsApp* é fechado. Você não sabe o que foi compartilhado. E o *Facebook* não entrega dados da pessoa, no máximo o telefone. Você segmenta via grupos, esse grupo é só de militares, esse grupo é só de direita. A sofisticação vai até aí. Não teve nenhum aprendizado de máquina (TORRETA, 2019).

Ao ser questionado como havia chegado até a *Cambridge Analytica*, Torreta (2019) comentou que saiu do marketing político em 2008 e voltou no final de 2016 quando recebeu o convite da Cambridge:

Eu não cheguei, eles chegaram até mim. Apareceu um cara no meu escritório e ofereceu uma proposta. Ninguém acredita, mas essa é a verdade. Eu não saio daqui para procurar ninguém. Eu tenho preguiça, eu sou baiano (TORRETA, 2019).

Sobre a atuação da mídia de massa não ter tido tanta influência nesta última eleição como ocorreu em campanhas anteriores, Torreta (2019) ponderou:

A gente ainda não pode dizer que acabou. Porque essa campanha foi atípica. Então assim, o *WhatsApp* provavelmente continua; a questão é que a imprensa tradicional não morreu (TORRETA, 2019).

Ao ser abordado sobre a demanda por gestão de imagem do candidato, Torreta (2019) considera que essa busca está em franca expansão. Quanto ao monitoramento ele ponderou que atualmente não é satisfatório, pois não alcança amplamente mídias como *Facebook* e *Instagram*, apenas o *Twitter*. O uso de dados do *Facebook* ficou mais restrito, mas segundo ele, não ocorreu em função do escândalo com a *Cambridge Analytica*:

⁴⁰ Bradesco Inteligência Artificial (BIA)

O escândalo da *Cambridge* não mudou em nada o mercado, porque ela não estava fazendo nada ilegal, não tem ninguém preso. O que aconteceu foi mais uma briga entre eles. Não tinha nada de ilegal. Tanto que não foi roubado nenhum dado do *Facebook*. Você pode não gostar, achar que isso é manipulação, mas ilegal não é; aí há uma outra discussão, ética (TORRETA, 2019).

Torreta (2019) fez esse questionamento: “O que você tem do mercado político?” E apresentou os seguintes argumentos:

Você teve uma campanha eleitoral muito violenta e ela não passou despercebida. Essa campanha presidencial serviu para demonstrar que Bolsonaro estava trabalhando há quatro anos. As campanhas no Brasil, o senado e a câmara foi apertando para que a eleição durasse um mês. Isso é um absurdo, um mês não dá para conhecer ninguém. Outras eleições duravam de 45 dias a dois meses. Aí começaram a encurtar as eleições para que os novos não pudessem entrar. Por que se você é novo, você precisa de tempo para ser conhecido. Quanto mais você encurta o tempo, mais os velhos ficam (TORRETA, 2019).

Para Torreta (2019), no Brasil, faltam pessoas que saibam ler dados. Segundo ele, a *Cambridge* o procurou, justamente por ele ter experiência com pesquisa qualitativa e por saber ler dados. De acordo com ele, a *Cambridge* não perguntou se ele sabia sobre tecnologia, pois o que eles queriam era alguém que soubesse ler os dados brasileiros, uma vez que a *Cambridge* não tinha expertise para fazer isso. Torreta (2019) chama isso de narrativa cultural, em que alguém só consegue fazer um trabalho de leitura de dados se compreender a narrativa cultural. Quanto à parte tecnológica, explicou o seguinte:

Quando a *Cambridge* me chamou, eu não era de sistema, não sabia nada disso. Eu lia as pesquisas, que é ler dados. Então fui entender o que era este mundo que eu não conhecia. A *Cambridge* me disse: “vá conversar com cientistas de dados para que você tenha uma ideia, mas sem ser técnico”. Eu não sou um técnico no assunto e eu acho que falta tudo. (TORRETA, 2019).

Para Torreta (2019), quem inovou no desenvolvimento tecnológico foi o Obama e não a *Cambridge Analytica* e nem a campanha do Trump. Segundo ele, a inovação da *Cambridge* e da Campanha de Trump não teve nada relacionado com tecnologia. Para Torreta (2019) “o que eles fizeram de inovador foi colocar a metodologia de comportamento nessa tecnologia utilizada na campanha do Obama”, e acrescenta:

A *Cambridge* não tinha um *software* que ninguém já não tivesse, ou alguma outra dessas tecnologias disruptivas do Vale do Silício. As pessoas que trabalharam para o Obama 90% delas eram de

tecnologia. Já as pessoas que trabalharam para o Trump eram pessoas de comunicação e de marketing. (TORRETA, 2019).

De acordo com Torreta (2019), esse tipo de inovação está ainda começando no Brasil. Segundo ele, “a parte comportamental”, como foi feita pela *Cambridge*, “inexistiu no país”. Para ele, os responsáveis pelas campanhas eleitorais no Brasil não usaram esse tipo de método, “nem sabem do que se trata”. Já nos Estados Unidos, o *microtargeting* comportamental é uma realidade. Ele acredita que, nos Estados Unidos começou-se a “pensar comportamentalmente na campanha presidencial de Bill Clinton, em 1992”, e desde então isso foi evoluindo até chegar ao *microtargeting* comportamental. No segmento privado, Torreta mencionou que “várias empresas já pensam comportamentalmente, até mesmo no Brasil” (TORRETA, 2019).

6.3.4 O que sua marca pode aprender com as Campanhas de Trump e Bolsonaro?

Em maio de 2019, Torreta apresentou o painel “O que sua marca pode aprender com as campanhas de Trump e Bolsonaro?” durante o evento ProXXima 2019, promovido pelo jornal Meio & Mensagem.

Para Monteiro (2019), a definição de novas regras promovida pela Justiça Eleitoral em 2018, como a autorização para impulsionamento pago nas redes sociais, permitiu ao marketing político novas estratégias para as campanhas eleitorais no Brasil.

Segundo Monteiro (2019) levantou, Torreta considera que Donald Trump, atual presidente dos Estados Unidos, e Jair Bolsonaro, presidente do Brasil se beneficiaram com o uso dessas novas estratégias:

Esses dois caras inventaram coisas que nenhum profissional de marketing estava vendo em termos de tecnologia e marketing cultural. A internet é um espaço que tem mais de dois bilhões de pessoas, então é impossível não existir política (TORRETA, apud MONTEIRO 2019, *online*).

De acordo com Monteiro (2019), Torreta apontou que ambos os presidentes compreenderam que a política atual requer saber falar com o público utilizando uma linguagem simples que atinja a todos, “inclusive não letrados”, e estar em contato com o “público-alvo de forma constante”, principalmente nas redes sociais (MONTEIRO, 2019, *online*).

O Brasil é um País ultraconectado. Nós usamos o *WhatsApp* para tudo. O brasileiro entendeu que é possível fazer negócio e campanha por meio de um aplicativo gratuito. (TORRETA, apud MONTEIRO 2019, *online*).

Monteiro (2019) cita que esse movimento, no entanto, teve início antes da eleição de 2018 e que Torreta considera que no Brasil “o movimento político ganhou força por meio das redes sociais em 2014 com o Movimento Vem Pra Rua” (MONTEIRO, 2019, *online*).

Em seguida, o Movimento Brasil Livre (MBL) começou a comprar domínios para ganhar alcance. Enquanto isso, a esquerda saiu dos jornalzinhos e foi para os *blogs*, mas analfabetos funcionais não leem *blogs*, assistem a vídeos no *YouTube*. Só agora a esquerda percebeu isso, enquanto a direita está usando esses recursos faz tempo (TORRETTA, apud MONTEIRO 2019, *online*).

De acordo com Monteiro (2019) Torreta citou seis aprendizados da eleição de Bolsonaro:

Manter frequência de contato com o público através das redes sociais; entender o País e cultura em que vivemos; usar linguagem acessível e direta; usar *big data* para compreender o público, como fez a *Cambridge Analytica* na eleição de Trump ao esmiuçar, por exemplo, que fãs de Lady Gaga não gostavam de Trump e consumidores da marca de *jeans Wrangler* eram de maioria brancos e simpatizantes do porte de armas; segmentar a comunicação; e aprender que público e privado não existe mais (TORRETTA, apud MONTEIRO 2019, *online*).

6.4 Empresas/Startups

Este item contempla as principais informações coletadas por meio do levantamento feito com as empresas/*startups*: Polis Consulting, Tail, Elife, Stilingue e MarketData.

O levantamento feito com a Polis e a Tail aconteceu por meio de entrevista presencial. O contato com a Elife transcorreu em duas etapas, primeiro por e-mail e depois por telefone. A coleta de informações com a Stilingue ocorreu via *WhatsApp* e pela análise de matérias veiculadas na mídia sobre essa *startup*. O contato com a MarketData se deu por meio de participação em uma palestra.

6.4.1 Polis Consulting

Em relação à coleta de dados, Kohlitz (2019) comentou que as soluções revendidas pela Polis Consulting consideram somente dados públicos. De acordo com ele:

Hoje a API do *Facebook*, é um canal restrito, quase nenhuma informação é pública e toda informação é anônima. Ou seja, hoje eu não vejo como você pode abusar de dados do *Facebook*, *Instagram*. O *Twitter* já é diferente, *Twitter* por definição é mais aberto e, neste caso, eu sei o gênero, o sexo, de onde vem, mas quem participa do *Twitter* sabe disso (KOHLITZ, 2019).

Ao ser questionado se o caso da *Cambridge Analytica* teve alguma influência nesse fechamento de dados do *Facebook*, ele acredita que não. Kohlitz (2019) comentou que acompanhou, de muito perto, o caso da *Cambridge Analytica*. Sobre o fato de a *Cambridge* ter realmente todo o poder por ela divulgado, Kohlitz (2019) argumentou o seguinte:

Eu acho que foi um aproveitamento político. De um lado, na Europa, se usava isso porque supostamente a *Cambridge Analytica* teria assessorado o pessoal do Brexit e do outro lado, nos Estados Unidos, o pessoal da *Cambridge Analytica* teria assessorado o Trump, que é da direita. Como a imprensa lá é da esquerda e a maior parte da opinião pública é da esquerda, então foi isso que “rolou”. O que aconteceu com a *Cambridge Analytica*, na verdade, foi que houve um aplicativo que usou dados do *Facebook*, do pessoal que aderiu, que baixou esse aplicativo, e eles revenderam esses dados. Isso era uma “zona cinza”, não era proibido e nem permitido, mas como o pessoal da *Cambridge Analytica* polarizou bastante e como ajudaram o Trump e o Brexit, eles pagaram um preço (KOHLITZ, 2019).

Quanto ao uso da Internet e redes sociais em campanhas eleitorais Kohlitz (2019) ponderou o seguinte:

Hoje o que acontece, você tem mais de 50% da população brasileira com acesso à Internet por meio do *mobile device*; há 10 anos, era só 5%, ou seja, só a classe A tinha um celular, um *smartphone*, iPhone 3... Mas, não é a elite de São Paulo que decide as eleições, é a classe C, e hoje como todo mundo acessa a Internet por meio do celular isso virou uma ferramenta muito potente (KOHLITZ, 2019).

De acordo com Kohlitz (2019), as ferramentas revendidas pela Polis monitoram não só as redes sociais, mas tudo que está disponível publicamente na Internet. Kohlitz (2019) citou as seguintes ferramentas: *Crimson Hexago*, *Brandwatch*, *v-tracker*, *TrackMaven* e *Netbase*. Ao ser questionado sobre quais são os algoritmos que essas ferramentas utilizam, Kohlitz respondeu o seguinte:

O NLP (*Natural Language Process*) é utilizado dentro do *NetBase*, *V-tracker*, *Crimson Hexago* e *Brandwach* para fazer análise semântica, ou seja, iPhone é bom? Se você consultar o dicionário Aurélio bom é positivo. iPhone nunca foi bom, tem uma palavra negativa tem uma positiva, e agora? iPhone nunca foi tão bom quanto agora. Sem NLP, se você simplesmente olhar o que é positivo ou que é negativo, não

consegue analisar a pluralidade de uma frase, é impossível (KOHLITZ, 2019).

Quanto ao nível de assertividade dessa análise semântica ser alto, Kohlitz explicou o seguinte:

Essa é uma pergunta que todo mundo quer saber, o nível de assertividade. Isso não é uma pergunta boa, por quê? Depende. Vou te dar alguns exemplos: ironia e sarcasmo, esquece, nem o ser humano sabe; tenta monitorar um jogo de futebol entre Palmeiras e Corinthians, se você não estiver por dentro do assunto, você não sabe se é xingamento, se é elogio, se é isso ou se é aquilo. Se você não sabe, como a “coitada” da ferramenta vai saber? Não sabe. Obviamente, você consegue depois no processo de treinamento do algoritmo fazer uma calibragem para ajustar esses pontos. Por exemplo, aqui no Brasil, há 2, 3 anos, surgiu o termo sqn (só que não), ou seja, a pessoa fala uma coisa positiva e acrescenta sqn. Isso acaba mudando a polaridade para negativo, muito fácil implementar no algoritmo. Por isso, para cada cliente, cada marca tem regras linguísticas que precisam ser implementadas. Ou seja, o NLP já tem, na verdade, certa assertividade de base, mas depois você precisa adaptar a cada caso (KOHLITZ, 2019).

Kohlitz (2019) foi questionado se as ferramentas revendidas pela Polis utilizam aprendizado de máquina (*machine learning*) e a resposta foi afirmativa. Quanto ao fato de muitas empresas divulgarem que utilizam Inteligência Artificial destacando isso como algo muito inovador, ele faz algumas considerações:

É preciso analisar e olhar isso muito detalhadamente. A IA utilizada para análise de dados, nada mais é do que modelos estatísticos e deterministas, e esses algoritmos já existem há mais de 50 anos, ou seja, não são novos. Algoritmos de redes neurais também não são novidade. Mas, agora isso virou a onda, a *buzz* do momento. Tudo usa Inteligência Artificial, a gente respira Inteligência Artificial. Mas, o que é Inteligência Artificial? Esse termo é muito abrangente, diz tudo e nada ao mesmo tempo, e como não é bem definido, todo mundo pode dizer que faz e cobrar mais caro por isso. E o pior é que ninguém analisa, deixa eu ver exatamente o que você está fazendo. É muito mais fácil ficar na superfície e dizer Inteligência Artificial é legal (KOHLITZ, 2019).

Kohlitz (2019) concordou que o uso do aprendizado de máquina é um recurso estratégico em uma campanha publicitária, e citou alguns exemplos:

Tem ferramentas nacionais de monitoramento que não têm nenhuma inteligência atrelada. O que acontece? Você precisa de uma sala de 20 pessoas, fazendo *tagueamento* manual, para acrescentar inteligência ao monitoramento. Isso gera dois problemas, primeiramente o erro humano, erro de interpretação, mesmo se você definir regras bem claras, tem “n” casos que mostram que 5 pessoas, tem 5 opiniões diferentes. Ou simplesmente o erro humano é por falta de atenção. De outro lado, você não consegue escalar, tenta analisar o que nós fazemos durante as eleições, durante os debates,

onde você tem um *buzz* de 20 milhões? Eu tenho um projeto, faço o dimensionamento da mão de obra, só que aconteceu uma crise, e o *buzz* triplica, o que eu faço? Eu vou à Avenida Paulista contratar 3 vezes mais gente para poder monitorar? Ok, mas, de novo, tenta analisar 20 milhões, 40 milhões. Impossível fazer isso manualmente, isso sem falar do custo da mão de obra e dos erros humanos que já comentei, ou seja, você precisa de inteligência para poder analisar dados em grandes volumes. Isso é óbvio, e hoje em dia você tem cada vez mais dados e precisa cada vez mais de inteligência para tomar decisões inteligentes (KOHLITZ, 2019).

De acordo com Kohlitz (2019), numa campanha eleitoral os políticos querem saber os seguintes dados:

Quem são os brancos, nulos e os indecisos, para fazer um corpo a corpo e tentar trazer esses eleitores para seu lado. Político quer saber se o programa que está defendendo se adequa bem quando a maior parte dos cidadãos está dentro do distrito dele, ou seja, é esse tipo de coisa que um político quer saber. Neste cenário, o Obama foi o pioneiro e umas das ferramentas que ele usou para fazer isso foi a *Netbase* que a Polis tem em seu portfólio (KOHLITZ, 2019).

Sobre exemplo de um case de campanha eleitoral, Kohlitz (2019) considera que atualmente o maior case é o da campanha do Bolsonaro que utilizou o *WhatsApp*. Ao ser questionado se a Polis teve contato com Bolsonaro, Kohlitz (2019) disse que não. Ele destaca que a campanha do Bolsonaro foi simples e eficaz, sobretudo levando em conta que Bolsonaro, a princípio, não tinha muita chance, principalmente por não dispor de muito espaço na mídia de massa:

Mas, você viu o resultado e o terremoto que isso aí deu, em termos de tecnologia e tudo mais. Isso vai render cases em Stanford e Harvard para os próximos 20 anos (KOHLITZ, 2019).

Ao ser questionado se a tecnologia tem um papel importante em uma campanha política, Kohlitz (2019) disse que não resta a menor dúvida e deu o seguinte exemplo:

Com Mitt Romney e Obama, o que aconteceu? O Romney não usava tecnologia, ele fez a campanha pela base dos republicanos que ele conhecia. Já o Obama fez totalmente diferente: “os democratas que vão votar em mim, OK! Agora o que eu quero saber é quem são os indecisos, são esses que eu vou tentar trazer para o meu lado”. Foi isso que decidiu as eleições, e isso foi possível por meio da tecnologia. Ou seja, já no segundo mandato do Obama, a tecnologia teve um papel muito importante (KOHLITZ, 2019).

Em relação ao monitoramento que a Polis realiza nas eleições, Kohlitz (2019) citou um case de monitoramento feito na campanha presidencial de 2014:

Antes do segundo turno, havia saído uma pesquisa do IBOPE indicando que Marina Silva iria para segundo turno. Aí você vê

porque o *big data* hoje em dia é cada vez mais importante. O IBOPE se enganou em 6 pontos percentuais e 6 pontos é muita coisa. Pelo monitoramento que a Polis havia feito na época, a gente já sabia que ela não chegaria ao segundo turno. E como a Polis conseguiu fazer isso? Porque em vez de ter um estagiário na Paulista e outro sei lá aonde captando dados na mão, fazendo pesquisa em campo, a Polis captou em cima do Brasil inteiro *online* informações que mostraram que ela não chegaria ao segundo turno (KOHLITZ, 2019).

Ao ser questionado se nessas eleições de 2018, o fato do candidato Alckmin não ter ido para segundo turno se constituiu uma grande surpresa, Kohlitz (2019) prefere falar sobre assunto daqui a 4 anos. Ele acredita que ainda é cedo para tirar conclusões, uma vez que essa campanha de 2018 foi atípica. Para Kohlitz, as eleições de 2018 foram:

Um “*black swan event*”⁴¹ porque a campanha do Bolsonaro cruzou de uma maneira extremamente inovadora a tecnologia e foi uma mudança de paradigma “*black swan*”. Alguma coisa aconteceu e ninguém, “*nobody saw coming*” (KOHLITZ, 2019).

Indagado se no monitoramento das eleições de 2018 a Polis não conseguiu prever isso? Kohlitz respondeu que não foi possível e fez alguns comentários a respeito:

A gente não trabalha com *WhatsApp*, esses dados móveis, é “zona cinza”. A Polis trabalha com coisas totalmente sem discussão que sejam aceitas e lícitas. Essa questão de correr atrás de dados de telefonia móvel é um pouco crítica. A Polis não mexe com isso. Nossas ferramentas não lidam com isso. Mas, nessa campanha circularam essas listas de pessoas, como conseguiram isso? (KOHLITZ, 2019).

6.4.2 Tail

Lopes (2019) explicou que a Tail é classificada como *startup*, porém é uma *startup* bem estruturada que atende grandes clientes, como LG, Tim, Unilever e vários bancos. Ao ser questionada se a Tail atua com marketing político, Lopes (2019) responde o seguinte:

A Tail não trabalha com marketing político. Inclusive, o João Amoedo procurou a Tail na época da eleição, mas a resposta foi não, porque é a política dos donos da empresa de não se envolver em política (LOPES, 2019).

⁴¹ É um evento ou ocorrência que se desvia do que normalmente é esperado em uma situação sendo muito difícil de prever, pois são eventos aleatórios e inesperados.

De acordo com Lopes (2019), a Tail atua com duas grandes áreas internas que foram descritas da seguinte forma:

Uma área de tecnologia, denominada engenharia e uma área de serviços e negócios. A engenharia é responsável pelo desenvolvimento do *backend* e do *frontend* dos produtos, além da área de ciência de dados, que cuida dos robôs e dos algoritmos de *machine learning*. A Tail, portanto, tem uma tecnologia própria. Na área de serviços e negócios estão os líderes de negócios, toda a parte de operações, venda, atendimento ao cliente, gerenciamento de projeto e os arquitetos de soluções que são as pessoas que entendem do negócio do cliente e fazem a ponte com a equipe técnica (LOPES, 2019).

Sobre como é feita a coleta dos dados utilizados pela Tail, Lopes (2019) explicou:

Hoje, a Tail captura mais de 390 milhões de *devices* únicos da internet brasileira. A gente não fala de usuários únicos porque, eu Lopes, por exemplo, posso ter três *devices*, o computador de casa, o computador do trabalho e o celular. A gente faz a coleta por meio de uma *tag* que fica dentro de vários sites da rede de parceiros. Então hoje, a Tail, tem 90% da internet brasileira mapeada. Temos parceiros como UOL, Terra, Globo e alguns sites menores. Os dados são capturados pelo *cookie* da máquina. Hoje se você instala uma *tag*, ela vai capturando o que aquele *cookie* está fazendo. O que muda com a GDPR e com as leis de proteção de dados, é o consentimento. Então, todo site que tiver uma *tag* da Tail tem que colocar: “nós usamos coleta de *cookies* para entender perfil de comportamento e melhorar sua navegação”. A Tail tem todo manual de boas práticas, inclusive estamos fazendo uma série de palestras com a assessoria jurídica para preparar os clientes para ficarem pontuais com a lei. A Tail sempre teve *compliance* e nunca fez nenhuma prática que não fosse permitida (LOPES, 2019).

Para facilitar o entendimento de como ocorre essa coleta dos dados Lopes deu o seguinte exemplo:

Você entrou no UOL e leu alguma notícia sobre esportes. Depois você entrou em outro site e viu um batom. A *tag* da Tail está em todos esses sites. O que a gente faz com isso? Nossos algoritmos são programados para identificar o conteúdo de uma página. Esses algoritmos usam processamento de linguagem natural para treinar o robô. E como o robô é treinado? Eles são treinados para conseguir identificar qual é o tema daquela matéria e o que a pessoa está lendo daquele conteúdo que ela está consumindo. Para isso, o robô é configurado com palavras-chaves e recebe vários treinos, por exemplo, me devolve mais de 10 matérias de esportes, para ver se esse robô está acertando o que é esporte. Somente os robôs que têm 98% de assertividade é que são considerados para rodar o algoritmo. É isso que diz o comportamento do consumidor e vai criando padrões de acordo com o que esses *cookies* têm interesse. Mas para considerar que um *cookie* tem interesse em esporte e colocá-lo na caixinha de esportes, a pessoa precisa ter passado, pelo

menos, três vezes na semana no conteúdo de esportes. Então, toda essa informação coletada faz com que a Tail tenha hoje mais de 2.000 segmentos comportamentais (LOPES, 2019).

Lopes deu um exemplo mais prático, da LG, para explicar melhor esse processo:

A LG brifa com a Tail e uma agência: Preciso falar com quem tem interesse em compra de televisão e moda. A Tail então, consegue levantar 5 milhões de pessoas, ou melhor 5 milhões de *cookies* que possuem esses interesses. Após a Tail ter levantado esses dados, a agência diz: quero usar isso na campanha do *Google*. A Tail tem uma integração com o *Google* para que esses dados cheguem até as plataformas de compra de mídia. Então, a agência vai comprar a mídia no *Google* com esses dados. A Tail, até esse momento, só conhece o *cookie* do dispositivo, ela não sabe quem é a pessoa, pois um *cookie* de dispositivo não identifica o usuário. Até porque, nesse exemplo, não é preciso saber quem é a pessoa, pois o que eles querem é fazer uma campanha para quem quer comprar televisão, não importando quem esteja comprando. Neste caso eles usam os dados da Tail apenas como segmentação para que um *banner* apareça toda vez que entrar um *cookie* que tenha interesse em comprar televisão e moda. Somente após a pessoa clicar, comprar, preencher um cadastro e confirmar essa operação é que é possível saber quem é a pessoa. Por isso que a Tail não infringe nenhuma lei de privacidade de dados (LOPES, 2019).

De acordo com Lopes (2019), além desse trabalho de segmentação, a Tail também atua com enriquecimento de dados para torná-los mais relevantes. Lopes (2019) detalhou esse processo:

No exemplo do caso de uso da LG, quando uma pessoa compra no site da LG e preenche um cadastro, essa pessoa dá consentimento, para a LG ter informações como CPF, *e-mail*, telefone, etc. Ou seja, quando a marca tem um CRM configurado e informações que o cliente autorizou que ela tivesse, eles mandam isso de forma criptografada para a plataforma da Tail encontrar o *cookie* dessas chaves. Mas, como isso é feito? Da seguinte forma: uma pessoa faz uma compra na Netshoes e informa o CPF. Depois essa pessoa compra uma TV da LG nas Casas Bahia e informa o CPF lá também. A LG tem o CPF dessa pessoa por causa da compra nas Casas Bahia. Quando essa pessoa fizer um *login* no site da Netshoes, e a base das Casas Bahia estiver na DMP da Tail, a Tail a partir desse momento vai conseguir identificar esse *cookie*. Mas isso não infringe nenhuma lei, porque a pessoa deu o consentimento. De posse da identificação do *cookie*, é possível enviar um *e-mail* marketing com promoções direcionadas para quem comprou produto LG. Para deixar isso ainda mais robusto, a Tail tem parceria com Serasa e com Boa Vista Serviços, SCPC, que fazem cruzamentos também para dizer como é a renda das pessoas. Toda essa integração é denominado de *audience center* que integra todo tipo de audiência e transforma isso é um dado relevante (LOPES, 2019).

Em relação se há uma evolução do uso de aprendizado de máquina no Brasil, Lopes (2019) acredita que sim. Ela tece algumas considerações sobre quando se começou a fazer uso dessa tecnologia e como está o cenário no momento:

Quando comecei a trabalhar com mídia digital, a gente comprava mídia assim: preciso falar com quem gosta de esportes. Então, vou ligar para a Globo e pedir para colocar um *banner* no Globo Esporte. Não existia uma compra com base em dados, apenas no conteúdo que se queria atingir. Como é que isso evolui para chegar até hoje? Teve um momento em que o *machine learning* começou a dizer que uma determinada pessoa gosta de esporte, mas não fica o dia inteiro no site do Globo Esporte, ela também olha UOL, *Facebook* e outros sites. A máquina consegue por meio dos *cookies* saber que essa pessoa gosta de esporte mesmo quando ela não está em páginas de esportes. Então foi aí que o negócio começou a evoluir. Mais ou menos, vou te dizer que em 2013 começou de fato a evoluir. Hoje a gente está em um nível que a Tail já pluga parceiro do *score* de crédito para dizer que esse *cookie* ganha mais de R\$ 10.000. Assim não preciso mostrar a televisão da LG mais barata, eu posso mostrar a mais cara que essa pessoa tem intenção de compra e vai comprar. Então, isso vem evoluindo por conta das integrações e vai evoluir mais ainda, porque agora, por exemplo, existem alguns estudos para colocar informações do IBGE e mesclar dados *online* com os dados *offline* (LOPES, 2019).

Lopes (2019) considerou que o caso da campanha do Bolsonaro foi emblemático, assim como o caso do Partido Novo, pois, segundo Lopes (2019), eles tiveram apenas 3 segundos de TV, e em ambos os casos o uso da internet foi primordial. Para Lopes (2019), a TV atualmente se tornou um ponto de apoio, ou seja, um complemento da Internet:

O grande ponto é a comunicação certa para a pessoa certa. Numa campanha de TV, por exemplo, o Bolsonaro diz que é de direita e quer implementar a direita no Brasil. Para quem é de direita, essa mensagem será suficiente, no entanto, existe aquela pessoa que não é direita nem esquerda, mas considera que o porte de arma é importante, então essa comunicação não vai funcionar para esta pessoa (LOPES, 2019).

Lopes (2019) considera importante pesquisar sobre as bolhas de conteúdo que são geradas pelos algoritmos:

Plataformas como *Netflix*, por exemplo, se você vê um filme de romance, ela vai cada vez mais te recomendar filmes desse gênero. Se eu vejo só ação ela vai cada vez mais me recomendar ação. No *Facebook* acontece o mesmo, porque a lógica do algoritmo utilizada para classificação do que é importante é muito parecida. Daí o que acontece, você tende a ficar só na sua zona de conforto. Então você deixa de ter acesso a aquelas coisas de massa, isso é um problema também (LOPES, 2019).

Ao ser indagada se a lei de proteção de dados pessoais ajuda nos casos em que a maioria das pessoas não possuem conhecimento a respeito do uso de dados pessoais pelos algoritmos, Lopes (2019) considera que a lei atual protege em termos e cita a seguinte situação: “Pense nos termos que as pessoas aceitam para criar um *e-mail* ou uma conta no *Facebook*, quem lê aquilo? Mas, está suportado juridicamente” (LOPES, 2019).

Para Lopes (2019), não tem como frear todo esse processo que está em andamento, mas é importante criar uma resistência:

Por exemplo, no caso das farmácias, toda vez que você vai comprar algo, eles pedem o CPF. Para que querem isso? Para usar dados de comportamento para vender mais. E aí quando as pessoas resistem o que eles fazem? Se você der o CPF vai ter desconto. A gente está falando de um país pobre que um real de desconto faz a diferença. Então precisaria ter um consentimento também em lojas físicas, mas a lei de proteção de dados só fala do digital (LOPES, 2019).

Ainda em relação aos dados pessoais, Lopes (2019) comentou que a maioria das empresas, incluindo a Tail, tem um conselho de ética que define alguns dados pessoais como sensíveis:

Preferência sexual, informações de saúde, religião e também de negatificação. A Serasa não é autorizada a mandar, por exemplo, uma base de pessoas negativadas. O mercado segue esse tipo de ética, não é assinado em pedra, mas as grandes e as boas empresas acabam adotando esse conceito de dados sensíveis (LOPES, 2019).

6.4.3 Elife

Ao ser questionado se o grande volume de informação gerado e compartilhado em ambientes digitais impõe novos olhares e estratégias de marketing, Lima (2019) respondeu o seguinte:

Sim, acreditamos que o volume de informações compartilhadas pode ajudar as marcas a compreender em profundidade o comportamento do consumidor, sua jornada de consumo e ainda auxiliar no treinamento de máquina na área de inteligência artificial. As próprias redes sociais estão usando estes dados para treinarem algoritmos que permitirão direcionar mídia de maneira mais precisa ao consumidor, além de antever com precisão seus passos na jornada de compra (LIMA, 2019).

Em relação à evolução do uso de *machine learning* no Brasil, Lima (2019) tem a seguinte opinião:

Nos últimos anos, a principal evolução foi a adoção por empresas e especialistas brasileiros de soluções de aprendizado de máquina disponíveis na nuvem. Empresas como *Google* e *Microsoft* criaram

aplicações rápidas e simples, disponíveis na nuvem, para que qualquer empresa possa criar aplicações que usem aprendizado de máquina. É o caso do *Google Dialog Flow*. Com este *software* nós da Elife, por exemplo, estamos criando *bots* e *bots* de voz de maneira muito mais rápida para nossos clientes. Esta evolução é recente. Há uma década os custos dessas ferramentas eram proibitivos (LIMA, 2019).

Sobre a importância estratégica do aprendizado de máquina para o marketing, Lima (2019) ponderou o seguinte:

O aprendizado de máquina ainda é utilizado como uma curiosidade dentro do marketing. As aplicações, muitas vezes, não estão inseridas ou não são criadas a partir de um problema real de negócios. Busca-se a perfeição da ficção científica e isso gera com frequência desapontamento por conta do cliente. Se usado de maneira a focar em problemas reais de negócios e de maneira a economizar recursos, o aprendizado de máquina pode automatizar processos repetitivos e ineficazes que hoje ocupam o tempo das pessoas. Não serão grandes saltos, mas melhorias em pequenos processos que podem trazer grandes economias de recursos. As empresas que possuem ou contratam um *call center* para atendimento ao consumidor ou para cobrança, por exemplo, podem mapear todos os seus processos mais frequentes e que ocupam a maior parte dos recursos humanos, e com a ajuda do aprendizado de máquina podem automatizar muita coisa. Várias áreas de fato podem ser beneficiadas (LIMA, 2019).

Para Lima (2019), a aplicação de técnicas de *microtargeting* e aprendizado de máquina numa campanha publicitária vem evoluindo e faz o seguinte comentário:

O *microtargeting* é uma técnica que está evoluindo e que serve para aperfeiçoar a experiência do consumidor com personalização, automação do relacionamento e automação de conteúdo, de acordo com um histórico prévio de interações do consumidor. Creio que mais e mais ferramentas irão oferecer automação destes aspectos a partir de uma base treinada sobre o comportamento prévio do consumidor. Quanto mais rastros de dados o consumidor deixa, mais os algoritmos conseguirão automatizar não apenas mídia, mas também conteúdo (LIMA, 2019).

Sobre as plataformas de aprendizado de máquina utilizadas nas campanhas publicitárias e no marketing, no Brasil, Lima (2019) cita que “o próprio *Facebook* e o *Google* utilizam aprendizado de máquina para sugerir opções de segmentação de mídia” (LIMA, 2019). Quanto às inovações que o aprendizado de máquina pode trazer para o marketing, Lima (2019), observa:

Creio que a inovação se dará pouco a pouco. A mais importante será a capacidade de resolver, em curto prazo, problemas simples que hoje custam milhares de reais para serem solucionados. Exemplo: o atendimento ao consumidor através de um *call center*. As empresas devem começar a atacar exatamente os processos ineficientes e que

comprometem grande parte do orçamento. Creio que já estamos vendo inovações como *bots* e *voice bots* de atendimento. Na Elife, por exemplo, estamos apostando em um SaaS⁴² para criar robôs de voz que poderão fazer cobranças, agendamento de consultas e confirmação da participação em eventos. Qualquer pequena ou média empresa poderá sozinha criar seu próprio robô *online*, economizando com o trabalho repetitivo e com isso alocando os recursos humanos para tarefas mais complexas (LIMA, 2019).

Sobre o uso de aprendizado de máquina nas campanhas eleitorais, para Lima (2019) o escândalo da *Cambridge Analytica* mostrou o quanto à tecnologia pode ter influenciado as eleições dos Estados do Unidos, o *impeachment* da Dilma e a eleição presidencial no Brasil em 2018. De acordo com Lima (2019), no Brasil o uso de aprendizado de máquina ainda está em fase embrionária. Mas, a análise dos dados para definição de perfis e a divulgação direcionada de mensagens podem ser utilizados em campanhas eleitorais, independente do uso do *machine learning*.

De acordo com Lima (2019), as campanhas políticas foram as mais rápidas em utilizar análise de dados com essa finalidade do que as marcas e campanhas publicitárias. Segundo ele, as marcas possuem uma ética mais consistente do que o campo político que é já é mais fechado e restrito.

Lima (2019) chamou a atenção para aqueles testes e brincadeiras que estão disponíveis no *Facebook* e que usam o *social login*, que é o *login* que o usuário se *loga* com o *Facebook* em vários aplicativos. Para Lima (2019), antes do escândalo da *Cambridge Analytica*, era permitido que esses aplicativos acessassem os dados dos usuários compartilhados no *Facebook* (nome, data de nascimento, *e-mail*, profissão, escolaridade) além de coleta dos *likes*. Após o escândalo da *Cambridge Analytica*, o *Facebook* passou a ter um controle maior em relação aos aplicativos de terceiros que usam a plataforma, restringindo esse tipo de coleta.

Lima (2019) também explicou que uma vez que os dados são coletados, pode ser feita uma clusterização que é uma técnica antiga que não requer necessariamente aprendizado de máquina. Segundo ele, a clusterização é semelhante à técnica da análise da cesta de compras (*basket*), e deu o seguinte exemplo:

Com a técnica de análise da cesta de compras, pode-se verificar que um homem que compra cerveja e fralda, portanto, é pai e pode direcionar para esta pessoa outras opções de compra que tenham

⁴² *Software as an Service ou Software como serviço.*

relação com o fato dele ser pai. Essa mesma técnica pode ser utilizada na política para saber se um grupo é a favor ou não de um determinado político e, com base nessas definições, produzir conteúdo específico para este grupo, pois se sabe que dessa forma o conteúdo terá mais alcance.

De acordo com Lima (2019), os dados não são tão importantes para o uso de aprendizado de máquina na política, pois uma campanha política pode utilizar as informações do próprio candidato, o seu plano de governo e treinar *bots* para dialogar com os eleitores, abrir um canal direto para responder perguntas e obter *insights* dos próprios eleitores, o que diminuiria o custo da campanha. Segundo Lima (2019), o uso aprendido de máquina na política será uma prática cada vez mais frequente.

Lima (2019) considera que o aprendizado de máquina possa ser utilizado para fazer gestão de imagem do candidato:

Treina-se o algoritmo para reconhecer o candidato e verificar quem mais aparece. Pode ser uma informação para ajudar no direcionamento da campanha e também para entender o cenário. É importante pensar em aplicações simples que vão resolver o problema. Por exemplo, verificar todos os *posts* do candidato no *Facebook* e descobrir que o candidato B tem mais eleitoras mulheres, isso é uma informação importante. Para as empresas, o foco é diminuir custos, e na política criar novos canais de relacionamento e entender o cenário político. E, neste contexto, entender a imagem do político na campanha com os eleitores ajuda bastante (LIMA, 2019).

6.4.4 Stilingue

6.4.4.1 Matérias veiculadas sobre a Stilingue

De acordo com Câmara (2018), a Stilingue começou em 2014, na Campus Party, um importante evento de tecnologia, quando dois cientistas da computação mineiros que apresentavam um aplicativo para resumo de notícias se encontraram com Rodrigo Helcer, atual CEO da Stilingue. Nesse encontro eles identificaram que faltava uma aplicação de Inteligência Artificial para o idioma brasileiro. Para resolver esse problema decidiram criar a Stilingue.

Segundo Stocco (2016), o *software* da Stilingue resume automaticamente textos, faz sugestão de *insights* para marcas ou campanhas, “compara influenciadores por tópico de afinidade e cruza resultados da mídia tradicional com a social” (STOCCO, 2016, *online*).

Com a tecnologia da Stilingue, pode-se reduzir até 50% do tempo operacional de times de *Business Intelligence* (BI) e fortalecê-los a interpretar um volume de dados impossível sem o poder da computação cognitiva (HELCER apud STOCCO, 2016).

Segundo Renner (2017), João Dória utilizou a tecnologia de processamento de linguagem natural da Stilingue para avaliar como influenciar eleitores na internet. De acordo com Mota (2017a), a influência digital de João Dória começou a ser validada nas prévias do partido PSDB para o governo de São Paulo que ocorreu em 18 de março de 2018, e cada vez mais foi se aperfeiçoando com o uso de ferramentas tecnológicas que combina *softwares* para monitoramento, além de uma equipe de análise que “avalia o impacto nas redes sociais de tudo o que ele fala” (MOTA, 2017a, *online*). Para Daniel Braga, que acompanha Dória desde agosto de 2015:

Nossa função é alinhar o discurso do João para ele ser bem entendido. Ele não vai mudar o que pensa, as propostas, mas vai falar da melhor forma (BRAGA, apud MOTA, 2017a, *online*).

Mota (2017a) levantou que a Social QI, empresa criada por Daniel Braga com outros sócios, opera com cinco *softwares* de monitoramento, cujos nomes não foram revelados. Em paralelo, utiliza o *software* da Stilingue.

A tecnologia permite o escrutínio das redes sociais *Facebook*, *Twitter*, *Instagram*, de influenciadores e de tudo o que é publicado na imprensa. Um volume exponencial de informações que dificilmente seria administrável sem a ajuda de um *software* (HELCER, apud MOTA, 2017a, *online*).

Nas campanhas políticas, Mota (2017a) verificou que este tipo de tecnologia está sendo utilizado para fazer a gestão de imagem de um candidato, uma vez que o *software* “consegue capturar tudo o que é escrito sobre qualquer assunto”, além de fazer reconhecimento de imagem para “identificar memes” e também obter dados para traçar psicometria. Segundo Mota (2017a), a “psicometria faz uma espécie de análise de personalidade dos eleitores”, além de identificar perfis que “vão muito além de direita e esquerda” e com isso ajudar a formular e direcionar o discurso político de forma personalizada.

Tudo o que o João Dória fala é monitorado. A ideia é avaliar como o discurso do prefeito é recebido para reduzir os impactos negativos e potencializar os positivos. Se ele fala sobre desestatização, nós checamos o que as pessoas falam sobre isso para vermos qual a melhor forma de elas absorverem o discurso (BRAGA apud MOTA, 2017a, *online*).

De acordo com Mota (2017a), Braga comentou que voltou a cuidar pessoalmente da campanha de João Dória, após um pedido do próprio Dória, e ressaltou que não presta serviços somente para o partido PSDB, a empresa tem sido "assedada por muita gente". Sobre a campanha de Dória, Braga disse o seguinte: "A gente tem que entregar o João Trabalhador que prometeu" (BRAGA, apud MOTA, 2017a, *online*).

Para Mota (2017b), a influência digital na disputa política se destacou nas eleições de 2014 e conquista mais espaço nas eleições de 2018 com a opção de soluções que "varrem" a Internet com *softwares* de Inteligência Artificial para mapeamento dos perfis de eleitores e direcionamento de discurso personalizado. De acordo com Mota (2017b), a prática de adaptar o discurso dos candidatos para elevar o alcance e conseguir votos é uma estratégia antiga dos consultores de marketing. "Pessoas com orientação progressista dificilmente dão atenção a discursos autoritários". Por isso é necessário adaptar o discurso dos candidatos. Atualmente, a grande vantagem desse tipo de tecnologia é:

Ir além da divisão demográfica e ideológica, da direita e esquerda, e agrupar os brasileiros usando como critérios seus sentimentos, medos, desejos e ambições (MOTA, 2017b, *online*).

Mota (2017b) ressalta que, na política, essa "vigilância robotizada" não é utilizada somente para fazer a gestão de reputação do candidato e analisar personalidade dos eleitores para formulação e direcionamento de discurso político personalizado. Ela também é capaz de "identificar aqueles que são a favor e contra determinado tema" e ajudar os candidatos a fazer um "corpo a corpo" virtual para procurar "convencer os indecisos" (MOTA, 2017b, *online*).

Segundo Rodrigo Helcer, CEO da Stilingue, essa empresa tem dois clientes pré-candidatos a cargos do Executivo, que contrataram o "pacote completo", porém, ele não mencionou mais detalhes (HELKER, apud MOTA, 2017b, *online*). Além da indústria de bens de consumo e das assessorias de Comunicação, Mota (2017b) cita que a Stilingue também monitora as votações no Congresso.

6.4.4.2 Principais pontos da Entrevista

Helcer (2019) revelou que entre as tecnologias adotadas pela Stilingue estão: *machine learning*, *deep learning* e *natural language processing* para o idioma brasileiro e linguagem informal. Em relação ao uso dessas tecnologias aplicadas ao

marketing e às campanhas publicitárias, ele explicou que elas podem ser utilizadas para categorizações automáticas, sumarização textual, *clustering* de audiências e narrativas.

Sobre a evolução do aprendizado de máquina no Brasil, Helcer (2019) considera que teve início em 2017, porém, é mais recentemente que o assunto tem ganhado atenção. Para Helcer (2019), de forma geral, ainda estamos no *hype* do tema e sem muitos profissionais com experiência no assunto. Em relação à especialidade desses profissionais, Helce (2019), cita que eles são: cientistas de dados; taxonomistas; linguistas; e especialistas no domínio de negócio.

Helcer (2019) mencionou o seguinte caso de sucesso referente ao uso de aprendizado de máquina no marketing e/ou em campanhas publicitárias:

No meio de uma das maiores crises da história, conseguimos antecipar assuntos críticos antes de chegar à imprensa. Otimizamos investimentos publicitários com base na opinião espontânea em torno de campanhas. Descobrimos novos segmentos de mercado que nunca se destacaram em pesquisas por métodos tradicionais. Apoiamos na priorização e auditoria de influenciadores digitais (HELCER, 2019).

Ao ser questionado se o uso de aprendizado de máquina no marketing e em campanhas publicitárias é um diferencial competitivo, Helcer pondera que “sem o apoio do computador, o profissional de marketing e comunicação perderá seguramente competitividade” (HELCER, 2019). Para ele, o volume de informação gerado e compartilhado em ambientes digitais impõe novos olhares e estratégias, pois “nunca houve acesso a tanta informação e em tamanha velocidade (*real time*) como hoje” (HELCER, 2019).

Sobre as inovações que o aprendizado de máquina pode trazer, Helcer analisa:

É como o microscópio. Antes da sua existência, não conhecíamos os efeitos de bactérias e micro-organismos que mudavam para o bem ou mal nossa saúde. Com *machine learning* e *natural language processing*, conseguimos enxergar em torno das marcas comportamentos impossíveis de se observar a “olho nu” (HELCER, 2019).

Helcer (2019) explicou que os dados utilizados pelos algoritmos de aprendizado de máquina da Stilingue são obtidos principalmente via APIs abertas. Quanto à privacidade dos dados, regulamentações e boas práticas Helcer (2019) menciona que respeita as políticas de cada API/Rede social.

6.4.5 MarketData

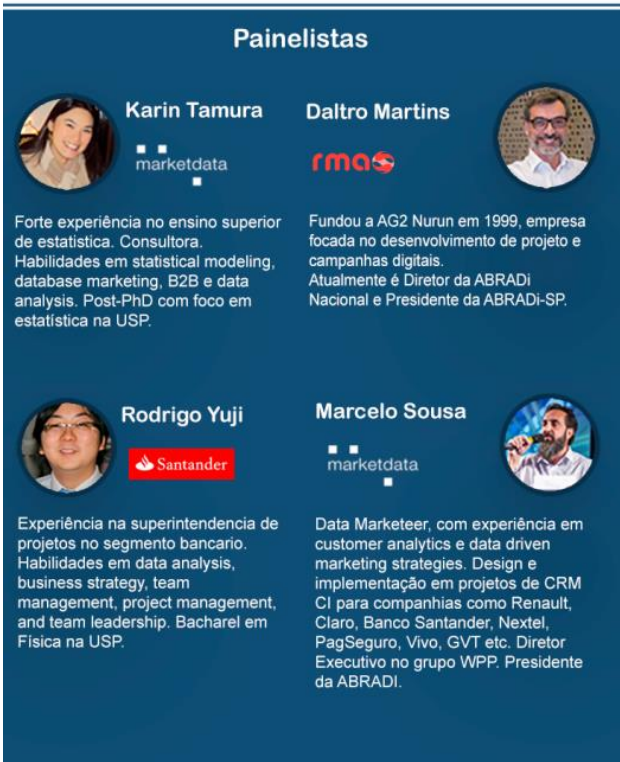
A palestra “O uso da informação como vantagem competitiva” aconteceu no dia 12 de fevereiro de 2019, no Campus *Google*, em São Paulo. Ela foi apresentada por Marcelo Sousa, da MarkeData.

Figura 25 - Palestra “O uso da informação como vantagem competitiva”

Painel “Como transformar a sua empresa a partir de dados”

Esse será o tema do nosso Painel com Guido Sarti, da GloboSat, Daltro Martins, da RMA, Karin Tamura, da MarketData e Rodrigo Yuji, do Santander Holdings.

O uso da informação como vantagem competitiva
Assista a palestra sobre O uso da informação, com Marcelo Sousa



Painelistas

Karin Tamura
marketdata
Forte experiência no ensino superior de estatística. Consultora. Habilidades em statistical modeling, database marketing, B2B e data analysis. Post-PhD com foco em estatística na USP.

Daltro Martins
rmas
Fundou a AG2 Nurun em 1999, empresa focada no desenvolvimento de projeto e campanhas digitais. Atualmente é Diretor da ABRADI Nacional e Presidente da ABRADI-SP.

Rodrigo Yuji
Santander
Experiência na superintendencia de projetos no segmento bancario. Habilidades em data analysis, business strategy, team management, project management, and team leadership. Bacharel em Física na USP.

Marcelo Sousa
marketdata
Data Marketeer, com experiência em customer analytics e data driven marketing strategies. Design e implementação em projetos de CRM CI para companhias como Renault, Claro, Banco Santander, Nextel, PagSeguro, Vivo, GVT etc. Diretor Executivo no grupo WPP. Presidente da ABRADI.

Fonte: ABRADI⁴³

A palestra apresentada por Sousa (2019) teve a seguinte agenda: *Data-Driven companies*; como se tornar uma delas; níveis e técnicas de *Analytics* e Inteligência Artificial.

Para mostrar a importância da análise de dados, Sousa (2019) listou alguns líderes analíticos atuantes em vários setores: P&G, *FedDex*, *Amazon*, *Netflix*, Vivo, Magazine Luiza, Santander, Folha de São Paulo.

⁴³ https://www.sympla.com.br/como-transformar-sua-empresa-a-partir-de-dados__438462

De acordo com Sousa (2019), os líderes analíticos possuem como principais características o comprometimento com a alta administração para obter os recursos e o apoio necessário, além de possuir uma cultura analítica que reconhece a importância e o benefício do uso dos dados. Sousa (2019) citou a *Netflix* que consegue prever a preferência de filme de seus clientes; a *Uber* que chega o mais rápido possível até o usuário e a *Amazon* que faz entregas em menor tempo. Para ele: “Em uma empresa *data-driven* a gestão é baseada em fatos e feita por pessoas que fazem uso extensivo dos dados” (SOUSA, 2019).

Sousa (2019) apresentou os seguintes estágios (níveis) de maturidade analítica nas empresas de marketing: ADHOC, Informativo, *Real Time*, Preditivo e Prescritivo.

Figura 26 - Estágios da maturidade analítica nas empresas de marketing

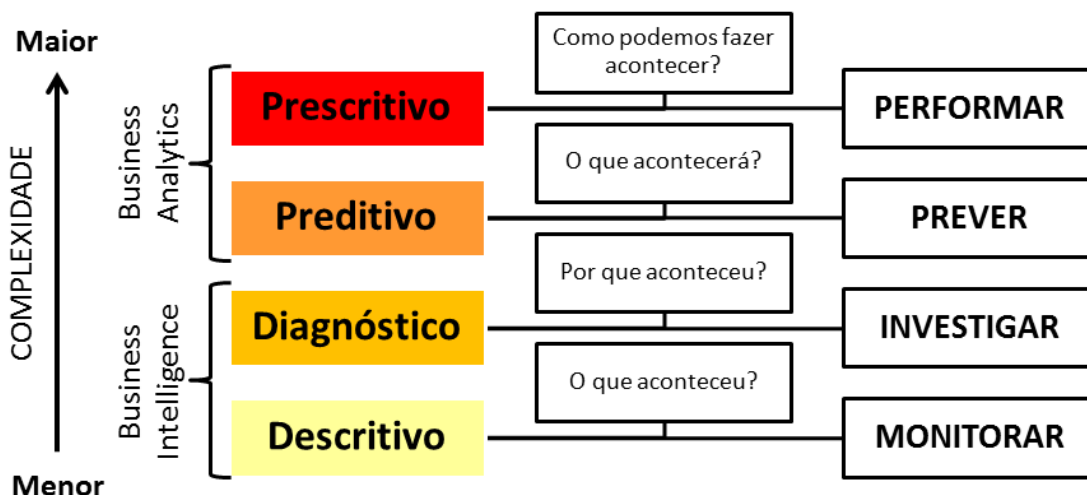
	ADHOC	Informativo	Real Time	Preditivo	Prescritivo
Organização	Reduzida e organizada por canais (e-mail, e-commerce, etc).	Por funções (mensuração de campanhas, web analytics, etc).	Processo analítico formalmente estabelecido.	<i>Analytics</i> como referência interna de equipe.	Processo decisório suportado pelo uso de dados.
Processos	Ações massificadas, sem segmentação, estudos eventuais.	Relatórios atualizados sistematicamente, realização de testes A/B.	Web <i>dashboards</i> atualizados com alta frequência, integrando dados de todos os canais.	Visão multicanal, processos sofisticados de segmentação, algoritmos.	Segmentação em tempo real, personalização de conteúdo em todos os canais.
Competências	Poucos profissionais com <i>skills</i> analíticos.	Alguns talentos trabalhando isolados.	Equipe de <i>analytics</i> estabelecida.	Experiência e métodos de trabalho estabelecidos.	Equipe capaz de utilizar soluções estatísticas avançadas.
Sistemas	Nenhuma automação de marketing e relatórios em Excel.	Baixa automação, dados compilados de várias plataformas.	Ferramentas de BI, alto nível de integração.	Alto nível de automação de marketing, integração em tempo real.	Sistemas inteligentes e adaptáveis, consumindo modelos e algoritmos.
Dados	Armazenados em silos e ausência da <i>data quality</i> .	Dados transacionais históricos consolidados.	Integração entre dados transacionais e web.	Integração entre dados transacionais, web e DMPs.	Dados de todos os canais alimentando o processo decisório e o sistema integrado de automação de marketing.
	O que aconteceu?	Por que aconteceu?	Por que está acontecendo?	O que acontecerá?	O que podemos fazer acontecer?

Fonte: Gartner apud Sousa (2019)

Segundo Sousa (2019), os níveis ADHOC e Informativo estão relacionados com o *Business Intelligence*. O nível ADHOC é descritivo e responde à pergunta “O que aconteceu?”. No nível Informativo é possível fazer diagnóstico e responder “Por que aconteceu?”. Já os níveis Preditivo e Prescrito são mais sofisticados e estão relacionados com o *Business Analytics*. O nível Preditivo permite responder a

pergunta “O que acontecerá?” enquanto o nível Prescritivo responde “O que podemos fazer acontecer?”. De acordo com Sousa (2019), a complexidade é muito maior nos níveis Preditivo e Prescritivo como mostra a Figura 27.

Figura 27 - Complexidade e ações de cada tipo de processo analítico



Fonte: Sousa, 2019

Para Sousa (2019), as vantagens do uso de análise de dados como diferencial competitivo no marketing são as seguintes:

Cliente mais fieis, ter o melhor preço, otimizar portfólio de produtos, otimizar cobertura e localização das agências, maior disponibilidade e logística, entrega mais rápida, 100% digital, ter os melhores profissionais do mercado (SOUSA, 2019).

Sousa (2019) citou quatro pilares para uma análise de dados bem-sucedida:

Estratégia, dados e tecnologia, processos e pessoas. A estratégia para o uso de dados precisa ser muito bem definida. Os dados e a tecnologia são o bloco fundamental, sem isso o processo inteligente e em tempo real de análise não acontece. Os dados devem ser intrínsecos aos processos de tomada de decisão. Assim como os dados e tecnologia, as pessoas são fundamentais (SOUSA, 2019).

De acordo com Sousa (2019), é preciso ter um plano estratégico para trabalhar com dados:

Que dados são necessários? Onde os dados podem ser obtidos? Qual o volume necessário? Quais as regras e processos para capturar, manter e consumir os dados? Uma vez de posse dos dados, os mesmos devem ser integrados e essa integração deve considerar diversas bases de dados diferentes. Deve haver um processo de qualidade para normalizar esses dados e em seguida disponibilizá-los em repositório de forma que ferramentas de *Business Intelligence* (BI) e *Business Analytics* (BA) possam utilizá-los (SOUSA, 2019).

A Figura 28 ilustra as etapas desse processo de tratamento dos dados.

Figura 28 - Etapas do tratamento dos dados



Fonte: Sousa, 2019

De acordo com Sousa (2019), o processo analítico é dividido em 5 etapas: Definição de objetivos; preparação dos dados; desenvolvimento do estudo ou modelo; plano de teste e *rollout* e mensuração. A Figura 29 ilustra as etapas do processo analítico definidas por Sousa (2019).

Figura 29 - Etapas do processo analítico



Fonte: Sousa, 2019

6.5 Especialistas

Foram realizadas entrevistas por e-mail com dois especialistas: Ambriola e Ikeda.

6.5.1 Ambriola

Sobre o aprendizado de máquina e exemplos de algoritmos, Ambriola (2019) concedeu as seguintes informações:

O aprendizado de máquina é uma evolução do reconhecimento de padrões e aprendizado computacional em inteligência artificial. Esse aprendizado explora o estudo e construção de algoritmos que podem aprender através de seus erros e fazer previsões sobre dados de forma retroalimentada, extraíndo padrões de grandes volumes de dados. Gosto de falar que trabalho com aprendizagem estatística, uma vez que são aplicados modelos com base no que se conhece de estatística: regressão linear, regressão logística, *clustering*, análise fatorial, análise discriminante e séries temporais. Depende muito do tipo de problema abordado. Dentre os tipos básicos de aprendizagem de máquinas temos a supervisionada e a não supervisionada. Os tipos comuns de aprendizagem supervisionada que tenho experiência são as árvores de decisão, regressão logística, florestas aleatórias e *support vector machines* (ou máquinas vetores de suporte); e de aprendizagem não supervisionada, tenho experiência com *clustering* e análise de componentes principais (AMBRIOLA, 2019).

Em relação à aplicação desse tipo de tecnologia, Ambriola (2019) considera que ela pode ser empregada em diversas áreas. Citou exemplos de aplicação na área da saúde e *bureaus* de crédito.

Na área da saúde, ela forneceu um exemplo de aplicação de aprendizado de máquina e explicou como os dados utilizados na análise são obtidos:

Trabalhei com modelagem matemática de risco a saúde de pacientes de convênios médicos. A partir de dados cedidos de forma consentida pelos beneficiados do programa para monitoramento de vidas, era possível retroalimentar um modelo de regressão logística que classificasse o risco de internação dos pacientes com base em diversas covariáveis: resultados de exames de sangue, informações sobre hábitos de consumo, qualidade de vida eram preditoras para avaliar o risco dos pacientes com doenças crônicas. Esse monitoramento e classificação auxiliavam as equipes de cuidados médicos a dedicarem maior atendimento para os casos mais urgentes e também colaborava na avaliação da qualidade da saúde desses pacientes (AMBRIOLA, 2019).

Na área de *bureau* de crédito, Ambriola (2019) deu o seguinte exemplo de aplicação de aprendizado de máquina e informou como os dados utilizados na análise são obtidos:

A regressão logística é muito utilizada nos *bureaus* de crédito, diria ainda que é o mais comum no mercado: a classificação de riscos de crédito, recuperação, fraude e até mesmo propensão a consumo de produtos de crédito. Esses produtos são vigorosamente utilizados para a tomada de decisão a respeito das pessoas e, muitas vezes,

utilizados de forma única na tomada de decisão quanto à concessão de crédito para um determinado indivíduo. De um lado, a empresa fica entre dispor de mais recursos para avaliação da concessão de crédito ou perder a chance de vender um empréstimo a um consumidor com bom comportamento e, do outro lado, o consumidor deixa de ter uma oportunidade sem saber quais fatores foram responsáveis pela rejeição. Em *bureaus* de crédito, os dados são coletados por meio dos bancos, que possuem o consentimento do uso de dados de forma mandatária a partir do momento em que o cliente fecha um contrato com a instituição, cedendo seus dados comportamentais (pagamentos) e cadastrais (dados pessoais) para as instituições (AMBRIOLA, 2019).

No que diz respeito ao questionamento se o grande volume de informação gerado e compartilhado em ambientes digitais impõe novos olhares e estratégias de marketing, Ambriola (2019) alerta que a sua maior preocupação está diretamente ligada à forma como os dados são adquiridos. Para ilustrar esse temor, deu o seguinte exemplo:

A captura de imagens em espaços públicos é um exemplo grave. Mesmo sem vincular um número de documento, através de reconhecimento facial é possível identificar as pessoas, claro que baseado em probabilidade. Contudo, o nível de acertos desses modelos, a capacidade técnica em aprimorar câmeras, processamento de visão computacional e associação com outras características como expressão facial, emoções, temperatura corporal, captação de movimentos da pupila, permitem a aquisição de informações muito invasivas que somente em raros casos os usuários estão cientes (AMBRIOLA, 2019).

A especialista também chama a atenção para a forma como muitas empresas de marketing realizam a aquisição dos dados:

Muitas empresas de marketing adquirem os dados por meio de aplicativos, dados de navegação dos usuários na rede e de uma forma que acho bem complicada: através das famosas "empresas parceiras" que são citadas nos contratos. No momento em que o cliente firma um contrato, ele cede autorização de seus dados para os "parceiros", de forma completamente vaga e superficial. Nesses contratos normalmente não há relação de quais são os parceiros, quais seriam as formas de uso dos dados e sob quais condições os dados são cedidos (AMBRIOLA, 2019).

Em relação ao uso de aprendizado de máquina no marketing como vantagem competitiva, Ambriola (2019) ponderou:

Não há como proibir a implementação de modelos matemáticos e seus usos. Controlar seus limites também não faz sentido do ponto de vista matemático, pois muitos modelos possuem como natureza explorar novas combinações de dados de forma automatizada. O maior problema mesmo está na forma da aquisição de dados e seu uso. As pessoas não possuem acesso a todas as covariáveis que

são utilizadas nesses modelos, e não sabem de que forma são categorizadas e estruturadas para criação desses algoritmos. Muitos dos dados são cedidos a terceiros e o consumidor não fica nem sabendo quem adquiriu os dados e pode até mesmo vir a ser vítima de extorsão, além de ser fortemente assediado por empresas com as quais nem possui vínculo (AMBRIOLA, 2019).

Sobre os riscos e pontos negativos proporcionados pelo uso de aprendizado de máquina nas campanhas eleitorais, Ambriola chamou a atenção para os seguintes aspectos:

O problema não é a tecnologia em si, mas sim a natureza encoberta da campanha e a insinceridade da mensagem política e o uso indiscriminado e obscuro de informações sem consentimento da população. É muito séria essa questão da aquisição e uso de dados de massas populacionais sem o consentimento. As pesquisas e comportamentos em redes, análise de sentimentos e sensações são armas perigosas para estratégias de marketing eleitoral. Por exemplo: se um candidato sabe que em determinado local a população está insatisfeita com um determinado tipo de serviço, as estratégias indicam aumento de publicidade para consertar esse problema. Porém, ao avaliar o tipo de comportamento pessoal admirado, tipo de situação que incomoda uma massa populacional, é possível criar novos personagens, criando políticos com comportamentos e postura simulados de acordo com o estudado previamente. (AMBRIOLA, 2019).

Ainda nesse cenário de campanhas políticas, Ambriola (2019) chama a atenção para o uso de robôs:

O uso de robôs também é uma aplicação de inteligência artificial com aprendizagem, e podem ser utilizados para disseminação de notícias falsas; também é programado através de algoritmos que detectam assuntos polêmicos e ampliam sua disseminação. Os *bots* são contas autônomas que são programadas, na arena política, para espalhar mensagens políticas unilaterais, criando a ilusão de apoio público. Normalmente disfarçados de contas humanas comuns, os *bots* têm sido responsáveis por espalhar desinformação e contribuir para um clima político amargo. A tecnologia permite que eleitores diferentes recebam mensagens diferentes com base em previsões sobre sua suscetibilidade a diferentes argumentos (AMBRIOLA, 2019).

Segundo Ambriola, a forma com que os políticos adquirem os dados não é transparente e em alguns casos podem até ser antiética e ilegal. E ilustrou essa observação com o seguinte cenário:

Políticos em campanha na periferia com dados de negativação (registro de débito em *bureau*) das pessoas vendendo votos em troca de alívio financeiro. Esse volume de dados não é autorizado para uso de pessoa física e os casos de marketing eleitoral utilizam os dados sem divulgação (AMBRIOLA, 2019).

De acordo com Ambriola (2019), antigamente os políticos tinham ferramentas limitadas para avaliar os sentimentos e as opiniões dos eleitores durante as eleições. Mas, com os avanços tecnológicos nas campanhas eleitorais o cenário se modificou:

O mais impressionante é que as soluções de inteligência de máquina agora estão sendo cuidadosamente implantadas em campanhas eleitorais para envolver os eleitores e ajudá-los a serem mais informados sobre as principais questões políticas. O problema é como isso pode ser utilizado para manipulação de massas, dando uma falsa sensação de estado democrático. No Brasil, os *bots* não foram implementados na última eleição presidencial de forma inteligente (me referindo ao quesito de complexidade computacional), mas é um princípio de implementação que já foi suficiente para surtimento de resultado, combinado a outros fatores econômicos (AMBRIOLA, 2019).

6.5.2 Ikeda

Sobre o uso de *machine learning* (aprendizado de máquina) no marketing e o que de importante este conceito trouxe para esse segmento, Ikeda (2019) considera o seguinte:

Machine learning é usado em marketing há quase 20 anos, especificamente em *websites*. Dentre as inúmeras aplicações de *machine learning* em *websites*, cito uma delas: *A/B testing*⁴⁴ quando um usuário acessa uma página, ela pode ser modificada dinamicamente, desde a formatação, até itens de publicidade, imagens e o próprio texto. Fica registrada a versão da página visualizada pelo usuário, por quanto tempo ele ficou vendo a página, quais itens ele clicou, etc. Com base nas reações dos usuários às diversas versões apresentadas, um modelo de *machine learning* procura identificar quais características podem influenciar positivamente uma reação do usuário, como clicar em um anúncio, realizar uma compra (IKEDA, 2019).

Em relação à evolução do uso de *machine learning* (aprendizado de máquina) no Brasil, a partir de quando isso pode ser observado e em que direção está o especialista tem a seguinte opinião:

Como em qualquer parte do mundo, sim há evolução e crescimento no interesse por *machine learning* no Brasil. Um aumento

⁴⁴ De acordo com Mathias (2018), o teste A/B tem por finalidade analisar dois modelos e verificar qual deles apresenta melhores resultados. Nas estratégias de marketing, esta ferramenta compara variáveis e ajuda a elevar a taxa de conversão das campanhas. Ele cita, por exemplo, que o teste A/B é usado no nos *layouts* e conteúdos de *e-mail* marketing. Esta ferramenta ajuda a equipe de marketing digital a determinar qual texto poder ter maior potencial de conversão ou mesmo qual arte é a mais atraente.

significativo no interesse, em minha opinião, que pode ser observado a partir de 2015. Assim como outras tecnologias relacionadas à Computação, observa-se um certo atraso entre o que acontece nos Estados Unidos, Europa e Japão e aqui (IKEDA, 2019).

Na visão de Ikeda (2019), do ponto de vista estratégico, o que uso de *machine learning* (aprendizado de máquina) poderá contribuir para o marketing é:

Personalizar a experiência do usuário em alguma plataforma digital *online*, *Apps* principalmente, procurando maximizar a conversão e a concretização de vendas, por exemplo. (IKEDA, 2019).

Ao ser questionado se o uso de *machine learning* (aprendizado de máquina) no marketing e no marketing político tende a aumentar, ele é categórico:

Definitivamente sim. Muitos avaliam que a reeleição de Obama se deve a uma lista de eleitores que foram contactados com base em análise preditiva via *machine learning* (algoritmos determinaram eleitores que com maior chance de votar nele), mas que talvez não compareceriam para votar. A eleição de Trump também usou largamente *machine learning*. Aparentemente eles veicularam 10.000 variações de propagandas para influenciar diferentes grupos de pessoas *microtargeting* (IKEDA, 2019).

Sobre os riscos e aspectos negativos proporcionados pelo uso de *machine learning* (aprendizado de máquina) nas campanhas publicitárias e eleitorais, Ikeda (2019) chamou a atenção para os seguintes pontos:

Pode ser uma poderosa ferramenta de manipulação. *Deep learning* já permite a criação de vídeos *fakes* muito realistas. Não sei ao certo o que pode ser feito para evitar isso, mas certamente se utilizará *deep learning* para combater *deep learning* (IKEDA, 2019).

6.6 Análise das Soluções Levantadas

Ao todo foram levantadas e analisadas 22 soluções, conforme mostra a Figura 30. Esta análise considerou as informações obtidas nas entrevistas e, principalmente, as informações disponibilizadas nos sites, *blogs*, casos de uso das empresas/*startups*.

Figura 30 - Soluções levantadas por empresa/*startup*

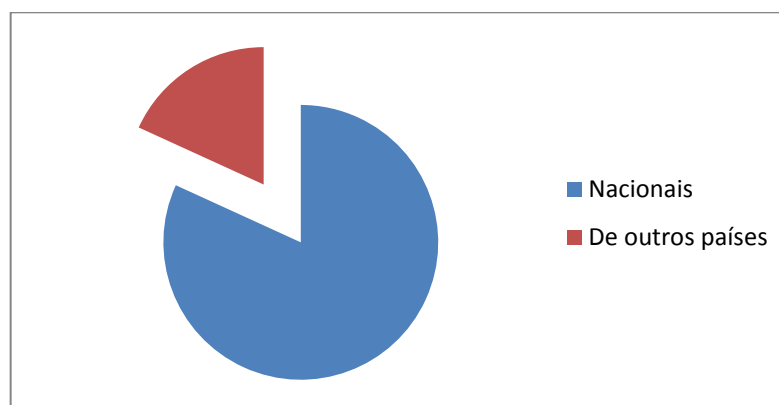
Polis	Tail	Elife	Stilingue	MarketData
<ul style="list-style-type: none"> • Brandwatch Analytics • Brandwatch Image Insights • Brandwatch Audiences • Crimson Hexago • NetBase • v-tracker Monitoramento • v-tracker SAC 2.0 	<ul style="list-style-type: none"> • Premium DMP - Audience Center • Onboard de CRM e Enriquecimento • Monitoramento de Campanha • Dados para segmentação de campanhas online • Personalização do conteúdo • Location Data e Geo Marketing 	<ul style="list-style-type: none"> • Social Intelligence • Brand Logo Detection • Consumer centered Monitoring • Social CRM - e-mail SAC Inteligente • Social CRM - Facebook Bots 	<ul style="list-style-type: none"> • ENERG.IA 	<ul style="list-style-type: none"> • Database Management (DBM) • Analytics and Decision Science • Creative Services

Fonte: Autora

Todas as soluções são de código fonte proprietário. Não foram identificadas nenhuma solução *open source* nessas empresas/*startups*.

A maioria das soluções são desenvolvidas no Brasil. Mas, há algumas tecnologias que não fazem parte dessa lista: *Brandwatch Analytics*, *Brandwatch Image Insights*, *Brandwatch Audiences*, *Crimson Hexago*, *NetBase*. A *Brandwatch* é uma tecnologia de uma empresa de monitoramento de mídia social com sede em Brighton, na Inglaterra. A *Crimson Hexago* é uma tecnologia de uma empresa de *insights* para consumidores sediada em Boston, Massachusetts. Já a *Netbase*, que foi utilizada na campanha de Barack Obama de 2012, foi desenvolvida na Califórnia, nos Estados Unidos. A Figura 31 mostra as soluções nacionais versus as soluções de outros países.

Figura 31 - Soluções nacionais versus soluções de outros países



Fonte: Autora

A maioria das soluções é utilizada para monitoramento da Internet e redes sociais, visando inteligência de mercado, análise de audiência, análise de tendência, análise de marca, gestão de crise e para criação de perfis psicográficos. Entre as soluções utilizadas para estas finalidades estão: A *Brandwatch Analytics*, *Crimson Hexago*, *Netbase* e *v-tracker* Monitoramento que são representadas pela Polis; Monitoramento de Campanha, da Tail; *Social Intelligence*, *Consumer-centered Monitoring*, da Elife e *Creative Services*, da MarketData.

As soluções de segmentação são empregadas em várias aplicações como, segmentação e clusterização para *insights* em tempo real, para ativar em mídia programática, portais e mídias sociais; segmentação demográfica, interesses, estilos de vida, intenção de compra e dados do comportamento de navegação (*browsing behavior*); segmentações de audiência por comportamento geográfico e de utilização de apps (*geo behavior* e *app behavior*) para estratégias de marketing. Entre as soluções utilizadas para estas finalidades estão: *Netbase*, que é representada pela Polis; A *Premium DMP: Audience Center*, Dados para segmentação de campanhas online e *Location Data* e *Geo Marketing* da Tail; *Social CRM*, da Elife e a *Analytics and Decision Science*, da MarketData.

As soluções com foco em personalização de conteúdo combinam a capacidade de executar ideias de forma única com *insights* estratégicos derivados dos dados para criar, projetar e produzir campanhas criativas *one to one*, integradas e personalizadas. As soluções voltadas para este tipo de aplicação são: a Personalização do Conteúdo, da Tail; a *Creative Services*, da MarketData; a *Netbase*, representada pela Polis e *Social CRM*, da Elife.

As soluções de análise de logomarcas e imagens são usadas para ver as menções das imagens; entender dados de autores; entender o sentimento das menções; verificar hábitos de consumo; utilização indevida de marca; associação de marcas e para obter *insights* de conteúdo. Entre as soluções adotadas para estas finalidades estão: *Brandwatch Image Insights* e *Crimson Hexago* representadas pela Polis e *Brand Logo Detection*, da Elife.

A gestão de influenciadores permite obter informações de forma instantânea sobre qualquer tipo de público para entender e encontrar microinfluenciadores. Entre as soluções utilizadas para estes propósitos estão: *Brandwatch Audiences* e *Crimson Hexago* representadas pela Polis, e a *Social Intelligence*, da Elife.

Nas mídias sociais, a análise de sentimento e gestão de reputação para revelar as emoções e desejos da audiência são feitas pelas soluções *Crimson Hexago*, *Netbase* e *v-tracker* representadas pela Polis.

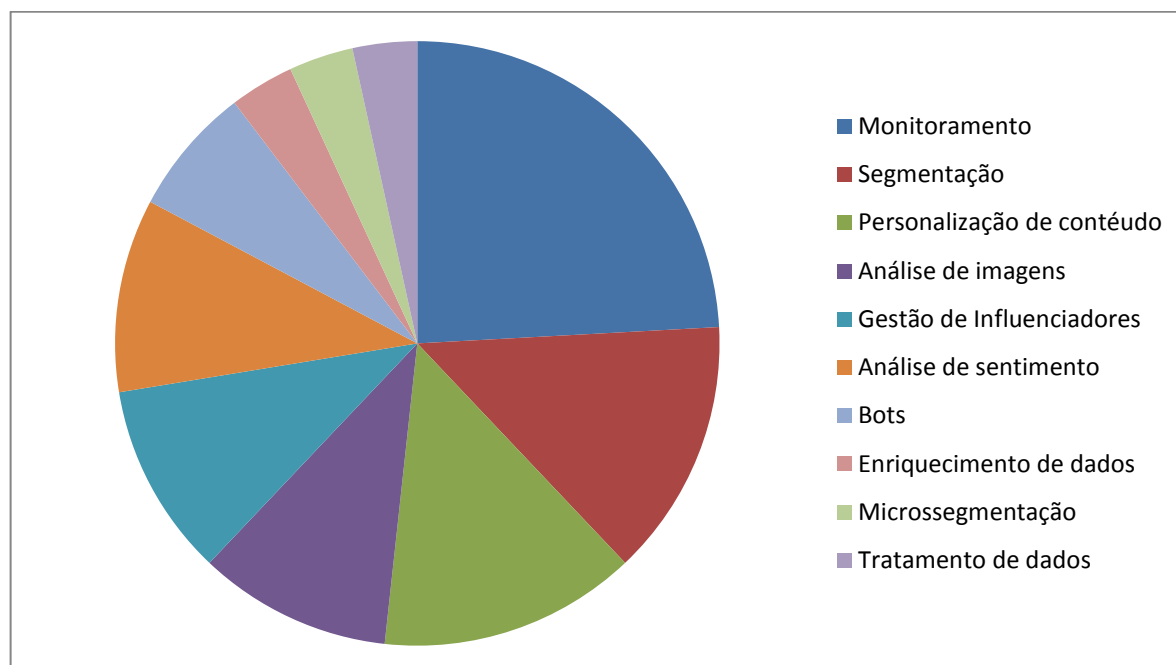
O uso de *bots* de autoatendimento é feito pelas seguintes soluções: *v-tracker SAC 2.0* representado pela Polis e *Social CRM - Facebook Bots*, da Elife.

O enriquecimento de dados por meio do cruzamento de informações *online*, dados de *cookies*, dados de CRM⁴⁵ e de *bureaus* de crédito como Serasa, Boa Vista Serviços e SCPC é feito apenas pelo *Onboard* de CRM e Enriquecimento, da Tail. Além do enriquecimento de dados, esse cruzamento também permite que a Tail consiga fazer *microsegmentação*.

O tratamento de dados que inclui extração, tratamento, carga de dados (ETL), gestão do processo de *database management*, identificação de origens, auditoria de processos, qualidade, integração de dados transacionais e *web behavior* é realizado pela solução *Database Management (DBM)*, da MarketData.

A Figura 32 apresenta de forma sintetizada as aplicações das soluções analisadas.

Figura 32 - Aplicações das soluções analisadas



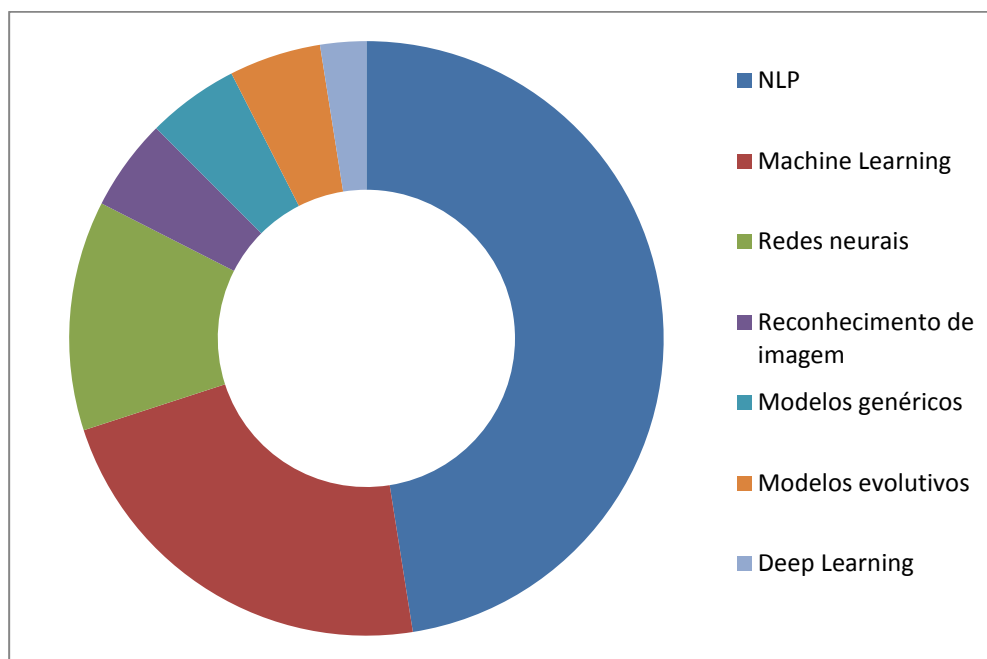
Fonte: Autora

⁴⁵ *Customer Relationship Management* ou em português: Gestão de Relacionamento com o Cliente.

Quanto à obtenção dos dados, todas as empresas/*startups* analisadas coletam informações públicas disponibilizadas na Internet e nas redes sociais. Além desse tipo de coleta, a Tail, a Elife, a Stilingue e a MarketData citaram em seus sites que também obtêm dados de *cookies*. A Tail foi a única a mencionar que faz cruzamento de dados *online* com dados de CRM e *bureaus* de crédito.

Em relação aos algoritmos empregados nas soluções analisadas, destacam-se: NLP (*natural language process*), *machine learning*, rede neurais, reconhecimento de imagem, modelos genéricos, modelos evolutivos e *deep learning*. Sendo que o mais utilizado é o NLP (*natural language process*), seguido de *machine learning* e redes neurais, conforme mostra a Figura 33.

Figura 33 - Algoritmos utilizados nas soluções



Fonte: Autora

7 CONSIDERAÇÕES FINAIS

A pesquisa empírica mostra que o aprendizado de máquina está sendo utilizado em campanhas eleitorais aqui no Brasil, e que esse uso teve início em 2014, quando a empresa Polis Consulting trouxe ao país a plataforma *Netbase*, que foi uma das soluções utilizadas na campanha de Barack Obama, nas eleições de 2012. Outras soluções que se utilizam do aprendizado de máquina para campanhas comerciais e eleitorais estão surgindo, e inclusive são tecnologias nacionais. Empresas com a Elife, Tail, MarketData e Stilingue utilizam aprendizado de máquina em suas soluções. A Stilingue, assim como a Polis, também atua em campanhas comerciais e campanhas políticas. Aliás, foi verificado que na campanha eleitoral de 2018, João Dória, na época candidato ao governo de São Paulo, utilizou as soluções da Stilingue em paralelo com outros softwares de monitoramento, que não foram declarados.

Esta pesquisa constatou que essas ferramentas tecnológicas estão sendo aplicadas nas campanhas políticas com as seguintes finalidades: monitoramento dos eleitores por meio da Internet e redes sociais para identificação de influenciadores, análise de audiência, gestão de reputação do candidato, análise de sentimento para revelar as emoções e desejos e segmentação de grande volume de dados para direcionamento do discurso de forma personalizada.

Também foi levantado que algumas empresas estão atuando com o chamado SAC 2.0 e com a utilização de *bots* de autoatendimento. Para Torreta (2019) o marketing político é bem mais sensível do que o marketing comercial, e que numa campanha eleitoral o eleitor não vai aceitar ser atendido por uma máquina. Na contramão desse pensamento, Lima (2019) acredita que uma campanha política pode utilizar os dados do próprio candidato, o seu plano de governo e treinar *bots* para dialogar com os eleitores; abrir um canal direto para responder perguntas e obter *insights* dos próprios eleitores diminuindo o custo da campanha.

A Tail trabalha com o cruzamento de dados (*online*, *cookies*, de CRM e de *bureaus* de crédito) permitindo definir microsegmentações. Se essa solução for utilizada em conjunto com técnicas de psicometria, é possível traçar o *microtargeting* comportamental que, além de auxiliar no direcionamento do discurso, permite personalizar o conteúdo de modo individualizado; prática essa que vem do *marketing one to one* (individualizado). Foi possível verificar que o *microtargeting*

comportamental permite diferenciar não apenas o conteúdo, mas a linguagem que deve ser utilizada; diferente da segmentação que possibilita somente a personalização do conteúdo de um discurso. Já com o uso do *microtargeting* comportamental é possível personalizar conteúdo e forma. De acordo com o levantamento feito, a Tail não atua em campanhas políticas. Embora tenha sido procurada pelo candidato João Amoedo, na época da campanha eleitoral à presidência de 2018, a empresa não aceitou esse trabalho.

Apesar de alguns publicitários acreditarem que o *microtargeting* comportamental, da forma como foi executado pela *Cambridge Analytica*, não esteja sendo aplicado ainda no Brasil em campanhas eleitorais, há uma série de denúncias junto à Justiça Eleitoral que permite questionar tais crenças. Conforme Kotler (1975) esclareceu, dificilmente técnicas do marketing comercial não são empregadas no marketing político. Nos Estados Unidos, o *microtargeting* comportamental é uma realidade tanto no marketing comercial quanto no político. No Brasil, algumas empresas já trabalham com microssegmentações para o setor privado. Como as técnicas de psicométrica e microssegmentação são conhecidas por diversos consultores políticos, elas podem já estar a serviço de campanhas políticas.

Um ponto importante investigado neste estudo foi o processamento de linguagem natural, um subcampo da Inteligência Artificial, que entre as várias aplicações, é utilizado para a operação dos chamados robôs sociais. Ambriola (2019) chamou a atenção para o uso de robôs ou *bots*, que podem ser utilizados para disseminação de notícias falsas, que por meio de algoritmos detectam assuntos polêmicos e ampliam sua disseminação. Outro aspecto observado por ela, é que os *bots* são contas autônomas programadas, e que na arena política podem ser utilizados para espalhar mensagens unilaterais, criando a ilusão de apoio público. Normalmente disfarçados de contas humanas comuns, os *bots* têm sido responsáveis por propagar desinformação e contribuir para um clima político polarizado. Essa tecnologia permite que eleitores diferentes recebam mensagens diversas com base em previsões sobre sua suscetibilidade a diferentes argumentos. Para Ambriola (2019), a forma com que os políticos adquirem os dados para tais feitos não é clara, nem um pouco transparente e nada ética. Já Ikeda (2019) considerou que o uso negativo do aprendizado de máquina nas campanhas publicitárias e eleitorais pode ocorrer quando estas são direcionadas à manipulação. Segundo ele, o *deep learning* já permite a criação de vídeos *fakes* muito realistas.

Embora a campanha eleitoral à presidência de 2018, no Brasil, tenha sido atípica, como observaram alguns entrevistados, e necessite de tempo para uma análise mais precisa, uma questão importante a ser observada foi o intenso uso das redes sociais e do *WhatsApp*, principalmente na campanha do então candidato à presidência Jair Bolsonaro. Em 2015, Antoniutti (2015) já havia sinalizado uma tendência de uso do *WhatsApp* como uma nova estratégia de comunicação com os eleitores e essa tendência se concretizou nas eleições de 2018. Outro ponto observado por Lopes (2019) sobre essa atípica campanha eleitoral, é que houve uma inversão de papéis e a mídia de massa, que nas eleições anteriores era protagonista, acabou atuando como apoio e complemento para a Internet e para as redes sociais. Torreta (2019) acredita na consolidação do *WhatsApp* como ferramenta de campanha publicitária e eleitoral, porém pondera que ainda é cedo para concluir que a mídia de massa perdeu a sua importância.

É essencial destacar a existência daquilo que Kohlitz (2019) denominou de zona cinza, ou seja, aquela zona sem lei, onde não está claramente definido o que é permitido e o que não é, abrindo uma brecha para que pessoas se aproveitem da situação. Segundo Kohlitz (2019), foi isso o que aconteceu com a *Cambridge*, quando usaram o aplicativo “*This is your Digital Life*” para coletar dados dos usuários do *Facebook* e dos amigos dos usuários. Como apontou Kohlitz, isso na época, era uma zona cinza, ou seja, não havia uma delimitação sobre a licitude ou não das operações. Kohlitz (2019) também chama a atenção para outro exemplo de zona cinza que ocorreu na campanha de desinformação em 2018, no Brasil, em que as listas com os números de *WhatsApp* foram adquiridas e trabalhadas por grupos políticos.

A questão ética é outra grande implicação que mereceu uma atenção especial neste estudo. Conforme destacou Torreta (2019), quem inovou no desenvolvimento tecnológico em campanhas eleitorais foi o Obama e não a *Cambridge Analytica* ou a campanha do Trump. A novidade da *Cambridge* e da Campanha de Trump não teve nada relacionado com tecnologia. O que eles fizeram de inovador foi colocar a metodologia de comportamento nessa tecnologia que já tinha sido utilizada pelo Obama. A campanha de Obama, em 2012, utilizou toda essa tecnologia para subsidiar a definição da estratégia da campanha e auxiliar nas táticas traçadas. Segundo Kohlitz (2019), Obama, ao contrário de Mitt Romney, empregou esse conhecimento para saber quem eram os indecisos e tentar trazê-los para o seu lado,

e essa tática foi essencial para definir o resultado da eleição, porém essa estratégia não ultrapassou e nem feriu nenhum limite ético.

A estratégia da campanha do Trump, por sua vez, usou a tecnologia em conjunto com a metodologia de comportamento; não para fortalecê-lo enquanto candidato, e sim para prejudicar a campanha da concorrente, por meio de mensagens repletas de desinformação, que acabou fazendo com que várias pessoas desistissem de votar em Hillary Clinton. Legalmente isso não é proibido, porém em termos éticos é algo que não deveria ser aceito.

Tudo o que foi visto até aqui traz à tona muitas reflexões, e uma delas é que os dados são fundamentais para as tecnologias de aprendizado de máquina. Segundo Bigonha (2018), “se os algoritmos são o motor, certamente dados são o combustível dessa revolução tecnológica” (2018, p. 2). Ou seja, sem esse combustível o motor não gira e, conseqüentemente, toda essa engrenagem não anda. Para Bigonha (2018), a inserção de algoritmos de aprendizado de máquina que necessitam de um grande volume de dados, em processos decisórios importantes, como avaliação de crédito, diagnósticos médicos, campanhas políticas, entre outros, deve levantar questionamentos que incluam desafios de transparência, inclusão, privacidade e *accountability*.

Foi possível perceber que, conforme aumenta a utilização de algoritmos de aprendizado de máquina, também cresce a necessidade de se refletir sobre o papel e o alcance dessas soluções tecnológicas. É necessário que haja uma colaboração entre líderes da indústria, do segmento privado, pesquisadores, especialistas, governantes e sociedade civil a fim de garantir que o desenvolvimento dessas tecnologias seja benéfico para a sociedade com um todo. É fundamental que sejam definidos princípios éticos e padrões técnicos para que essas soluções tecnológicas sejam embasadas pela preocupação e cuidado com os efeitos que possam acarretar à sociedade. Além disso, é necessário se antecipar aos riscos para identificar essas zonas cinzas, antes que situações antidemocráticas e falsificações da realidade aconteçam.

8 REFERÊNCIAS BIBLIOGRÁFICAS

ALVARENGA, Geoflávia Guilarducci de. *Aplicações Práticas de Inteligência Artificial para Modelagem Conceitual de Banco de Dados*. 2005. Disponível em: <http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_001-05.pdf>.

Acesso em: 09 maio 2018.

AMORIM, Paula Fernanda Patrícia de. *A crítica de John Searle à Inteligência Artificial: Uma abordagem em filosofia e mente*. 2014. 99f. Dissertação (Mestrado em Filosofia) – Universidade Federal da Paraíba, João Pessoa, PB, 2014.

ANTONIUTTI, Cleide Luciane. *De operário a presidente. A imagem política de Lula no Horário Eleitoral da televisão nas campanhas presidências de 1989 e 2002*. 2004. 158f. Dissertação (Mestrado em Sociologia Política), Universidade Federal do Paraná, Curitiba, PR, 2004. Disponível em: <http://www.educadores.diaadia.pr.gov.br/arquivos/File/2010/artigos_teses/SOCIOLOGIA/cleide.PDF>. Acesso em: 29 maio 2019.

ANTONIUTTI, Cleide Luciane. *Usos do Big Data em Campanhas Eleitorais*. 2015. 271f. Tese (Doutorado em Ciência da Informação), Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, 2015. Disponível em: <<http://ridi.ibict.br/handle/123456789/849>>. Acesso em: 29 maio 2019.

BAUMAN, Zygmunt. *A Cultura no Mundo Líquido Moderno*. Rio de Janeiro: Zahar Editores, 2013.

BERNARD, H. R. *Research methods in anthropology: qualitative and quantitative approaches*. Lanham, MD: AltaMira Press, 2005.

BIDDLE, Sam. *Facebook usa inteligência artificial para prever o comportamento de usuário para anunciantes*. Tradução: Bernardo Tonasse. The Intercept Brasil, 13 abr. de 2018. Disponível em: <<https://theintercept.com/2018/04/13/facebook-inteligencia-artificial/>>. Acesso em: 08 jun. 2018.

BIERNARCKI, Patrick e WALDORF, Dan. Snowball sampling-problems and techniques of chain referral sampling. *Sociological Methods and Research* v. 10, n. 2, p. 141-163, nov. 1981.

BIGONHA, Carolina. Panorama setorial da Internet. *Inteligência Artificial e Ética*, Ano 10, n. 2. 2018.

BRANDÃO JÚNIOR, Francisco de Assis Fernandes. *Palanques Virtuais: A campanha presidencial pela Internet nas eleições de 2006*. Dissertação (Mestrado em Ciência Política), Universidade de Brasília, UnB, Brasília – DF, 2008.

BROOKSHEAR, J. Glenn. *Ciência da Computação: Uma visão abrangente*. Bookman. Porto Alegre. 2001.

CADWALLADR, Carole. The Cambridge Analytica Files - 'I made Steve Bannon's psychological warfare tool': meet the data war whistleblower. *The Guardian*, 18 mar. 2018.

Disponível em: <<https://www.theguardian.com/news/2018/mar/17/data-war-whistleblower-christopher-wylie-facebook-nix-bannon-trump>>. Acesso em: 11 jun. 2018.

CADWALLADR, Carole e GRAHAM-HARRISON, Emma. Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The Guardian*, 17 mar. 2018. Disponível em: <<https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>>. Acesso em: 11 jun. 2018.

CÂMARA, Isabella. Como a Stilingue criou uma Inteligência Artificial sob medida em português. *Startse*, 3 maio 2018. Disponível em: <<https://www.startse.com/noticia/startups/49203/como-stilingue-criou-uma-inteligencia-artificial-sob-medida-ao-portugues>>. Acesso em: 17 maio 2019.

CANAVILHAS, João. *A comunicação política na era da internet*. Corvilhã, Portugal, Universidade da Beira Interior, 2009. Disponível em: <<http://www.bocc.ubi.pt/pag/canavilhas-joao-comunicacao-politica-na-era-da-Internet.pdf>>. Acesso em: 06 abr. 2019.

CASTELLS, Manuel. *Communication Power*. New York: Oxford University Press, 2009.

CASTLEMAN, Dan. CHAPTER 1 - Essentials of Modeling and Microtargeting. In: THERRIAULT, Andrew. *Data and Democracy How Political Data Science Is Shaping the 2016 Elections*. Sebastopol, CA: O'Reilly Media, 2016.

CAVALHEIRO, Glauco e BRANDÃO, Carolina Gandon. Ensaio 3, “Comunicação e retórica: Um contexto teórico para pensar a pós-verdade”. In: GUARESCHI, Pedrinho A; AMON, Denise e GUERRA, André. *Psicologia, Comunicação e Pós-verdade*. Porto Alegre: Abrapso, 2017. P. 83-100.

CHALEGRE, A.; ALMEIDA, J.; PASSOS, R. A importância das pesquisas de opinião para a elaboração de uma estratégia eleitoral. *Diálogos & Ciência - Revista Eletrônica da Faculdade de Tecnologia e Ciências*. Ano 4, n. 7, jun. 2006. ISSN: 1678-0493. Disponível em: <http://dialogos.ftc.br/index2.php?option=com_docman&task=doc_view&gid=46&Itemid=15>. Acesso em: 25 jan. 2019.

COUTINHO, Marcelo. Opinião pública, pesquisas eleitorais e a internet em 2010: possíveis cenários. In: *Em Debate*. V.2, n.5: Belo Horizonte, 2010.

DEAN, Jeff e NG, Andrew. Using large-scale brain simulations for machine learning and A.I. Google. 26 jun. 2012. Disponível em: <<https://www.blog.google/technology/ai/using-large-scale-brain-simulations-for/>>. Acesso: 09 jul. 2019.

DELEUZE, Gilles. *Conversações*. São Paulo: Editora 34. 1992.

DOMINGOS, Pedro. *O algoritmo mestre*. São Paulo: Novatec. 2017.

FERRAZ, Francisco. *Manual completo de campanha eleitoral*. São Paulo: L&M Editores, 2010.

FLUSSER, Vilém. *O mundo codificado: por uma filosofia do design e da comunicação*. São Paulo: Cosac Naify, 2007.

FOUCAULT, Michel. *Microfísica do Poder*. Rio de Janeiro: Paz & Terra, 2015.

FOUCAULT, Michel. *Resumo dos Cursos do Collège de France (1970-1982)*. Rio de Janeiro: Jorge Zahar Editor, 1997.

FRANÇA, Luiz De. Políticos devem se render às redes sociais. *VEJA*, 15 jun. 2009. Disponível em: <<https://veja.abril.com.br/brasil/politicos-devem-se-render-as-redes-sociais/>>. Acesso em: 01 jun. 2019.

GOMES, Neuza Demartini. *Formas persuasivas de comunicação política: propaganda política e publicidade eleitoral*. 1. ed. Porto Alegre. EDUPUCRS, 2000.

GOMES, W., FERNANDES, B., REIS, L., & SILVA, T. " Politics 2.0": Barack Obama's on-line 2008 campaign. In *Revista de Sociologia e Política*, 2009, 17(34), 29-43.

GRAÇAS, Ana Gabriela. O que podemos aprender com Bia, assistente virtual do Bradesco? *Elife*, 16 ago. 2018a. Disponível em: <<https://elifemagazine.com.br/index.php/2018/08/16/o-que-podemos-aprender-com-bia-assistente-virtual-do-bradesco/>>. Acesso em: 16 maio 2019.

GRAÇAS, Ana Gabriela. Social Media na Prática: Bot Suviniil. *Elife*, 22 mar. 2018b. Disponível em: <<https://elifemagazine.com.br/index.php/2018/03/22/social-media-na-pratica-bot-suviniil/>>. Acesso em: 16 maio 2019.

GRAÇAS, Ana Gabriela. Social Media na Prática: o Bot da Opel Espanha. *Elife*, 3 jul. 2017a. Disponível em: <<https://elifemagazine.com.br/index.php/2017/07/03/social-media-na-pratica-o-bot-da-opel-espanha/>>. Acesso em: 16 maio 2019.

GRAÇAS, Ana Gabriela. Social Media na Prática: Case Bot Albert Einstein. *Elife*, 31 out. 2017b. Disponível em: <<https://elifemagazine.com.br/index.php/2017/10/31/social-media-na-pratica-case-bot-albert-einstein/>>. Acesso em: 16 maio 2019.

GRAÇAS, Ana Gabriela. Elife adota Inteligência Artificial em serviços de Inteligência e Social CRM. *Elife*, 8 jun. 2017c. Disponível em: <<https://elifemagazine.com.br/index.php/2017/06/08/elife-adota-inteligencia-artificial-em-servicos-de-inteligencia-e-social-crm/>>. Acesso em: 16 maio 2019.

GRASSIA, Audra. CHAPTER 2 - Data Management for Political Campaigns. In: THERIAULT, Andrew. *Data and Democracy How Political Data Science Is Shaping the 2016 Elections*. Sebastopol, CA: O'Reilly Media. 2016.

GUARESCHI, Pedrinho A; AMON, Denise e GUERRA, André. *Psicologia, Comunicação e Pós-verdade*. Porto Alegre: Abrapso, 2017a.

GUARESCHI, Pedrinho. Ensaio 5 – “Psicologia e Pós-Verdade”. In: GUARESCHI, Pedrinho A; AMON, Denise e GUERRA, André. *Psicologia, Comunicação e Pós-verdade*. Porto Alegre: Abrapso, 2017. P. 161-193.

GUARESCHI, Pedrinho A; AMON, Denise e GUERRA, André. Ensaio 12, *Psicologia, comunicação e pós-verdade: Uma dimensão crítica, ética e propositiva*. Abrapso, 2017b. P. 337-356.

GUERRA, André e BARBOSA, Cláudia. Ensaio 4, “Crítica e Pós-Verdade”. In: GUARESCHI, Pedrinho A; AMON, Denise e GUERRA, André. *Psicologia, Comunicação e Pós-verdade*. Porto Alegre: Abrapso, 2017. P. 101-160.

HAAS, Guilherme. *Singularidade tecnológica: os seres humanos serão dominados pelas máquinas?* 2014. Disponível em: <<https://www.tecmundo.com.br/futuro/50264-singularidade-tecnologica-os-seres-humanos-serao-dominados-pelas-maquinas-.html>>. Acesso em: 29 mai. 2018.

HALPER, Fern. Best practices report Q3: Advanced Analytics: Moving Toward AI, Machine Learning, and Natural Language Processing, 2017, TDWI Research. Disponível em: <https://www.sas.com/content/dam/SAS/en_us/doc/whitepaper2/tdwi-advanced-analytics-ai-ml-nlp-109090.pdf>. Acesso em: 27 fev. 2019.

HAZELWOOD, Kim et al. *Applied Machine Learning at Facebook: A Datacenter Infrastructure Perspective*. Facebook, 24 fev. 2018. Disponível em: <<https://research.fb.com/wp-content/uploads/2017/12/hpca-2018-facebook.pdf?>>. Acesso em: 07 jun. 2018.

HOWARD, Philip N. *Pax Technica: How the Internet of things may set us free or lock us up*. Yale University Press, 2015.

IASULAITIS, Sylvia. *Internet e campanhas eleitorais: experiências interativas nas cibercampanhas presidenciais do cone sul*. 2017. 376f. Tese (Doutorado em Ciências Políticas). Universidade Federal de São Carlos, São Carlos, São Paulo, 2012.

JQUES, Rafael. O que é um Framework? Para que serve? *PHPIT*, 7 jan. 2016. Disponível em: <<http://www.phpit.com.br/artigos/o-que-e-um-framework.phpit>>. Acesso em: 26 maio 2019.

JÚNIOR, Gilberto Timótheo e LIMA, Sérgio Muinhos Barroso. Algoritmos Genéticos Aplicados a Jogos Eletrônicos. *Revista Eletrônica da Faculdade Metodista Granbery*, Curso de Sistemas de Informação - N. 8. 2010. Disponível em: <<http://re.granbery.edu.br/artigos/MzU2>>. Acesso em: 28 set. 2018.

KOSINSKI, M.; STILLWELL, D. e GRAEPEL, T. Private traits and attributes are predictable from digital records of human behavior. 2013. Disponível em: <<http://www.pnas.org/content/pnas/110/15/5802.full.pdf>>. Acesso em: 05 jun. 2018.

KOSINSKI, M. et al. Mining Big Data to Extract Patterns and Predict Real-Life Outcomes. 2016. Disponível em: <<http://www.apa.org/pubs/journals/features/met-met0000104.pdf>>. Acesso em: 05 jun. 2018.

KOTLER, Philip. *Administração de marketing*. São Paulo: Atlas, 1975.

KUNTZ, Ronald A.; LUYTEN, Joseph M. *Marketing político: a eficiência a serviço do candidato*. São Paulo: Global, 1982.

LATOURE, Bruno. *Reagregando o Social: uma introdução à teoria Ator-Rede*. Salvador: Edufba, 2014.

LAZZARATO, Maurizio. *As revoluções do capitalismo*. Rio de Janeiro: Civilização Brasileira, 2006.

LAZZARATO, Maurizio. *Signos, Máquinas, Subjetividades*. São Paulo: Edições Sesc, 2014.

LEVY, Pierre. *As Tecnologias da Inteligência: O futuro do pensamento na era da informática*. Tradução de Carlos Irineu Costa, Rio de Janeiro: Editora 34, 2ª. Edição 2011.

LIMA, Alessandro. Chocolates e refrigerantes estão entre as guloseimas de indulgência mais compartilhadas no Instagram. *Elife*, 21 mar. 2017a. Disponível em: <<https://elifemagazine.com.br/index.php/2017/03/21/chocolates-e-refrigerantes-estao-entre-guloseimas-de-indulgencia-mais-compartilhadas-no-instagram/>>. Acesso em: 16 maio 2019.

LIMA, Alessandro. Cerveja e Catuaba se destacam entre as bebidas mais consumidas no Carnaval. *Elife*, 7 mar. 2017b. Disponível em: <<https://elifemagazine.com.br/index.php/2017/03/07/cerveja-e-catuaba-se-destacam-entre-as-bebidas-mais-consumidas-no-carnaval/>>. Acesso em: 16 maio 2019.

LIMA, Alessandro. Consumer Trends: novo produto da Elife analisa padrões de consumo em diversos segmentos. *Elife*, 27 jan. 2017c. Disponível em: <<https://elifemagazine.com.br/index.php/2017/01/27/consumer-trends-novo-produto-da-elifemagazine-analisa-padroes-de-consumo-em-diversos-segmentos/>>. Acesso em: 16 maio 2019.

LIU, Bing. Sentiment Analysis and Subjectivity. *Handbook of Natural Language Processing*, Second Edition, editors: N. Indurkha and F. J. Damerau. 2010.

LUNDRY, Alex. CHAPTER 4 - Data-Driven Media Optimization. In: THERRIAULT, Andrew. *Data and Democracy How Political Data Science Is Shaping the 2016 Elections*. Sebastopol, CA: O'Reilly Media. 2016.

MANHANELLI, Carlos Augusto. *Eleição é guerra*. São Paulo: Summus, 1992.

MANHANELLI, Carlos Augusto. *Estratégias eleitorais: marketing político*. São Paulo: Summus, 1988.

MANIN, Bernard. As metamorfoses do governo representativo. *Revista Brasileira de Ciências Sociais*, n. 29, p. 5-34, 1995.

MANOVICH, Lev. Estudos do Software: o motor das sociedades contemporâneas. São Paulo, IMESP/FILE, 2008. Disponível em: <<http://lab.softwarestudies.com/2008/08/estudos-do-software-por-lev-manovich.html>>. Acesso em: 05 jun. 2018.

MARQUESONE, Rosângela. *Big Data: Técnicas e tecnologias para extração de valor de dados*. São Paulo: Casa do Código. 2017.

MATHIAS, Lucas. Teste A/B: o que é e como pode melhorar seus resultados. *Mindminers*, 5 fev. 2018. Disponível em: <<https://mindminers.com/blog/o-que-e-teste-ab/>>. Acesso em: 15 maio 2019.

MATZ, S. C et al. Psychological Targeting as an Effective Approach to Digital Mass Persuasion. 2017. Disponível em: <<http://www.pnas.org/content/pnas/114/48/12714.full.pdf>>. Acesso em: 05 jun. 2018.

MCCARTHY, J. et al. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. 1955. *AI Magazine* Volume 27 Number 4 (2006).

MIGUEL, Luis Felipe. Os meios de comunicação e a prática política. *Lua Nova* n. 55-56, p.155-184. 2002. Disponível em: <http://www.scielo.br/pdf/ln/n55-56/a07n5556.pdf>. Acesso em: 02 jun. 2019.

MONTEIRO, Renato Leite. Cambridge Analytica e a nova era Snowden na proteção de dados pessoais. *El País*, 20 mar. 2018. Disponível em: https://brasil.elpais.com/brasil/2018/03/20/tecnologia/1521582374_496225.html. Acesso em: 11 jun. 2018.

MONTEIRO, Thaís. Como a tecnologia impactou a política. Trump e, posteriormente, Bolsonaro, entenderam como dialogar por meio das redes sociais. *Proxima*, 7 maio 2019. Disponível em <https://evento.proxima.com.br/noticias2019/2019/05/07/como-a-tecnologia-impactou-a-politica/>. Acesso em: 08 maio 2019.

MORDVINTSEV, A.; OLAH, C. e TYKA, M. Inceptionism: Going Deeper into Neural Networks. *Google*, 17 jun. 2015. Disponível em: <https://ai.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html>. Acesso em: 09 jul. 2019.

MOREIRA, Daniel. O que é uma startup? *Exame*, 1 mar. 2018. Disponível em: <https://exame.abril.com.br/pme/o-que-e-uma-startup/>. Acesso em: 16 jul. 2018.

MOTA, Camilla Veras. *Dória usa cinco softwares de 'big data' para aumentar seu alcance nas redes*. *BBC Brasil*, 26 set. 2017a. Disponível em: <http://www.bbc.com/portuguese/brasil-41406420>. Acesso em: 11 out. 2017.

MOTA, Camilla Veras. Robôs e 'big data': as armas do marketing político para as eleições de 2018. *BBC Brasil*, 26 set. 2017b. Disponível em: <http://www.bbc.com/portuguese/brasil-41328015>. Acesso em: 11 out. 2017.

NASCIMENTO, Daniel. *Machine Learning - entendendo o processo de aprendizagem*. Medium, 6 maio 2017. Disponível em: https://medium.com/@Daniel_nasci/machine-learning-entendendo-o-processo-de-aprendizagem-80835b3ec2dc. Acesso em: 05 jun. 2018.

NICKERSON, David W. e TODD Rogers. Political Campaigns and *Big Data*. 2014. *Journal of Economic Perspectives*, 28 (2): 51-74. Disponível em: <https://www.aeaweb.org/articles?id=10.1257/jep.28.2.51>. Acesso em: 05 abr. 2019.

NORRIS, Pippa. *Democratic divide? The impact of the Internet on parliaments worldwide*. Harvard: Harvard University, John Kennedy School of Government, 2000.

PASSARINHO, Nathalia. Ex-sócio da Cambridge Analytica no Brasil diz que empresa não tinha banco de dados de brasileiros. *BBC Brasil*, 20 mar. 2018. Disponível em: <https://www.bbc.com/portuguese/brasil-43481279>. Acesso em: 08 maio 2019.

PENCHIKALA, Srini. Big Data com Apache Spark - Parte 1: Introdução. *InfoQ*, 14 nov. 2015. Disponível em: <https://www.infoq.com/br/articles/apache-spark-introduction/>. Acesso em: 29 maio 2019.

PENTEADO, Claudio. *Marketing político na era digital: perspectivas e possibilidades*. Revista USP, 2011, n. 90, p.6-23, 2011.

PICKLER, Maria Elisa Valentim. *Web Semântica: ontologias como ferramentas de representação do conhecimento*. 2007. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362007000100006>. Acesso em: 09 maio 2018.

PRAUDE, Carlos Correia. *Arte Computacional e Teoria Ator-Rede: actantes e associações intersubjetivas em cena*. 2015, 249f. Tese (Doutorado em Arte) - Universidade de Brasília. Brasília, 2015.

PRIMO, Alex. O aspecto relacional das interações na Web 2.0. *E- Compós* (Brasília), v. 9, p. 1-21, 2007.

PUGET, Jean Francois. What Is Machine Learning? *IBM*, 18 maio 2016. Disponível em: <https://www.ibm.com/developerworks/community/blogs/jfp/entry/What_Is_Machine_Learnin_g?lang=en>. Acesso em: 05 jul. 2019.

RENNER, Maurício. Startup mineira varre Internet para Doria. *Baguete*, 27 set. 2017. Disponível em: <<https://www.baguete.com.br/noticias/27/09/2017/startup-mineira-varre-internet-para-doria>>. Acesso em: 20 maio 2019.

RIBEIRO, Jeferson. Publicitário brasileiro suspende parceria com Cambridge Analytica. *O Globo*. 2018. Disponível em: <<https://oglobo.globo.com/mundo/publicitario-brasileiro-suspende-parceria-com-cambridge-analytica-22500990>>. Acessado em: 08 maio 2019.

ROSENBERG, M.; CONFESSORE, N. e CADWALLADR C. How Trump Consultants Exploited the Facebook Data of Millions. *The New York Times*, 17 mar. 2018. Disponível em: <<https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html>>. Acesso em: 11 jun. 2018.

RUFFINI, Patrick. CHAPTER 3 - How Technology Is Changing the Polling Industry. In: THERRIAULT, Andrew. *Data and Democracy How Political Data Science Is Shaping the 2016 Elections*. Sebastopol, CA: O'Reilly Media. 2016.

SANTOS, Bruna Martins dos e VARON, Joana. Relatório brasileiro para o projeto internacional 'Dados Pessoais e Influência Política', desenvolvido pela Tactical Technology Collective para a plataforma "Our Data, Our Selves". 2018. Coding Rights.

SAUSSURE, Ferdinand de. *Curso de Linguística Geral*. São Paulo: Cultrix, 2006.

SCARVALONE, Daniel. CHAPTER 6 - Digital Advertising in the Post-Obama Era. In: THERRIAULT, Andrew. *Data and Democracy How Political Data Science Is Shaping the 2016 Elections*. Sebastopol, CA: O'Reilly Media. 2016.

SEARLE, John R. *Minds, brains, and programs*. The Behavioral and Brain Sciences. Volume 3, pp. 417-457 1980. Disponível em: <<http://cogprints.org/7150/1/10.1.1.83.5248.pdf>>. Acesso em: 17 jul. 2018.

SILVA, Cícero. *Laboratório de Software Studies* (Estudos Culturais do Software). 2015. Disponível em: <<http://www.ufjf.br/sws/2015/03/30/inicial/>> Acesso em: 09 maio 2018.

SILVEIRA, Sergio Amadeu da. *Tudo sobre todos. Redes digitais, privacidade e venda de dados pessoais*. São Paulo: Edições Sesc, 2017.

SOUZA, Joyce Ariane de. *Saúde dos dados pessoais e o município de São Caetano do Sul*. 2018, 142f. Dissertação (Mestrado em Ciências Humanas e Sociais), Universidade Federal do ABC – UFABC, São Bernardo, SP, 2018.

STOCCO, Mariana. Stilingue recebe aporte de R\$ 3 milhões. *Proxima*, 12 set. 2016. Disponível em: <<https://www.proxima.com.br/home/proxima/noticias/2016/09/12/stilingue-recebe-aporte-de-r-3-milhoes.html>>. Acesso em: 20 maio 2019.

SUGOMORI, Yusuke. *Java Deep Learning Essentials*. Packt Publishing. UK. 2016.

THEOBALD, Oliver. *Machine Learning for Absolute Beginners: A Plain English Introduction (First Edition)* (Locais do Kindle 336-337). Scatterplot Press. Edição do Kindle. 2017.

TERRIAULT, Andrew. *Data and Democracy How Political Data Science Is Shaping the 2016 Elections*. Sebastopol, CA: O'Reilly Media. 2016.

TOMAZELLI, Luiz Carlos. *Marketing Político*. Porto Alegre, RS: Mercado Aberto. 1988.

TORRETA, André. Uso de dados pessoais como instrumento de campanha eleitoral e a persuasão da opinião pública. Palestra apresentada na Câmara dos Deputados, Seminário "Proteção de Dados Pessoais". 2018. Disponível em: <<http://www2.camara.leg.br/atividade-legislativa/webcamara/videoArquivo?codSessao=72720#videoTitulo>>. Acesso em: 19 nov. 2018.

TOZETTO, Cláudia. Eleições 2014 revelam novo cabo eleitoral: o WhatsApp. *VEJA*, 1 nov. 2014. Disponível em: <<https://veja.abril.com.br/tecnologia/eleicoes-2014-revelam-novo-cabo-eleitoral-o-whatsapp/>>. Acesso em: 01 jun. 2019.

TURING, A. M. *Computing Machinery and Intelligence*. *Mind* 49: 433-460. 1950. Disponível em: <<https://www.csee.umbc.edu/courses/471/papers/turing.pdf>> Acesso em: 21 maio 2018.

VINGE, Vernor. *The coming technological singularity: How to survive in the post-human era*. 1993. Disponível em: <<https://ntrs.nasa.gov/search.jsp?R=19940022856>> Acesso em: 29 maio 2018.

WAKKA, Wagner. Facebook compra empresa de inteligência artificial para combater fake news. *Canaltech*, 3 Jul. 2018. Disponível em: <<https://canaltech.com.br/negocios/facebook-compra-empresa-de-inteligencia-artificial-para-combater-fake-news-117176/>>. Acesso em: 27 set. 2018.

XAVIER, Gláucia do Carmo. *Significante e significado no processo de alfabetização e letramento: contribuições de Saussure*. *Revista Cadernos CESPUC*, 2014, n. 25. Disponível em: <<http://periodicos.pucminas.br/index.php/cadernoscespuc/article/viewFile/11089/8904>>. Acesso em: 09 maio 2018.

YOUYOUA, W.; KOSINSKI, M. e STILLWELL, D. *Computer-based personality judgments are more accurate than those made by humans*. 2014. Disponível em: <<http://www.richardbenjamintrust.co.uk/uploads/finalreports/2013/DStillwell.pdf>> Acesso em: 11 jun. 2018.

Entrevistas

AMBRIOLA, Y. Especialista. Entrevista 8 jun. 2019.

HELCER, R. Stilingue. Entrevista 2 maio 2019.

IKEDA, W. Especialista. Entrevista 3 jun. 2019.

KOHLITZ, A. Polis Consulting. Entrevista 15 jan. 2019.

LIMA, A. Elife. Entrevista 25 e 26 fev. 2019.

LOPES, B. Tail. Entrevista 17 jan. 2019.

TORRETA, A. Consultor de Marketing Político. Entrevista 6 fev. 2019.

Evento “Como transformar sua empresa a partir de dados”

SOUSA, M. MarketData. Palestra: O uso da informação como vantagem competitiva. 12 fev. 2019.

APÊNDICE

APÊNDICE A - Informações sobre as Soluções da Polis Consulting

O site da Polis foi acessado em 12 de maio de 2019. De acordo com esse site, além de representar as plataformas de *Social Intelligence: Brandwach, Crimson Hexago, NetBase e v-tracker*, essa empresa também é especializada em soluções de *Business Intelligence (BI)*. Nesse segmento, presta os seguintes serviços: planejamento, desenvolvimento e implantação de soluções para empresas, desde agências de comunicação, marcas ou administração pública. Também realiza terceirização de processos de negócio (BPO), uma prática que melhora a qualidade do serviço e reduz custos. A Polis também atua com monitoramento de redes sociais; gestão de crises; análise de campanhas; identificação de microinfluenciadores; pesquisa e tendências de mercado e análises de concorrência.

A Polis tem em seu site⁴⁶ um fluxo da metodologia adotada para realizar um trabalho de monitoramento. Após o recebimento do *briefing* é realizado um planejamento para definição das regras. Depois dessa etapa, passa-se à fase da configuração das regras necessárias antes de iniciar o monitoramento. Uma vez que o monitoramento é feito, há um tratamento dos dados antes da realização da análise. Após concluída a extração das métricas, é feita a análise e assim que esta é confirmada o processo é finalizado.

Figura 34 - Metodologia seguida para realizar um trabalho de monitoramento



Fonte: Polis Consulting

A seguir, será feita uma descrição das plataformas de *Social Intelligence: Brandwach, Crimson Hexago, NetBase e v-tracker*.

⁴⁶ <http://polisconsulting.com.br/servicos/>

Brandwatch

Segundo o site da Polis, a *Brandwatch*⁴⁷ é uma tecnologia de monitoramento de redes sociais e inteligência de negócios. Foi projetada para analisar milhões de conversas que acontecem diariamente pela Internet. Permite descobrir em tempo real o que estão dizendo sobre uma marca, produtos, concorrentes e o mercado, acessando mais de 80 milhões de fontes em 44 idiomas. Essa tecnologia possui as seguintes soluções: *Brandwatch Analytics*; *Image Insights*, *Audiences* e *Vizia Command Center*. Segue abaixo detalhamento dessas soluções com base nas informações disponibilizadas no site da Polis.

O *Brandwatch Analytics* permite coletar conversações sobre qualquer assunto (marcas, clientes) e obter *insights* importantes. Possui os seguintes diferenciais: coleta histórica; *dashboards* customizáveis; mais de 20 operadores booleanos que permitem maior assertividade das buscas; sistema de NLP (*Natural Language Processing*) para polarização automática das menções; sistema de geolocalização; *dashboards* de influenciadores e dados demográficos; sistema robusto de regras; possibilidade de comparação de desempenho de canais do *Twitter* e *Facebook* e Sistema completo de alertas, incluindo alerta por aumento de volume de menções. O *Brandwatch Analytics* possui o seguinte fluxo: primeiro a solução procura na *web* e coleta todas as postagens, comentários e conversas relevantes em função da sua configuração de busca. Em seguida, a solução segmenta automaticamente os dados em categorias que são relevantes para o negócio ou pesquisa. Após essa etapa, desagrega e analisa os dados para descobrir novos *insights* que ajudem a responder questões complexas. Por fim, tem relatórios automatizados e alertas por *e-mail*.

O *Image Insights* é uma solução de reconhecimento e análise de imagens e logos que permite obter *insights* de consumidores de forma que as equipes de marketing consigam medir o impacto de atividades. Com esta solução é possível: ver as menções das imagens; adicionar categorias e *tags*; entender dados de autores; visualizar informações demográficas; analisar nuvens de tópicos; comparar buscas de imagens diferentes; comparar busca de imagens e busca de texto; entender o sentimento das menções; criar alertas; visualizar menções geolocalizadas; filtrar por idioma e por localidade. Com essa tecnologia também é possível verificar: hábitos de consumo; utilização indevida de marca; associação de marcas; *insights* de conteúdo; posicionamento; planejamento e mídia.

O *Audiences* permite obter informações de forma instantânea sobre qualquer tipo de público que se queira pesquisar para entender e encontrar microinfluenciadores para uma marca ou cliente. Esta solução opera da seguinte forma: primeiro é realizada uma busca em uma base de dados ao vivo com mais de 300 milhões de usuários ativos do *Twitter*.

⁴⁷ <http://polisconsulting.com.br/ferramentas/brandwatch/>

Figura 36 - *Brandwatch* nuvem de palavras por frases na entrevista do Bolsonaro

Nuvem de palavras por frases

Podemos destacar os assuntos mais relevantes a respeito da entrevista do candidato Jair Bolsonaro à Rede Globo:

- Renata Vasconcellos
- Jair Bolsonaro
- Roberto Marinho

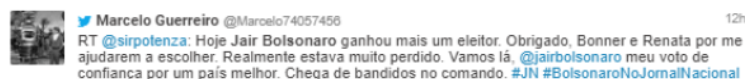


Renata Vasconcellos



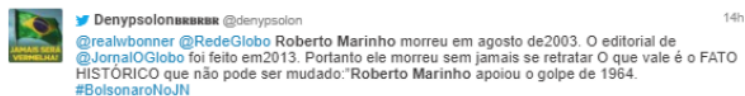
Renata Vasconcellos disse ao candidato Jair Bolsonaro que jamais aceitaria um salário menor que o de um homem para exercer a mesma função.

Jair Bolsonaro



Usuários retweetaram uma menção na qual diziam que Bonner e Renata ajudaram a escolher Bolsonaro como presidente.

Roberto Marinho



Bolsonaro disse que Roberto Marinho apoiou o golpe de 64. Isso fez com que a Rede Globo emitisse uma nota declarando o mesmo apoiou "Revolução de 64 editorialmente.

Fonte: Polis Consulting

Figura 37 - *Brandwatch* evolução de menções na entrevista do Bolsonaro

Evolução de menções

- O pico de menções ocorreu perto das 20 horas, foram **1.905** menções e por volta das 21 horas foram mais **2.059** menções.
- Usuários deram **171** retweets em uma menção da jornalista **Barbara Gancia**. No tweet ela diz que dói ver um candidato com um baixo nível de instrução.
- Outro ponto forte da entrevista foi o livro que Jair Bolsonaro mostrou para as câmeras e segundo ele, ficou conhecido como **“Kit gay”**.



Fonte: Polis Consulting

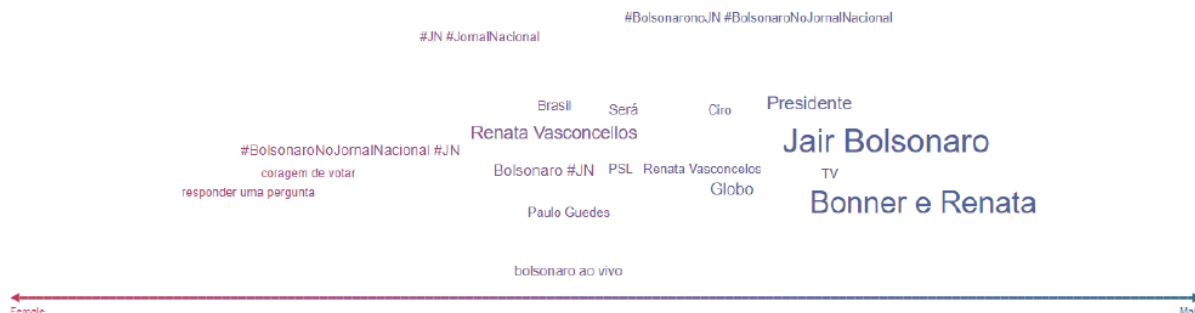
Figura 38 - *Brandwatch* informações demográficas na entrevista do Bolsonaro

Demografia

Os homens (60%) ainda são maioria em relação às mulheres (40), A **Renata Vasconcellos** foi mencionada por ambos os gêneros.

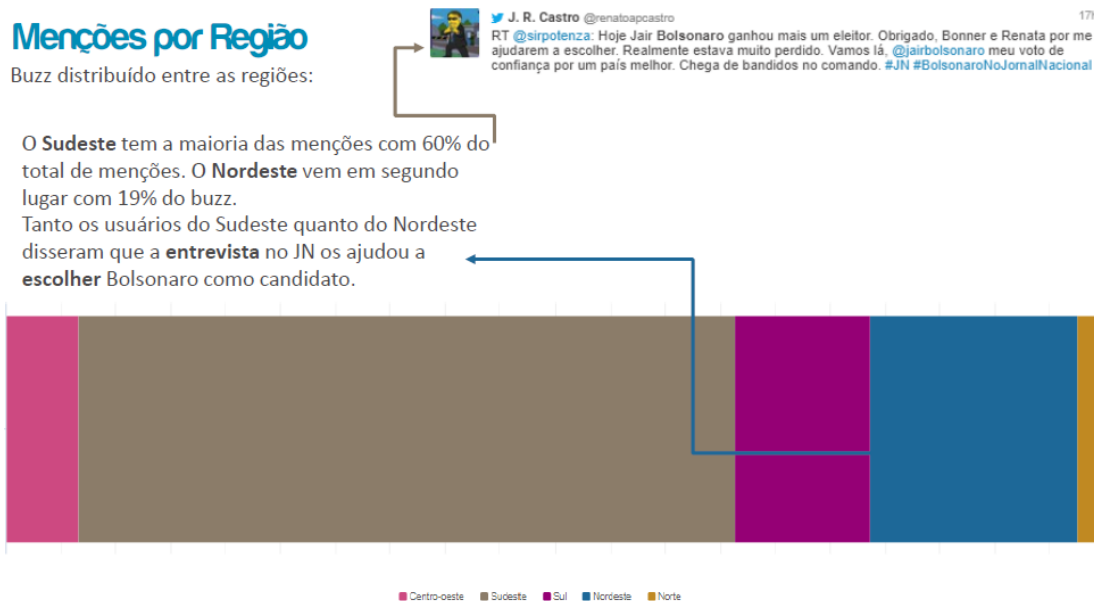


Os homens (60%) ainda são maioria em relação às mulheres (40), mas os assuntos falados foram praticamente os mesmos: “Renata Vasconcellos”, “Bolsonaro”, “JN”.



Fonte: Polis Consulting

Figura 39 - Brandwatch menções por região na entrevista do Bolsonaro



Caso de uso *Brandwatch* Entrevista Marina Silva no Jornal Nacional

Esse caso de uso⁴⁹ fez uma análise das menções a respeito da entrevista com Marina Silva no Jornal Nacional entre 30/08 a 31/08/2018. Foi utilizada a ferramenta de monitoramento *Brandwatch Analytics* para captação e segmentação dos dados durante a sabatina de Mariana Silva, ocorrida no dia 30/08, e sua repercussão no dia 31/08. Os termos usados na busca foram: Marina no JN e #marinanojornalnacional. Fonte de dados: *Twitter*.

Figura 40 - Brandwatch volume por hashtags na entrevista da Marina Silva

⁴⁹ <https://www.slideshare.net/polisconsulting/anlise-das-menes-da-entrevista-de-marina-silva-no-jornal-nacional>

Volume por hashtags

- As hashtags oficiais juntamente com a hashtag “votemarina18” foram destaque na entre os assuntos mais falados.

TOP 3 HASHTAG

61.690 #marinanojornalnacional
3.246 #votemarina18
1.784 #jn



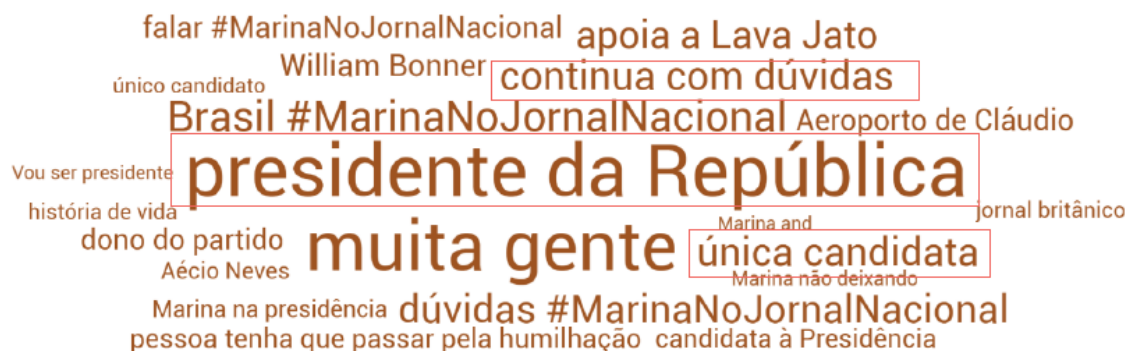
Fonte: Polis Consulting

Figura 41 - *Brandwatch* nuvem de palavras por frases na entrevista da Marina Silva

Nuvem de palavras por frases

Destacamos três, dos assuntos mais relevantes falados no Twitter a respeito da entrevista da candidata Marina Silva à Rede Globo:

- Continua com dúvidas
- presidente da República
- única candidata

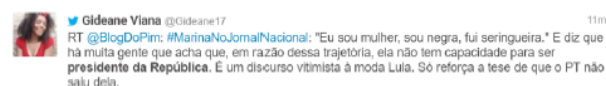


Continua com dúvidas



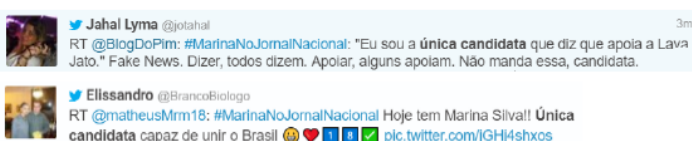
A reação da Marina depois que Bonner insistiu em cobrar uma posição da candidata sobre a reforma da Previdência, gerou uma grande repercussão e virou meme entre os internautas.

Presidente da República



O post do jornalista Felipe Moura onde ele fala sobre a Marina utilizar de um discurso vitimista, esteve entre os mais compartilhados durante a sabatina.

Única Candidata



Marina informou ser a única candidata que apoia a Lava Jato, porém os internautas apontaram como fake News, pois dizem que todos afirmam isso.

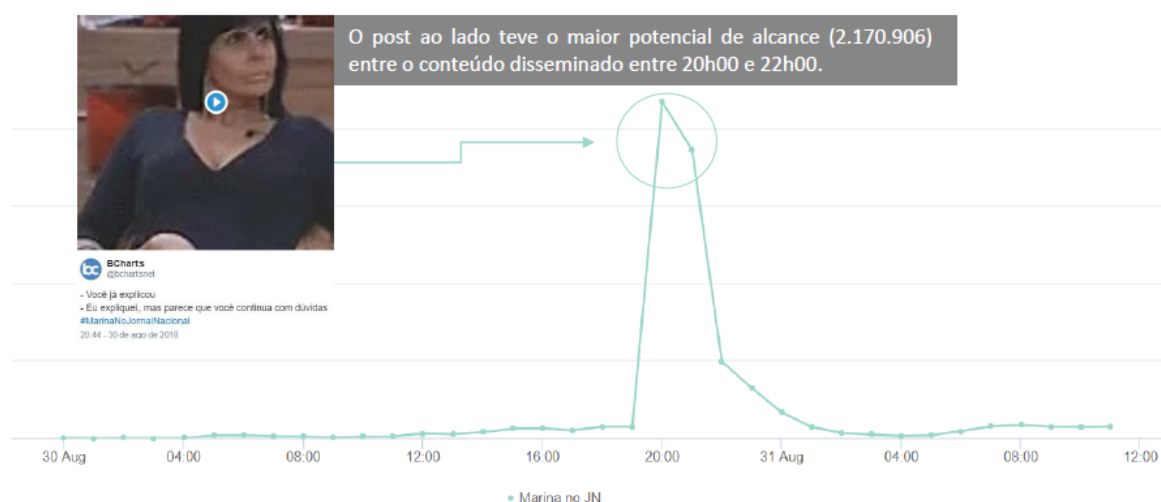
Por outro lado os eleitores de Marina a apontam como a única candidata capaz de unir o país

Fonte: Polis Consulting

Figura 42 - Brandwatch evolução de menções na entrevista da Marina Silva

Evolução de menções

- Durante a sabatina o buzz da candidata chegou a ter um pico de 1007%, se comparado o volume de menções da média dos dias anteriores, o meme gerado pela repostagem de Marina Silva em afronta a Bonner causou muita conversa entre os internautas.

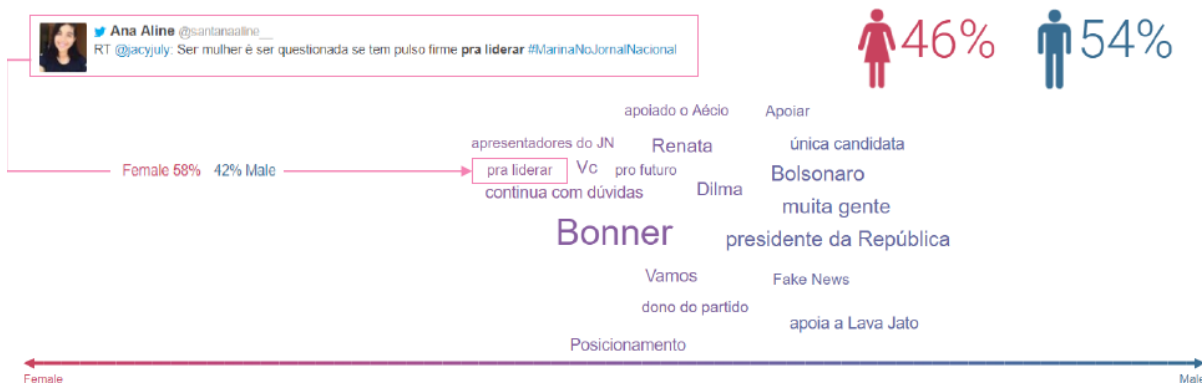


Fonte: Polis Consulting

Figura 43 - Brandwatch informações demográficas na entrevista da Marina Silva

Demografia

- Das quatro sabatinas a de Marina Silva, foi a que mais teve participação feminina no Twitter, os assuntos mais comentados pelas mulheres foram sobre a formato das perguntas realizadas, e que qualquer resposta da candidata iria gerar questionamentos contrários. As internautas compartilharam amplamente o post afirmando que mulher tem que ter pulso firme para liderar.



Fonte: Polis Consulting

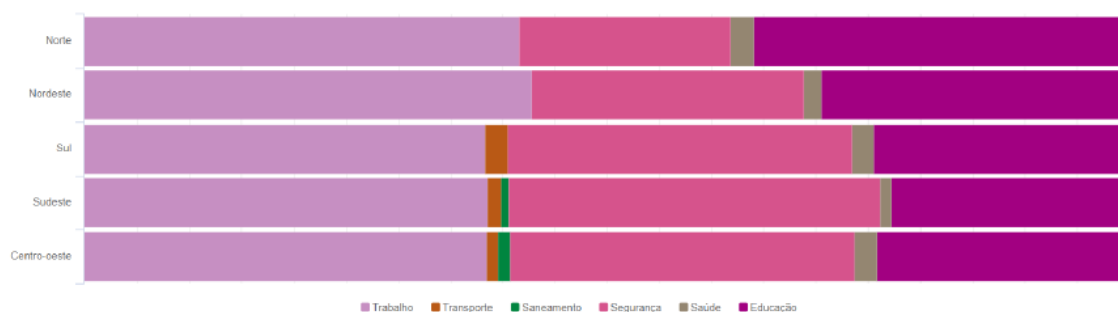
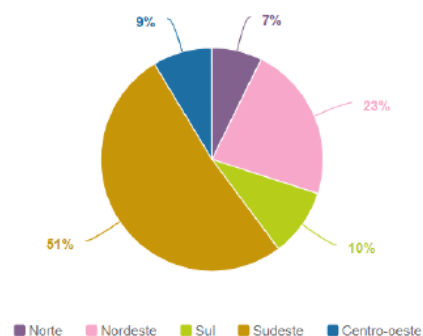
Figura 44 - Brandwatch menções por região na entrevista da Marina Silva

Menções por Região

Buzz distribuído entre as regiões:

O **Sudeste** teve a maioria das menções com 51% do total. O **Nordeste** vem em segundo lugar com 23% do buzz.

“Trabalho” foi a pauta que ficou em destaque entre os assuntos discutidos no Twitter, isso para todas as regiões do Brasil.



Fonte: Polis Consulting

Crimson Hexago

Segundo o site da Polis, o *Crimson Hexago*⁵⁰ é uma solução que permite análise de audiência, análise de marca, inteligência competitiva e análise de tendências. A análise de audiência identifica e compreende públicos, influenciadores e personas interessantes. Na análise de marca é feita uma avaliação da saúde de marca para descobrir o que causa os sentimentos positivos e negativos sobre seus produtos ou serviços. A inteligência competitiva entende o posicionamento de marca, atividades e público-alvo de concorrentes. Já a análise de tendências, identifica as tendências de consumo e entende o cenário de mercado em evolução.

De acordo com o site da Polis, essa solução possui os seguintes diferenciais: Inteligência Artificial que permite descobrir automaticamente *insights* de conversas de consumidores em grande escala usando tecnologias avançadas de IA; análise avançada de imagens para uma visão completa das conversas dos usuários ao incluir imagens para uma análise unificada de conteúdo visual e de texto; biblioteca de dados que permite acesso instantâneo a mais de um trilhão de conversas de consumidores em mídias sociais, fóruns, *blogs*, resenhas e muito mais; e experiência personalizada por meio de *insights* de consumidores com pesquisas instantâneas, recursos visuais intuitivos e painéis personalizáveis.

NetBase

Conforme descrito no site da Polis, a *Netbase*⁵¹ tem tecnologias sofisticadas e inovadoras que revelam as emoções e desejos de uma audiência nas mídias sociais. Ela permite gerenciar reputação em 42 idiomas, impulsionar campanhas e se conectar com um público globalmente. A *Netbase* possui as seguintes soluções: *Instant Search*; Plataforma *Netbase*; *Audience 3D*; *Alertas* e *Command Center*. Segue abaixo detalhamento dessas soluções com base nas informações disponibilizadas no site da Polis.

O *Instant Search* remove a complexidade das ferramentas tradicionais, permitindo que os profissionais busquem e analisem milhões de postagens obtendo *insights* imediatos. A Plataforma *Netbase* viabiliza a obtenção de informações em tempo real gerando mais resultados em suas análises; oferece análise de polaridade, comportamento de compra com resultados de fácil compreensão; possui uma tecnologia única de NLP que lê milhões de *posts* sociais em 42 idiomas.

⁵⁰ <http://polisconsulting.com.br/ferramentas/crimson-hexagon/>

⁵¹ <http://polisconsulting.com.br/ferramentas/netbase/>

A *Audience 3D* entende as conexões de grupos específicos de usuários e tem uma visão ampla da audiência como *lifestyle*, interesses e intenções de compra. O que possibilita a criação de conteúdo altamente interativo para segmentos específicos e criação de engajamento em publicações e mensuração da efetividade de cada campanha.

A *NetBase* possui uma solução abrangente de alertas em tempo real para gerenciar crises ou antecipar riscos em alguma campanha. Oferece cadastro de um alerta específico para notificação quando qualquer evento relevante acontecer. Permite também notificar a ocorrência de qualquer atividade incomum relacionada à marca ou concorrente, com base em histórico.

O *Command Center* é um painel que exibe indicadores e métricas em tempo real para que o usuário tome decisões de forma rápida a fim de entender o que estão falando sobre uma determinada marca e reproduzir conteúdo altamente relevante. Permite acompanhar a resposta dos consumidores e o efeito nos indicadores de negócios. É a forma mais fácil de visualizar o que estão falando sobre uma marca e se conectar com esses consumidores. Ajuda na tomada de decisões de negócios nas mais diversas áreas de uma agência.

Caso de uso Netbase monitoramento Rodrigo Maia

Em 17 de maio de 2017, o jornal *O Globo* publicou uma reportagem informando que o dono da JBS, Joesley Batista, gravou conversa com o presidente Michel Temer, na qual foi discutido o silêncio do deputado cassado Eduardo Cunha (PMDB-RJ). A gravação foi entregue ao Ministério Público Federal. Logo após tomar conhecimento a respeito dessa notícia, Rodrigo Maia encerrou a sessão que estava ocorrendo na Câmara dos Deputados⁵².

Nesse caso de uso⁵³ foi realizado monitoramento para menções referentes à Rodrigo Maia. Foi adotada a seguinte metodologia: utilização da ferramenta *Netbase* para fazer o monitoramento de Rodrigo Maia. Termo utilizado para as buscas: Rodrigo Maia. Idioma: Português. Intervalo: de 11/05 a 18/05/2017 (um dia após a divulgação da matéria). A ferramenta mostra as principais métricas e tendências: quantidade de publicações; quantidade de autores; impressões e sentimento geral.

Figura 45 - Netbase principais métricas e tendências no monitoramento do Rodrigo Maia

⁵² <https://g1.globo.com/politica/noticia/maia-encerra-sessao-da-camara-minutos-apos-noticia-sobre-audio-de-temer.ghtml>

⁵³ <http://www.polisconsulting.com.br/wp-content/uploads/2017/05/RodrigoMaia.pdf>



PRINCIPAIS MÉTRICAS E TENDÊNCIAS

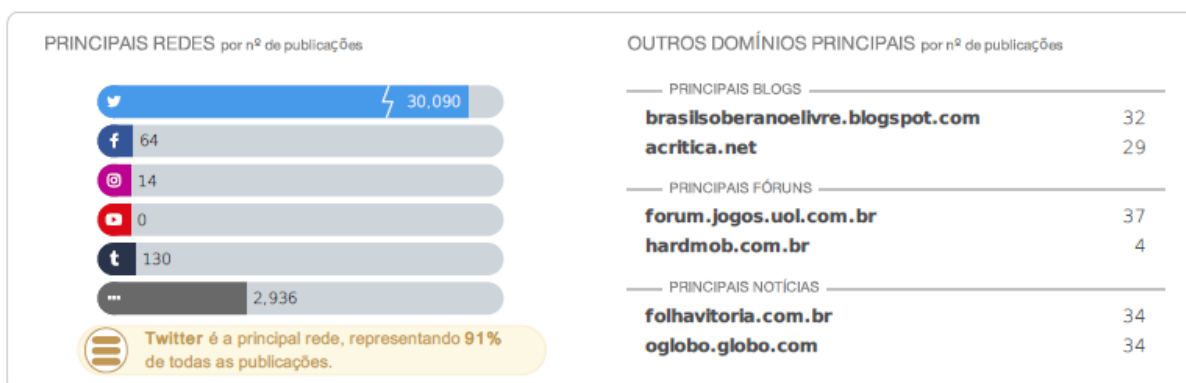
Analisando 33,234 publicações de 11/05/17 até 18/05/17 (GMT+01:00).



Fonte: Polis Consulting

Figura 46 - Netbase onde as pessoas estão falando no monitoramento do Rodrigo Maia

ONDE AS PESSOAS ESTÃO FALANDO?



Fonte: Polis Consulting

Figura 47 - Netbase o que as pessoas estão falando no monitoramento do Rodrigo Maia

SOBRE O QUE AS PESSOAS ESTÃO FALANDO E COMPARTILHANDO?



Fonte: Polis Consulting

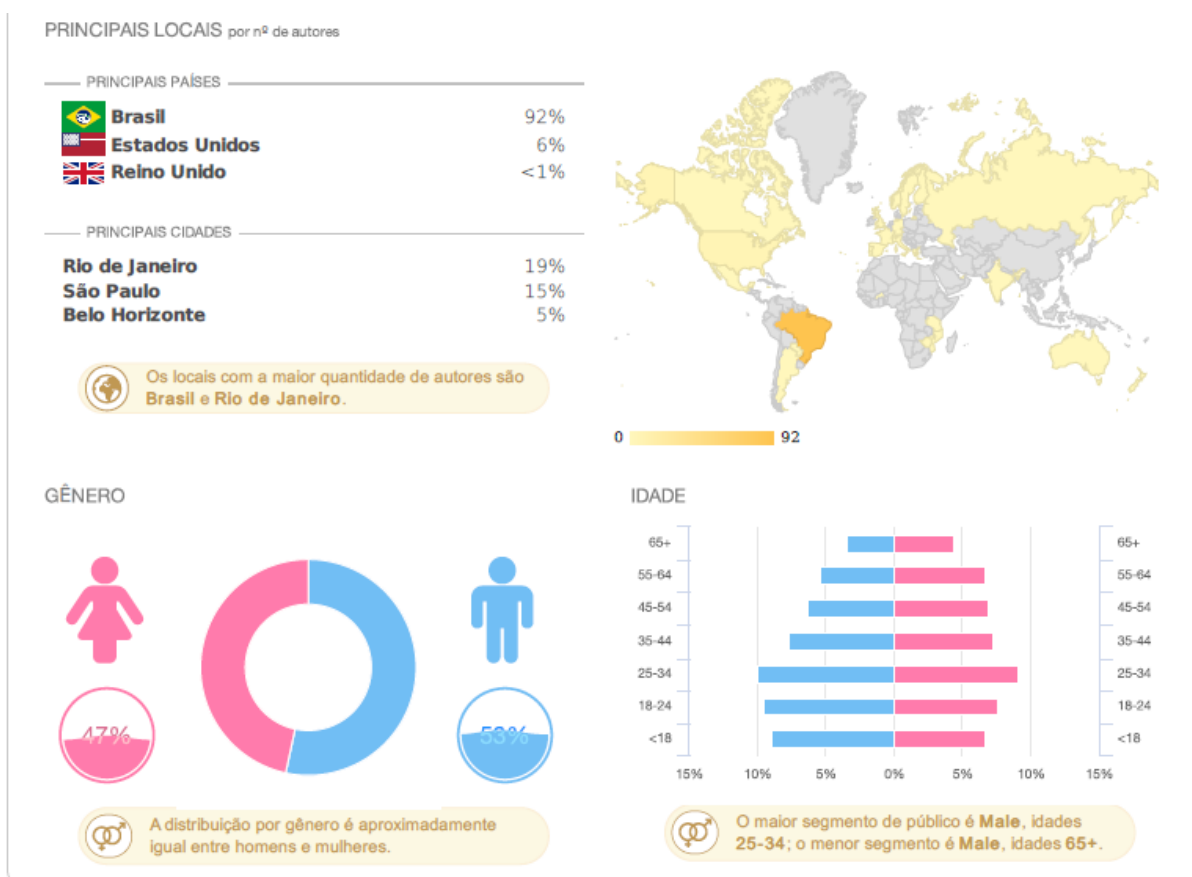
Figura 48 - Netbase quem são os autores no monitoramento do Rodrigo Maia

QUEM SÃO OS AUTORES?



Fonte: Polis Consulting

Figura 49 - Netbase informações demográficas no monitoramento do Rodrigo Maia



Fonte: Polis Consulting

v-tracker

Segundo o site da Polis, a *v-tracker*⁵⁴ é uma plataforma nacional de *social media management*. Possui as seguintes funcionalidades: Monitoramento; SAC 2.0; *Insights*; IA e API. Os *Insights* geram relatórios, *dashboards* e grafos. Possui Inteligência Artificial e NLP (*Natural Language Processing*) para sentimentalização, tagueamento, detecção de *spam*, detecção de gênero pelo nome. Possui API para integração com outras ferramentas/sistema. Segue abaixo detalhamento das soluções dessa plataforma com base nas informações disponibilizadas no site da Polis.

A funcionalidade de Monitoramento da *v-tracker* permite monitorar diversas redes sociais e mídias digitais e possibilita: coleta em mais de 40 serviços, entre redes sociais, portais, jornais, mídias digitais e lojas de apps; coleta automatizada de *dark posts*; regras completas com mais de 40 ações, condições e horários de execução; nuvens de palavras,

⁵⁴ <http://polisconsulting.com.br/ferramentas/v-tracker/>

bio, *hashtags*, *links*, grupos e citações diretamente na listagem de *posts*; exportações em Excel, CSV e imagem; inferência de gênero dos publicadores pelo nome.

O *v-tracker* SAC 2.0 faz atendimento em diversas redes. O SAC 2.0 possui as seguintes funcionalidades: atendimento no *Facebook*, *YouTube*, *Instagram*, *Twitter* e *Google*; abertura manual ou automática de *tickets* e atendimentos; associação e criação de níveis e setores de atendimento, associação de status e prioridades; relatórios customizáveis, relatórios de auditoria com todo o histórico dos *tickets* e interações; relatório de performance de atendentes; respostas pré-cadastradas; envio e recebimento de anexos e *emojis* e agendamento de postagens.

O *v-Tracker Insights* possui os seguintes recursos: relatórios customizáveis; grafos com redes de influência, relação e clusterização de assuntos e *tags*, grupos, *hashtags*, *posts*, respostas, comentários, publicadores e qualificações com exportação para Gephi e visualização no navegador; nuvens de palavras, bio dos publicadores, *links* e *hashtags* com algoritmos customizáveis; mapa de calor de geolocalizações; mapa de imagens mais compartilhadas e linhas de tendências em gráficos com datas.

O *v-Tracker* Inteligência Artificial possui as seguintes tecnologias: Inteligência Artificial em português para sentimentalização, tagueamento e detecção de *spam* (ocorrências com os termos de busca, porém, fora de contexto); modelos genéricos e customizados (taxa de acerto bem maior); modelos evolutivos, que aprendem com o analista; aplicável em itens de monitoramento e SAC e inferência de sexo dos publicadores pelo nome.

Análise do site da Polis

O site da Polis é bem organizado, possui *blog* e uma descrição das soluções e serviços bem clara e de fácil navegação. Ao gerar uma nuvem de palavras, as que mais se destacam são: marketing, social, *blog*, comentários, dados, plataforma, digital como mostra a Figura 50.

Figura 50 - Nuvem de palavras do site Polis Consulting

APÊNDICE B - Informações sobre as Soluções da Tail

O site da Tail foi acessado em 17 de janeiro de 2019. O site possui um *blog* onde foi publicado o texto “Inteligência Artificial para Todos”, no qual Cristiano Nobrega, CEO da Tail faz as seguintes considerações:

Toda startup que pretenda oferecer algum serviço disruptivo dificilmente não estará baseada em uma plataforma de *big data* e algoritmos de inteligência artificial, capazes de aprenderem sozinhos a medida que repetem as tarefas para as quais foram treinados ou programados. A razão é óbvia: máxima produtividade, sem depender de uma operação executada por humanos. Maior escalabilidade, margens de erro irrelevantes ou inexistentes, menor custo. Serviços que tornam-se possíveis e economicamente viáveis graças a máquinas inteligentes. Justamente essa revolução, que se espalha por todos os lados e populariza termos até então desconhecidos pelo grande público como “*machine learning*” e “*deep learning*”, traz à tona um novo tipo de demanda, na qual diversos negócios buscam se reinventar e surfar essa mesma onda (Cristiano Nobrega, CEO da Tail, *online*⁵⁵).

O CEO da Tail também menciona que a solução que estaria mais próxima do alcance de todos é a ferramenta de *machine learning*; disponível na nuvem. Ele cita duas ferramentas do *Google*, o *TensorFlow*, uma biblioteca *open source* e a *Cloud AutoML*, uma suíte de sistemas pré-treinados. Também menciona que o *AWS (Amazon Web Services)* fez parceria com a *Azure* da *Microsoft* para lançarem em conjunto a *Gluon*, uma biblioteca *open source* para algoritmos de *deep learning*.

De acordo com o site, a Tail é uma empresa de *Big Data Analytics*, voltada para soluções de audiência. Foi responsável pelo lançamento da primeira DMP (*Data Management Platform*) para dados *online* e *offline*, em *desktop* e *mobile*, utilizando uma tecnologia 100% proprietária. Resumidamente, a Tail é uma plataforma de dados para *insights* e segmentação de audiência em ambiente digital. No site também há um vídeo falando sobre a empresa⁵⁶. Segue a transcrição:

Você imagina quantos dados são gerados nesse momento? Milhões, bilhões, trilhões, talvez um pouco mais, mas para onde vai e para que serve tudo isso? Para nós que integramos a ciência de exatas com a sensibilidade de humanas, dados se tornam oportunidades. Oportunidades para marcas, para as pessoas e para o mundo. A tecnologia de dados muda, transforma e revoluciona a vida das pessoas. Mostrar o caminho mais curto, recomendar a série de acordo com que você já assistiu ou até mesmo aquele apartamento do outro lado do mundo que você se sente em casa. Talvez a primeira impressão é que todo esse emaranhado de tecnologia como DMP, *Data Provider*, *Data Science*, *Big Data*, ainda é algo muito complexo para o seu negócio. Apague esse dado. Pessoas não são somente números e códigos criptografados. Encontramos vida, comportamentos, sensações, experiências e transformamos isso em

⁵⁵ <https://blog.tail.digital/inteligencia-artificial-para-todos/>

⁵⁶ <https://www.tail.digital/#about>

insights estratégicos para os negócios. Nós estamos aqui para ajudar e simplificar tudo que você conhece e não conhece sobre o assunto. Acreditamos nas boas relações e não só aquela de empresa cliente, mas sim da relação de parceria, customizando estratégias de maneira didática para cada cliente, melhor dizendo, para cada parceiro. Uma relação em que você pode confiar de olhos fechados e computador aberto, para lidar com os dados de maneira correta e transparente. Se tudo parece complexo, confuso, escolhemos descomplicar e solucionar. Se as tecnologias mudam todos os dias, o tempo todo e em todo lugar, somos incansáveis para lhe entregar o melhor. Tail Target Audience & *Insights* Lab (Site da Tail, Quem Somos, *online*).

De acordo com o site, a Tail possui as seguintes soluções: Premium DMP: *Audience Center*; *Onboard* de CRM e Enriquecimento; Monitoramento de Campanha; Dados para Segmentação de Campanha, Personalização de Conteúdo e Ofertas e *Location Data* e *Geo Marketing*. Segue abaixo o detalhamento dessas soluções com base nas informações disponibilizadas no site da empresa.

Figura 51 - Soluções da Tail

SOLUÇÕES TAIL:
DMP, Dados em Mídia Programática e Soluções de Audiência

PREMIUM DMP: AUDIENCE CENTER
A DMP da Tail, através de seu Big Data Analytics, habilita uma central de audiência premium para cada projeto, capaz de criar regras de segmentação para encontrar audiências muito específicas. Clusters de audiência personalizados, e amplificados por Look Alikes, com insights em tempo real, para ativar em Mídia Programática, Portais e Mídias Sociais.

ONBOARD DE CRM E ENRIQUECIMENTO
Conectamos o CRM ou base de cadastros com a DMP, sincronizando cada registro ao seu respectivo Cookie ou Advertising ID, para ativação em Mídia Programática, Rede Sociais e Portais. Apenas pelo cpf ou email, a Tail usa o poder do big data analytics para enriquecer e manter atualizados os registros de um CRM, com dados instantâneos de perfil.

MONITORAMENTO DE CAMPANHA
As campanhas online são uma fonte de dados essencial para se conhecer melhor o perfil das audiências que interagem com sua marca. A Tail pode monitorar suas campanhas para trazer insights valiosos das audiências impactadas, e os perfis que clicaram no anúncio, através de relatórios detalhados e consolidados por campanha ou por canal.

DADOS PARA SEGMENTAÇÃO DE CAMPANHAS
Uma oferta de dados exclusivos para segmentar campanhas online. Dados do comportamento de navegação (Browsing Behavior), e um Data Marketplace completo, com dezenas de segmentações prontas para uso. Demografia, interesses, estilos de vida, intenção de compra e afinidades. Tudo pronto para uso nas principais DSPs.

PERSONALIZAÇÃO DE CONTEÚDO E OFERTAS
Uma experiência relevante para os visitantes de seu website é chave para se obter máximo engajamento e conversão. A Tail Showcase é uma Vitrine Inteligente, capaz de apresentar o conteúdo ou oferta certos para a audiência que está visitando uma landing-page naquele exato instante, baseado em tudo que a Tail sabe previamente sobre cada perfil.

LOCATION DATA & GEO MARKETING
O deslocamento das pessoas no mundo real e a utilização de Apps são aliados poderosos para se entender melhor o perfil das audiências. Com dados exclusivos de aplicativos instalados em smartphones, a Tail viabiliza segmentações de audiência por Geo Behavior e App Behavior, com insights essenciais para uma estratégia de Marketing vencedora.

*One Stop Shop
de Big Data Analytics
para Soluções de
Audiência*

SOLICITE UMA DEMO DA TAIL

Fonte: Site da Tail: < <https://www.tail.digital/#solutions>>

O Premium DMP: *Audience Center* é a DMP da Tail, que por meio de seu *Big Data Analytics*, habilita uma central de *audiência premium* para cada projeto, capaz de criar regras de segmentação para encontrar audiências muito específicas. Clusters de audiência personalizados, e amplificados por *Look Alikes*, com *insights* em tempo real, para ativar em Mídia Programática, Portais e Mídias Sociais.

O *Onboard* de CRM e Enriquecimento permite conectar um CRM ou base de cadastros com a DMP, sincronizando cada registro ao seu respectivo *cookie* ou *advertising ID*, para ativação em mídia programática, redes sociais e portais. Apenas pelo CPF ou e-

mail, a Tail usa o poder do *Big Data Analytics* para enriquecer e manter atualizados os registros de um CRM, com dados de perfil do usuário.

O Monitoramento *online* das campanhas é uma fonte de dados essencial para se conhecer melhor o perfil das audiências que interagem com uma marca. A Tail pode monitorar campanhas para trazer *insights* valiosos das audiências impactadas, e os perfis que clicaram no anúncio, mediante relatórios detalhados e consolidados por campanha ou por canal.

Conforme relatado no site, a Tail também oferta dados exclusivos para segmentação de campanhas *online*. Dados do comportamento de navegação (*browsing behavior*), e um *Data Marketplace* completo, com dezenas de segmentações prontas para uso, como segmentação demográfica, interesses, estilos de vida, intenção de compra e afinidades. Tudo pronto para uso nas principais DSPs.

A Personalização de Conteúdo, segundo a Tail, permite uma experiência relevante para os visitantes de *website*. Ela é a chave para se obter máximo engajamento e conversão. Esta solução da Tail é uma vitrine inteligente, capaz de apresentar o conteúdo ou oferta certos para a audiência que está visitando uma *landing-page* naquele exato instante, baseado em tudo que a Tail sabe previamente sobre cada perfil.

O deslocamento das pessoas no mundo real e a utilização de *Apps* são aliados poderosos para se entender melhor o perfil das audiências. Com dados exclusivos de aplicativos instalados em *smartphones*, a Tail viabiliza segmentações de audiência por *Geo Behavior* e *App Behavior*, com *insights* essenciais para uma estratégia de Marketing vencedora.

Análise do site da Tail

O site da Tail é bem organizado, possui *blog* e uma descrição das soluções e serviços bem clara e de fácil navegação. Por conta do uso massivo de dados pessoais, há destaque para a Política de Privacidade e o Aviso de *Cookies* que estão muito realçados no site. Ao gerar uma nuvem de palavras, as que mais se destacam são: audiência, dados, DMP, perfil, data, campanhas, segmentação, *behavior* e *insights* como mostra a Figura 52.

Figura 52 - Nuvem de palavras do site Tail

usuários e clientes possam ter acesso às informações, além de poder editá-las ou solicitar a exclusão dos dados pessoais. De acordo com a Tail, a finalidade dos dados coletados é legítima, específica e conhecida do titular dos dados coletados e processados. Detalhes sobre os diferentes usos dos dados dos usuários estão descritos na Política de Privacidade (*Privacy Policy*)⁵⁸.

Na Central de Privacidade há uma série de perguntas. Foi feita uma análise dessas perguntas e respostas. Segue um resumo das mais relevantes.

O que são dados pessoais? Dados pessoais são aqueles que podem identificar ou tornar identificável uma pessoa como, por exemplo, nome, RG, CPF, endereço, *cookies*, ou informações referentes a uma pessoa, como sua localização, preferências culturais, gostos, interesses.

Quais regulações de proteção de dados se aplicam aos serviços prestados pela Tail? As principais regulações que se aplicam aos serviços prestados pela Tail são: GDPR, LGPD, Marco Civil da Internet (MCI), Decreto 8.771/16 e Código de Defesa do Consumidor (CDC).

No site da Tail há a opção *Privacy Dashboard* para visualizar quais são os dados coletados e também editar e/ou remover esses dados. Os dados coletados são os seguintes: Gênero (Feminino e Masculino); Faixa Etária (Infantil, Teen, Jovem Adulto, Adulto e Idoso); Nível de Renda (Alta Renda e Baixa Renda); Times de Futebol (lista com todos os times de futebol do Brasil) e Interesses (Agronegócio, Arte, Artes Marciais, Automobilismo, Automóveis, Basquetebol, Beleza, Casa e Jardim, Ciclismo, Cinema, Corrida, Corrida de Aventura, Cozinha e Culinária, Dança, Educação, Empregos, Equestres, Família, Famosos, Finanças, Fotografia, Futebol, Games, Golf, História em Quadrinhos, Horóscopo e Esoterismo, Humor, Imóveis, Infantil, Informática, Jardinagem, Jurídico, Literatura, Mergulho, Meteorologia, Moda, Musculação, Música, Natação, Negócio, Pet, Política, Rúgbi, Saúde, Sexo, Skate, Surf, Teatro, Tecnologia, Televisão, Trânsito, Tênis, Vela, Viagens e Vôlei).

Foi feita uma análise da Política de Privacidade (*Privacy Policy*) da Tail. Segue resumo dos principais pontos. Sua última atualização foi realizada em 12/09/2018.

A Tail é filiada ao IAB-Brasil (*Internet Advertising Bureau*), que recomenda boas práticas e princípios éticos a serem seguidos pelo mercado em consonância com o Marco Civil da Internet e o Código Internacional NAI (*Network Advertising Initiative*).

A Política está organizada nos seguintes itens: 1) Que tipos de dados a Tail coleta; 2) Com quem a Tail compartilha os dados; 3) Transferência internacional de dados; 4) Quais são os direitos dos titulares de dados; 5) Por quanto tempo os dados serão armazenados; 6)

⁵⁸ <https://www.tail.digital/privacy-policy/>

Como funciona a Segurança da Informação na Tail; 7) Como falar com o DPO da Tail; 8) Mudanças na Política. Para esta pesquisa foram analisados os itens 1, 2 e 3.

Em relação ao item 1, a Tail coleta dados de quem visita o site deles, utiliza o *software* e os serviços dos clientes que utilizam o *software* da Tail. Também são armazenadas informações dos contatos realizados com os usuários, incluindo conteúdos baixados a partir de páginas da Tail e interações via *e-mail*. A coleta dos dados é segmentada em três partes: 1) Visitantes do site Tail; 2) Clientes Tail e 3) Usuários dos sites dos Clientes Tail.

Quando um visitante acessa o site da Tail é inserido um *cookie* no navegador por meio do *Google Analytics* e do TailDMP. O primeiro é usado para identificar quantas vezes uma pessoa retorna ao endereço e como navega no site; o segundo verifica o perfil e preferências inferidas, baseadas no comportamento de navegação. Podem usar *cookies* de sessão (que expiram depois de fechar o navegador) e *cookies* persistentes (que permanecem no computador até serem excluídos). Também são coletados os seguintes dados durante a navegação: IP, data e hora de acesso, localização geográfica, fonte de referência, tipo de navegador, duração da visita e páginas visitadas.

Quanto aos dados de clientes da Tail, são coletadas, além de dados de navegação mencionados na parte de usuários visitantes, outras informações submetidas pelo cliente na plataforma. Quando um cliente se cadastra para ter acesso a algum dos serviços e/ou produtos, são coletados os seguintes dados: nome, *e-mail*, senha, endereço, país, estado e cidade, dados da pessoa jurídica (nome empresarial, RFC e *e-mail*), dados de faturamento (nome, banco, agência, conta, *swift*, IBAN, ABA e endereço), dados para fins de marketing (suas preferências de marketing e informações sobre seu CRM). Para prestar os serviços, a Tail cruza os dados que coletam com outros dados que obtém sobre o mesmo titular nas bases de dados, inclusive os dados dos visitantes dos sites que o cliente cadastrar nas plataformas da Tail. Em relação aos dados pessoais de outros titulares submetidos pelos clientes, estes devem garantir que obtiveram o consentimento dos titulares ou se utilizaram de outra base legal legítima para realizar a coleta desses dados.

Quando um usuário utiliza um aplicativo de um cliente da Tail, via *smartphone* ou *tablet*, a Tail pode obter os dados desse usuário a partir do SDK dele que utiliza o Identificador Eletrônico de Publicidade do dispositivo (“ID de publicidade” do sistema operacional, como o *Android Advertising ID – AAID* e o *Apple ID for Advertising – IDFA*). Além de dados de navegação e utilização da aplicação, também podem ser coletados: dados sobre o dispositivo (modelo, fabricante, sistema operacional, operadora de telecomunicações, navegador móvel, tipo e velocidade da conexão com a internet, IP da conexão, fuso horário GMT e lista de aplicativos instalados); e dados de sensores do dispositivo (dados de sensores de *Wi-Fi*, GPS, *Bluetooth*, acelerômetro, magnetômetro, giroscópio e de proximidade quando esses acessos estiverem disponíveis no aplicativo

parceiro). A Tail pode receber dados anonimizados dos clientes, mediante técnicas de encriptação, que não permitem identificar o usuário no mundo real. A empresa também informa que não coleta e nem recebe de seus clientes e parceiros dados “sensíveis”, ou seja, dados pessoais que revelem orientação religiosa, política ou sexual, convicção filosófica, origem racial ou étnica, bem como informações sobre saúde.

Sobre o item 2, a Tail pode vir a compartilhar os dados coletados com as seguintes empresas: anunciantes ou agências atuando por conta e ordem de anunciantes; parceiros; *Analytics*; e para resguardar e proteger direitos da Tail.

Em relação ao item 3, a Tail coleta e transfere dados pessoais coletados no Brasil para países localizados na União Europeia e para os EUA. Essa transferência ocorre para parceiros da Tail que atuam no processamento de dados pessoais, e envolve apenas empresas parceiras da Tail que demonstraram estar em processo de conformidade ou em concordância com a GDPR e com as leis setoriais brasileiras. A Figura 53 mostra essa lista de parceiros.

Figura 53 - Lista de parceiros da Tail



Fonte: Site da Tail <<https://tail.digital/data-partners/>>

APÊNDICE C - Informações sobre as Soluções da Elife

O site da Elife foi acessado em 16 de maio de 2019. Segundo o site, a Elife é uma consultoria global criada em 2004. Especializada em inteligência de mercado e gestão de relacionamento digital, atua no Brasil, México, Espanha e Portugal. Em janeiro de 2019, a Elife apresentou sua nova marca para celebrar, não só os 15 anos de existência, como também a união dos pilares tecnológico e humano, com a eliminação do ponto que separava as palavras *eletronic* e *life*. A nova marca apresenta também a mudança do foco da empresa, que agora visa à gestão da experiência do consumidor, fornecendo serviços de inteligência de dados, além da gestão e automação do relacionamento no ambiente digital⁵⁹.

Figura 54 - Nova marca da Elife em 2019 e antiga de 2009



Fonte: Site da elife <<https://elife.com.br/index.php/2019/01/17/elife-apresenta-sua-nova-marca/>>

Segundo Graças (2017c), a Elife passou a adotar em seus serviços *premium* tecnologias de inteligência artificial, como processamento de linguagem natural e redes neurais; e com base nessas tecnologias está desenvolvendo aplicações para todos os seus clientes de inteligência de mercado e gestão de relacionamento com o consumidor.

De acordo com Graças (2017c), o primeiro segmento a adotar esse tipo de tecnologia foi a inteligência de mercado. Em 2016 foi apresentada a solução de identificação automática de logos. Segundo Graças, isso resolveu um antigo problema da área de monitoramento de redes sociais: “A invisibilidade das marcas em imagens compartilhadas” (GRAÇAS, 2017c, *online*). Se a marca não fosse escrita no texto do *post*, não era possível identificá-la em fotos ou vídeos.

A Elife também adotou inteligência artificial na área de atendimento ao consumidor. Para a criação de *bots* de autoatendimento para a plataforma *Facebook Messenger* usou a técnica de redes neurais. Com isso, os *bots* se tornaram mais eficientes em entender as demandas do consumidor, além de prever o contexto de uma demanda facilmente (GRAÇAS, 2017c, *online*).

Segundo Graças (2017c), os novos serviços baseados em Inteligência Artificial já estão disponíveis no portfólio de serviços da Elife.

⁵⁹ <https://elife.com.br/index.php/2019/01/17/elife-apresenta-sua-nova-marca/>

As novas técnicas de inteligência artificial e o uso de APIs de parceiros como *Amazon*, *Google* e *Facebook* permitem aplicações inimagináveis há 10 anos atrás. Estamos combinando o que já fazemos de melhor em inteligência e Social CRM para oferecer aos nossos clientes corporativos soluções criativas que aumentem a produtividade com baixo custo (Vitorino, apud Graças, 2017, *online*).

A Elife oferece três tipos de soluções: *Social Intelligence*, Social CRM e *Buzzmonitor*, conforme mostra a Figura 55. Segue abaixo detalhamento dessas soluções com base nas informações disponibilizadas no site da Elife.

Figura 55 – Soluções da Elife

The image shows a screenshot of the Elife website's 'SERVIÇOS' (Services) page. The page features a navigation bar at the top with links for 'SERVIÇOS', 'BLOG', 'NOTÍCIAS', 'PAPERS E EBOOKS', 'NEWSLETTER', and 'CONTATO'. Below the navigation bar, the word 'SERVIÇOS' is prominently displayed. Three service categories are presented in a grid layout, each with an icon, a title, a brief description, and a list of services. Each category also has a 'SAIBA MAIS' (Learn More) button. A WhatsApp icon is visible in the bottom right corner of the page.

SOCIAL INTELLIGENCE	SOCIAL CRM	BUZZMONITOR
<p>DIVERSAS FONTES DE DADOS DIGITAIS COMBINADAS PARA COMPREENSÃO DA AUDIÊNCIA DO CONSUMIDOR, SUA JORNADA E PERCEÇÃO DE MARCA.</p> <ul style="list-style-type: none"> • Monitoramento de redes sociais. • Gestão de crises. • Identificação e gestão de influenciadores. • Tendências e jornada do consumidor. • Pesquisa de mercado. • Consumer-centered Monitoring. • Brand Logo Detection. • Social Wall. 	<p>GESTÃO DE RELACIONAMENTO EM MEIOS DIGITAIS: E-MAIL, CHAT E REDES SOCIAIS DE FORMA INTEGRADA.</p> <ul style="list-style-type: none"> • Atendimento e relacionamento digital. • Cobertura de campanhas e ações promocionais. • Consultoria para gestão de Social CRM. • Gestão de bots e chats. • Atendimento multicanal. • Serviços de concierge digital. 	<p>UMA SUÍTE DE SOLUÇÕES DE SOFTWARE COM O MELHOR DA CONSULTORIA ELIFE PARA GESTÃO INTERNA DE ESTRATÉGIAS VENCEDORAS EM SOCIAL MEDIA.</p> <ul style="list-style-type: none"> • Social Listening. • Social CRM e SAC 2.0. • Social Analytics. • Social Ads. • Social News. • Gestão e monitoramento de influenciadores.

Fonte: Site da Elife <<https://elife.com.br/>>

A solução de *Social Intelligence* oferecida consiste em monitoramento de redes sociais, pesquisas de mercado, identificação e gestão de influenciadores, gestão de crises, tendências e jornada do consumidor. Além disso, a Elife oferece outras soluções com foco em inteligência de mercado como: *Brand Logo Detection*, *Social Wall* e *Consumer-centered Monitoring*.

Brand Logo Detection

O *Brand Logo Detection* (detecção de logo de marca) é uma tecnologia que identifica logotipos em fotos publicadas em rede sociais com imagens indexáveis, mesmo que não seja citado o nome da marca na descrição ou em *hashtags* e marcações. Lima (2017a) explica que essa tecnologia permite aumentar as oportunidades de interação em tempo real

ao ampliar buscas por *hashtags* que remetem ao domínio da marca, mesmo que a marca não seja citada. É uma ferramenta que explora a jornada do consumidor, permitindo criar índices de compartilhamentos sobre a marca ou concorrentes. Com isso é possível encontrar influenciadores que estão usando determinada marca, mas sem citá-la (LIMA, 2017a).

Casos de Uso *Brand Logo Detection*

Dentre os casos de uso do *Brand Logo Detection*, aqui será destacado o da cerveja. A Elife analisou 30 mil imagens publicadas no *Instagram* durante o Carnaval para descobrir qual a marca de cerveja mais consumida.

A elife analisou 30 mil imagens publicadas no *Instagram* durante o Carnaval buscando por logotipos em fotos da rede e descobriu que cerveja é a bebida mais consumida nos festejos, mas houve espaço também para outras bebidas e até para quem preferiu passar a festa popular assistindo à *Netflix* (LIMA, 2017b, *online*).

Segundo Lima (2017b), a tecnologia analisa mais de 1 milhão de imagens por vez sem a ajuda humana, expandindo a possibilidade de engajamento e as buscas no canal para além de termos associados à marca. O domínio das pesquisas é ampliado para momentos em que a marca está presente, mas não é citada. Exemplo: monitorando *#happyhour*, *#churrasco*, *#balada* no *Instagram*, é possível criar um relatório de *share of voice* das marcas de cervejas que mais aparecem nas imagens.

Por meio de busca de *hashtags* relacionadas ao Carnaval (*#carnaval*, *#carnaval2017*, *#carne* e *#carne2017*), a Elife gerou um banco de dados de 30 mil imagens, de fotos compartilhadas no *Instagram*. Depois utilizou seu algoritmo de reconhecimento de logos para identificar e analisar nessas fotos as marcas que mais apareceram durante o carnaval. Graças a essa tecnologia, não foi necessário citar a marca na postagem, bastando apenas que a marca estivesse presente na foto compartilhada. Dentro das imagens, 2.222 mil tiveram algum logo detectado. O resultado mostrou que as bebidas foram o primeiro grupo de imagens que mais trouxeram resultados. Cerveja foi a categoria mais citada, com 62,4% dos logos identificados. Entre as cervejas, destaque para Heineken, com 20% das imagens, e para Skol, com 11%. Budweiser, Brahma e Eisenbahn também aparecem entre as cinco primeiras colocadas em número de menções, respectivamente com 6, 5 e 4%. Este resultado foi consolidado em um gráfico com mostra a Figura 56.

Figura 56 - Gráfico com resultado da análise de imagens no Carnaval

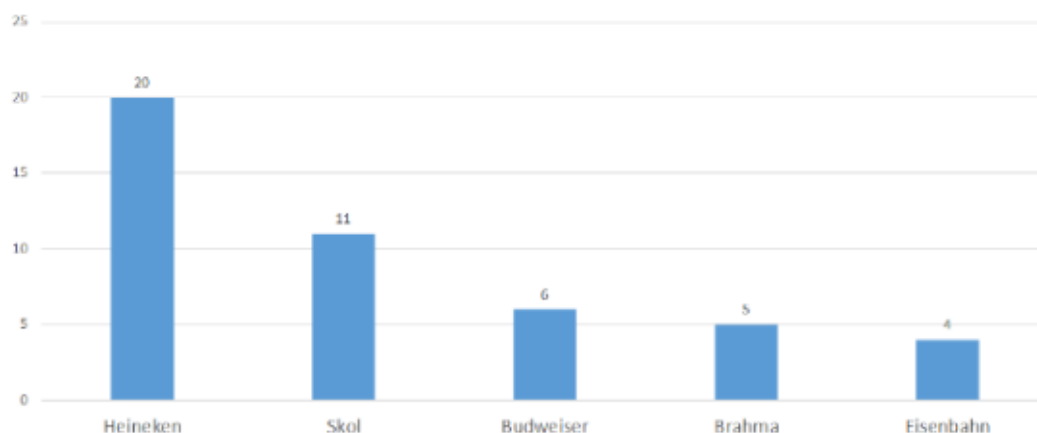
número de seguidores de 2 milhões, 5 e 4% respectivamente.

elife

SERVIÇOS ▾

BLOG *elife* NC

As 5 cervejas mais fotografadas – em porcentagem



Fonte: Lima (2017b)

Social Wall

O *Social Wall* é uma solução que monitora *hashtags*, *check-ins* ou termos de busca e entrega dados em interfaces customizadas especialmente para eventos e ações ao vivo.

Consumer-centered Monitoring

O *Consumer-centered Monitoring* é uma metodologia de monitoramento para a criação de perfis psicográficos a partir de informações e atividades compartilhadas nas mídias sociais pelos consumidores.

Buscando compreender padrões de consumo em diversos segmentos, a Elife desenvolveu uma nova metodologia chamada *Consumer Trends*. De acordo com Lima (2017b), esta metodologia combina técnicas psicográficas, como “*life events*”, valores, atitudes, interesses e estilos de vida dos consumidores com dados de redes sociais para compreender os novos desafios de consumo das marcas. A metodologia da Elife de estudos setoriais possibilita que a segmentação do público estudado vá além daquela tradicionalmente adotada na pesquisa *offline*, que parte de traços sócio-demográficos como renda, idade ou região do entrevistado.

Somada a essas novas possibilidades de segmentação de público, também é utilizada uma metodologia exploratória de análise de dados no ambiente em que este aprofundamento é potencializado: o discurso espontâneo de redes sociais. Um mergulho profundo na realidade do seu público sem pré-concepções de questionários ou roteiros. A seguir, um exemplo de monitoramento utilizando o *Consumer-centered Monitoring*.

Caso de Uso *Consumer-centered Monitoring*

A Figura 57 mostra um estudo feito para levantar a diferença entre os consumidores das diversas redes de supermercados brasileiros. Os consumidores foram classificados de acordo com os seguintes perfis: *The Intellectual*, *The Emotional*, *Internet Geek*, *The TV addicted* e *The Relaxed*. Nesse estudo, os consumidores classificados como “*The Intellectual*” frequentam o Pão de Açúcar, já os classificados como “*The Relaxed*” preferem o Makro.

Figura 57 - Exemplo de monitoramento de perfis psicográficos

Mapeando o perfil de quem faz check-in em supermercados.

BRIEFING











O MIT (Instituto de Tecnologia de Massachusetts) queria saber quais as diferenças entre os consumidores das diversas redes de supermercados brasileiros?

PROCESSO

Com a metodologia *Consumer-centered*, coletamos 5 mil usuários que fizeram check-in em cada um dos mercados estudados e levantamos suas últimas 200 publicações para traçar, assim, o perfil de cada um deles.

RESULTADO

Percebemos, por comportamentos relatados, os diferenciais de renda e de idade de cada supermercado, bem como as diferenças de dia, horário de visita e produtos mais relatados por cada um dos perfis.

	Attitude	Frequent words	Gender*	Tweet times
	The Intellectual Tweet about politics, economy, news, finance. Love Corinthians.	government, news, employment, corruption		Lunch and end of day
	The Emotional Tweet about emotions, music, relationship. Like Beyoncé.	Love, flower, happy, missing, music, night		Saturday night
	The Internet Geek Heavy users of tweet. Tweet a lot in English and internet slangs	follow, iphone, tech, android, netflix, microsoft		Sat and Wed (soccer)
	The TV Addicted No clear profile, likes soap operas and TV related news	TV, Claudia Leite, soap opera,		Night
	The Relaxed Light users. No clear profile, but tweets about Makro and other supermarkets	picture, artist, gym, university, makro, barbecue		No preferred time

Fonte: Site da Elife <<https://elife.com.br>>

Social CRM

O Social CRM faz a gestão de relacionamento em meios digitais como *e-mail*, *chat* e redes sociais de forma integrada. Engloba as seguintes funcionalidades: atendimento e relacionamento digital; cobertura de campanhas e ações promocionais; consultoria para gestão de Social CRM; gestão de *bots* e *chats*; atendimento multicanal e serviços de *concierge* digital. Dentre as soluções ofertadas nesse segmento, merece destaque o serviço de *E-MAIL SAC* inteligente, responsável por coletar o perfil social dos consumidores que é salvo em um banco de dados e se mantém sincronizado com as redes sociais. A partir dos dados coletados, é feita uma análise e segmentação dos usuários baseada em suas preferências, faixa etária, localização; posteriormente, esses dados do perfil social são utilizados para criar uma comunicação personalizada. O *Facebook Bots* possui serviços de *Bots* de serviços para pedir táxi, enviar flores, reservar hotel. *Bots* de compras para realizar compras diretamente pelo *Facebook Messenger*. *Bots* de atendimento que acessa o FAQ da marca ou tira dúvidas pontuais.

Casos de Uso com *Chats Bots*

Graça (2017a) cita três tipos de *bots* que permitem realizar atendimento ao cliente: 1) *Bots concierge* que recebe recomendações sobre serviços ou produtos; 2) *Bots de Serviços* que permitem, por exemplo, chamar um táxi, enviar flores, reservar um hotel e 3) *Shopping Bots* que fazem as compras diretamente pelo *Facebook Messenger*.

Em 2016, a Elife fez um estudo referente aos hábitos e comportamentos dos consumidores e constatou que mais da metade dos entrevistados usam o *Facebook* para interagir com as organizações. Utilizar *chatbot* neste ambiente é uma ótima maneira de relacionamento entre marca e consumidor (GRAÇAS, 2017b).

O importante no primeiro momento é conhecer a jornada do consumidor, mapear os pontos de contato (*touchpoints*) que o consumidor utiliza e criar aplicações que resolvam problemas simples, mas que principalmente aqueles que demandem maior custo ou demanda de atenção ao consumidor (GRAÇAS, 2018a, *online*).

Em julho de 2017 a Elife lançou um *chatbot* para o Hospital Israelita Albert Einstein. Após o lançamento, na segunda quinzena do mês, o *bot* já havia enviado mais de 1,9 mil respostas aos consumidores. Em agosto, o volume de respostas subiu para 2,3 mil. De acordo com Graças:

Para atingir esta produtividade, seriam necessários 5 profissionais trabalhando em média 6h por dia durante 22 dias úteis. Além da economia de tempo, os profissionais envolvidos puderam ser realocados a questões mais complexas do atendimento ao cliente (GRAÇAS, 2017b, *online*).

O *chatbot* do Hospital Israelita Albert Einstein tinha as seguintes opções de serviços: agendamento de exames, vacinas, horário de funcionamento, divulgação de vagas, entre outras opções. Esse projeto durou cerca de três meses e considerou as seguintes etapas: definição do objetivo do *bot* junto com o cliente; produção da árvore de decisões do *bot*, alinhamento com todas as áreas da empresa para consolidar as informações e funcionalidades, desenvolvimento de tecnologia, período de testes e correções; e otimização de investimento, visto o custo considerável em manter uma operação de atendimento humano nestas proporções.

Além do protótipo para o hospital Albert Einstein, a Elife também implementou um *chatbot* para auxiliar o processo de compra de tintas da marca Suvinil e criou junto com a agência Y&R Madrid o *chatbot* para a Opel Espanha, um *bot* que interagia com os internautas de forma dinâmica e divertida como se fosse o novo carro. Para Graças (2017b), a inteligência artificial possibilitará criar *bots* capazes de “processar a linguagem natural do ser humano” e assim realizar interações cada vez melhores (2017b, *online*).

meio de publicidade digital; identificação e ativação de influenciadores; *e-mail* marketing e disparo de *newsletters*. A Elife também considera os dados pessoais que os clientes e consumidores finais fornecem para ela, que são: nome e sobrenome; endereço; telefone; nome da empresa; *e-mail*; dados de conexão (*cookies*) e identificador de publicidade do telefone. A Elife coleta informação para fornecer serviços para os seus clientes. Segundo a política de privacidade, a Elife não vende, aluga ou transmite informações pessoais e dados privados com ou para outras entidades. Os dados podem ser usados em duas esferas de ação: uso geral e marketing.

APÊNDICE D - Informações sobre as Soluções da Stilingue

O site da Stilingue⁶¹ foi acessado no dia 19 de maio de 2019. No site é enfatizada a força da comunidade, onde as redes sociais transformaram a relação marca e consumidor. Conforme descrito no site, em vez de procurar a marca, o consumidor procura referência em uma rede de outros consumidores. De acordo com o site, são incontáveis as comunidades com o poder de promover o sucesso ou o fracasso de um negócio e nestas comunidades habitam diferentes agentes consumidores, influenciadores e diversos outros contribuidores. Para a Stilingue isso é uma nova arena, uma nova dimensão na forma como se formam opiniões, como as pessoas compram produtos, buscam atendimento e *feedbacks*.

A Stilingue considera 3 competências para essa transformação: 1) Cuidado com a comunidade utilizando canais unificados, automatizações por IA e banco de influenciadores; 2) Inteligência sobre a voz das comunidades em dados internos e externos. Para aprofundar tendências, padrões, oportunidades e *insights* baseados em evidências; e 3) Agilidade na coleta de *feedbacks* em *real-time*, nos aprendizados em ciclos curtos e na reação imediata.

Segundo o site, a plataforma da Stilingue pode ser utilizada nos seguintes segmentos de negócios: Inteligência, Marketing, Atendimento ao Cliente, Agências, Vendas, *E-commerce*, Comunicação corporativa, TI, Relações Governamentais e *Compliance*.

Análise do site da Stilingue

De todas as empresas analisadas o site da Stilingue é o com menos informações disponibilizadas. Ao gerar uma nuvem de palavras do site, as que mais se destacam são: comunidades, inteligência, plataforma, marca, consumidor, artificial, clientes, consumidor como mostra a Figura 59.

Figura 59 - Nuvem de palavras do site da Stilingue

⁶¹ <https://stilingue.com.br/>



Fonte: Autora

Sobre a Privacidade dos Dados

A política de privacidade a Stilingue está disposta de forma bem discreta no final do site, uma vez que o *link* da mesma encontra-se dentro do Termo e Condições de Uso⁶². Analisando o Termo e Condições de Uso foi possível encontrar restrição para segmentação e uso de dados sensíveis, como dados de saúde (gravidez) condição financeira negativa, crenças ou outros dados que violem a Declaração Universal dos Direitos Humanos⁶³.

A Política de Privacidade⁶⁴ da Stilingue contempla os seguintes itens: 1) Recolhimento das informações; 2) Como as informações são usadas; 3) Armazenamento e Segurança; 4) O direito dos usuários de acesso às informações; 5) Compartilhamento e divulgação das informações; 6) *Web Beacons* e *Cookies*; 7) Registro de Acesso a Aplicações; 8) Spam e 9) Responsabilidade Limitada. Para esta pesquisa foram analisados os itens 1, 2, e 6.

Sobre o item 1, a Stilingue pode recolher informações pessoais dos usuários da plataforma, como o nome empresarial, nome próprio, endereço, *e-mail*, telefone, RG, CPF, CNPJ, informações de contato e outros dados que permitam a identificação. Além disso, a Stilingue poderá recolher informações demográficas como, idade; o conteúdo que o usuário utilizar; dados e informações sobre o *hardware* de acesso e *softwares* instalados no

⁶² <https://stilingue.com.br/wp-content/uploads/2018/10/Termos-Cond%C3%A7%C3%B5es-Stilingue-v3.0.pdf>

⁶³ <https://www.un.org/en/universal-declaration-human-rights/index.html>

⁶⁴ <https://lp.stilingue.com.br/wp-content/uploads/2018/10/Pol%C3%ADtica-de-Privacidade-Stilingue-v3.0.pdf>

dispositivo; endereço de Protocolo de Internet, tipo de navegador, informações inseridas em formulários, tempos de acesso e referências a endereços de sites.

Em relação ao item 2, a Stilingue faz uso dos dados inseridos na plataforma para administrar, gerir, melhorar e personalizar a plataforma, direcionar campanhas publicitárias, serviços e/ou produtos, responder perguntas do usuário, evitar atividades potencialmente proibidas ou ilegais, melhorar o conteúdo, assim como para entrar em contato com o usuário.

Quanto ao item 6, a Stilingue utiliza *cookies* e deixa claro que o bloqueio dos mesmos irá gerar algumas limitações na plataforma.

APÊNDICE E - Informações sobre as Soluções da MarketData

O site da MarketData foi acessado em 16 de maio de 2019. De acordo com o site⁶⁵, a Marketdata, é uma empresa de serviços de tecnologia com *know-how* nas seguintes soluções: *Database Management*, *Data-drive Creative Services*, CRM e *Analytics*.

As pessoas estão expostas a um mundo de mensagens e influências que determinam as suas decisões de compra. Elas seguem jornadas de compra distintas, com muitos pontos de contato que incluem a loja, a internet, o *e-commerce*, a TV, o celular e as redes sociais. Ao longo de suas interações com a marca, elas deixam sinais, o tempo todo. Estes sinais são traduzidos em dados (MARKETDATA, *online*).

Figura 60 - Informações sobre a MarketData

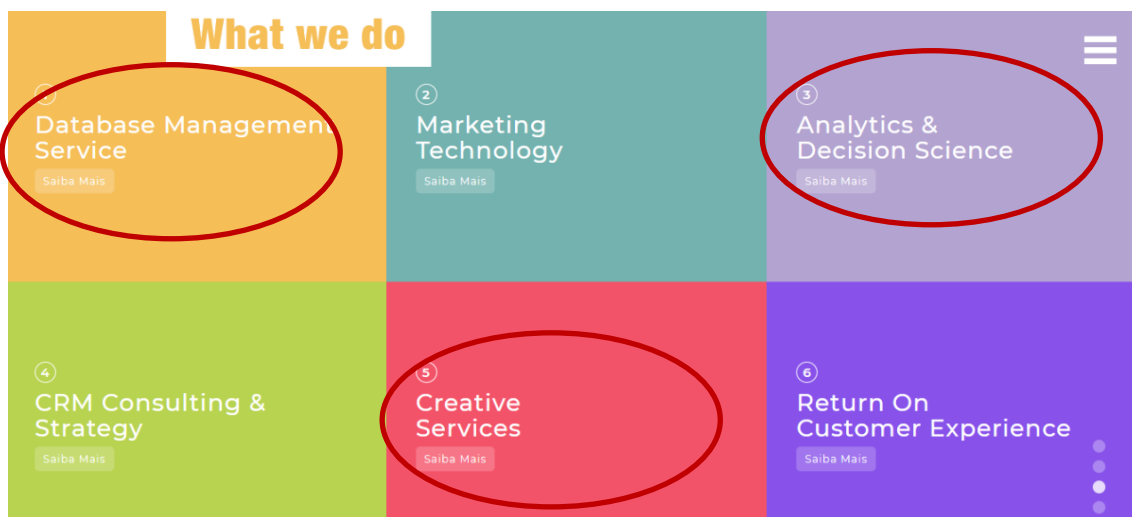


Fonte: Site da MarketData: <<http://www.marketdata.com.br/Marketdata>>

A Figura 61 mostra as soluções da MarketData. Para esta pesquisa foram analisadas as seguintes soluções: *Database Management*, *Analytics and Decision Science* e *Creative Services* que estão destacadas em vermelho. Segue abaixo detalhamento com base nas informações disponibilizadas no site da MarketData.

Figura 61 - Soluções da MarketData

⁶⁵ <http://www.marketdata.com.br/Marketdata>



Fonte - Site da MarketData: <<http://www.marketdata.com.br/Marketdata#sec-4>>

O *Database Management* (DBM) é um software de gerenciamento de Bancos de Dados que serve como interface entre o banco de dados e os usuários finais ou programas aplicativos. O DBM permite fazer construção de Banco de Dados de marketing; extração, tratamento e carga de dados (ETL); gestão do processo de *Database Management*; identificação de origens, auditoria de processos e qualidade; integração de dados transacionais e *web behavior* (DBM e *big data*).

O *Analytics and Decision Science*, ou estudos analíticos são utilizados para produzir *insights* a partir da base de clientes e permitir decisões mais precisas. Os modelos estatísticos utilizados indicam segmentos e os momentos exatos de ativação, relacionamento e retenção, permitindo prever resultados e otimizar investimentos de marketing. Este tipo de serviço inclui: *Business Intelligence*, portais e *dashboards*; Segmentação; Ferramentas de *Geomarketing*; *Next Best Offer* (NBO); *Search Engine Marketing* (SEM); Algoritmos de *Text Mining*; *Cluster Marketdata*® Enriquecimento com dados sócio-demográficos; Modelos estatísticos preditivos (*churn*, aquisição & reativação); *Lifetime Value* (LTV); *Media Investment Analytics*; e *Real-time predictions*.

O *Creative Services* permite combinar a capacidade de executar ideias de forma única com *insights* estratégicos derivados dos dados (*Data driven*) para criar, projetar e produzir campanhas criativas *one to one* integradas e personalizadas visando estimular os clientes a comprar de forma pessoal e em tempo real. Equilibrar o racional com o emocional é uma arte. Afinal, clientes não são todos iguais, estão sempre em movimento (MARKETDATA, *online*). Essa solução engloba: *User experience* e *user interface*; vídeos personalizados; criação de *mobile marketing* e apps; *design* e desenvolvimento de *websites*; *e-mail marketing*; campanhas *multi-channel*; conteúdo, ativação e monitoramento em redes sociais.

Análise do site da MarketData

O site da MarketData é organizado e possui uma descrição das soluções e serviços clara e de fácil navegação. Destaque para a Política de Privacidade e o Aviso de *Cookies* que estão realçados no site. Ao gerar uma nuvem de palavras, as que mais se destacam são: Marketing, Analytics, dados, informações, personalizadas, CRM, ROI, database, Digital, Automation como mostra a Figura 62.

Figura 62 - Nuvem de palavras do site da MarketData



Fonte: Autora

Sobre a Privacidade dos Dados

Ao acessar o site da MarketData há um aviso de que o mesmo utiliza *cookies* próprios e de terceiros que visam registrar o histórico dos usuários. Nesta página há o botão (Ver Política) que direciona o usuário para a Política de Privacidade e o botão (Concordo!) a fim de que o mesmo aceite o uso dos *cookies*. Caso o usuário não concorde, deve alterar as configurações para inibi-los.

A Política de Privacidade⁶⁶ tem por objetivo detalhar como os dados são coletados, tratados e processados. A última atualização desta política consta de novembro de 2018, e nela são tratados os seguintes itens: 1) Quais são as informações coletadas; 2) Como as informações são utilizadas; 3) Como eles se comunicam com os usuários; 4)

⁶⁶ <https://www.marketdata.com.br/contato/politica>

Compartilhamento de informações com terceiros; 5) Por quanto tempo as informações são mantidas; 6) Websites de Terceiros; 7) Como os dados são protegidos; 8) Dados de menores de idade; 9) Direitos dos usuários; 10) Como entrar em contato; e 11) Alterações a esta política de privacidade. Para esta pesquisa foram analisados os itens 1, 2 e 4.

Em relação ao item 1, quando alguém usa o site da MarketData e demais canais de comunicação mantidos por esta empresa são coletadas as seguintes informações: nome completo; sexo; endereço de *e-mail*; número de telefone (comercial, celular); *login* social (*linkedin*, *facebook* e outros); empresa; cargo; quantidade de funcionários (no caso de empresas); setor no qual a pessoa atua; assuntos profissionais de interesse; endereço comercial; data de nascimento; informações para recrutamento e seleção (o que pode incluir informações pessoais confidenciais). De acordo com o site:

A Marketdata pode receber, de suas afiliadas e parceiros, dados sobre organizações, indústrias, visitantes de sites, campanhas de marketing e outros assuntos relacionados aos nossos negócios e de nossos parceiros, afiliados ou outros, os quais usamos para tornar nossas próprias informações melhores ou mais úteis. Esses dados podem ser complementados com as informações coletadas diretamente com você (MARKETDATA, 2018, *online*).

Sobre o item 2, a MarketData pode usar as informações coletadas para: fornecer informações solicitadas pelo usuário, por exemplo, uma *newsletter*, boletins; fazer convites para eventos da MarketData; fornecer melhores maneiras de acessar informações do site e demais canais de comunicação mantidos pela MarketData; apresentar soluções e serviços da MarketData, afiliados e parceiros; realizar pesquisas profissionais dentro do contexto da atividade de marketing e tecnologia de dados; e processar e considerar candidatura a vaga de emprego (Recrutamento e Seleção).

No item 4 a MarketData descreve que eles podem enviar as informações pessoais para outras empresas do grupo WPP, afiliadas e parceiros, para ajudar a processar esses dados de acordo com os fins definidos nessa política, conforme mostra a Figura 63.

Figura 63 - Política de Privacidade: Compartilhamento de informações com terceiros

4. Compartilharemos suas informações com terceiros?

Podemos enviar suas informações pessoais para outras empresas do grupo WPP, afiliadas e parceiros, para nos ajudar a processar suas informações pessoais para os fins definidos nesta política. Mais detalhes sobre quais empresas fazem parte do grupo WPP podem ser encontrados [aqui](#).

Somos uma empresa que faz parte de um grupo global (WPP) e, portanto, eventual e estritamente para contatos profissionais, podemos transferir suas informações pessoais para países em todo o mundo, incluindo os EUA e outros países fora da Europa. Quando o país para o qual os seus dados serão transferidos, não apresentar uma legislação com nível adequado de proteção, estabeleceremos as devidas salvaguardas (usando cláusulas contratuais padrão) para garantir que suas informações sejam protegidas.

Podemos divulgar e compartilhar suas informações pessoais se nós ou qualquer uma das empresas do grupo WPP for objeto de uma venda/aquisição ou transação corporativa semelhante. Diante de tal situação, exigiremos de nossos parceiros que as informações pessoais sejam mantidas confidenciais.

Podemos divulgar informações pessoais a parceiros quando acreditarmos que somos obrigados por lei e para investigar, prevenir ou tomar medidas em relação a atividades suspeitas, ilegais ou de outra forma proibidas, incluindo, mas não se limitando à fraude.

Fonte - Site da MarketData: <<https://www.marketdata.com.br/contato/politica>>