

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE INFORMÁTICA  
CURSO DE CIÊNCIA DA COMPUTAÇÃO

RÉGIS EBELING

**Um Estudo do Impacto dos Debates  
Eleitorais da Eleição Presidencial de 2018  
na Opinião dos Usuários do Twitter**

Monografia apresentada como requisito parcial  
para a obtenção do grau de Bacharel em Ciência  
da Computação

Orientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Karin Becker

Porto Alegre  
2018

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Rui Vicente Oppermann

Vice-Reitora: Prof<sup>a</sup>. Jane Fraga Tutikian

Pró-Reitor de Graduação: Prof. Vladimir Pinheiro do Nascimento

Diretora do Instituto de Informática: Prof<sup>a</sup>. Carla Maria Dal Sasso Freitas

Coordenador do Curso de Ciência de Computação: Prof. Sérgio Luis Cechin

Bibliotecária-chefe do Instituto de Informática: Beatriz Regina Bastos Haro

## AGRADECIMENTOS

Agradeço primeiramente aos meus pais por todo o suporte que tive ao longo da vida e, mesmo que não seja mais necessária a ajuda, saber que eu tenha a quem recorrer em alguma hora de dificuldade.

Agradeço à minha namorada, Mirella Aguiar, por não desistir de mim mesmo em tempos mais sombrios. Finais de semestres e principalmente o final da graduação foram tempos de verdadeiros testes, principalmente por aturar minhas mudanças de humor e os diversos momentos de ausência. Talvez nem eu mesmo me aguentaria.

Agradeço aos meus amigos, em especial aos colegas da universidade por me ajudarem a desopilar o stress e as preocupações da vida acadêmica.

Agradeço à minha orientadora, professora Karin Becker, por ter guiado o fluxo de ideias, complexidade e prazos do presente trabalho com maestria. Ao iniciar o semestre já me passou de forma prática tudo o que eu deveria saber e como aconteciam geralmente as etapas para a realização de um Trabalho de Graduação de Curso. Sem sua ajuda, o caminho teria sido muito mais longo.

Agradeço aos professores Leandro Krug Wives e Viviane Moreira, integrantes da banca avaliadora, pelo tempo empregado em ler, sugerir e discutir alterações no presente trabalho.

Agradeço aos demais professores do Instituto de Informática da universidade, que ensinaram que a vida não é fácil e ao mesmo tempo foram solícitos em sanar dúvidas para que se chegasse ao objetivo.

Por último, agradeço à Universidade Federal do Rio Grande do Sul, que forneceu toda a estrutura necessária para a minha graduação no curso. Não precisei pagar deliberadamente pelo meu curso, pelos meus professores, pelas estruturas de prédios e salas, tive a oportunidade de fazer refeições por valores irrisórios nos Restaurantes Universitários dos campi, pude desfrutar férias nas instalações da universidade em Tramandaí, recebi auxílio para realização de bolsas, e inúmeras outras oportunidades que a UFRGS oferece aos seus alunos.

## RESUMO

A sociedade brasileira expressa cada vez mais suas opiniões em meios digitais, fomentando discussões nos mais diversos assuntos. O volume de dados gerados pelas redes sociais com este tipo de conteúdo, como o Twitter, apresenta inúmeras oportunidades de exploração, muitas delas ainda não aproveitadas. Neste trabalho busca-se identificar e mensurar o impacto causado por um debate eleitoral na intenção de votos para a presidência da República, a partir das opiniões sobre candidatos expressas na rede social Twitter. Baseado nos resultados obtidos, o custo e o tempo empregados em uma pesquisa eleitoral poderiam ser substituídos pela execução do experimento proposto no presente trabalho. O desenvolvimento do trabalho deu-se pela construção de um classificador para analisar as opiniões geradas via *tweets* para 3 candidatos à presidência monitorados ao final do 1º turno (Ciro Gomes, Fernando Haddad e Jair Bolsonaro), obtendo a porcentagem de aceitação de cada um para cada um dos debates. Foi mensurado para cada um dos debates o percentual de opinião positiva dos candidatos antes e após um debate, assim analisando a tendência de evolução da aceitação ou rejeição de cada um dos postulantes. Os classificadores dos candidatos monitorados obtiveram métricas de avaliação razoáveis na fase de treinamento e teste, porém os resultados alcançados pelo trabalho, ao serem comparados com as pesquisas de intenção de voto oficiais nos períodos dos debates, não apresentaram valores totalmente precisos. Uma análise adicional foi realizada neste trabalho para comparar as intenções de voto estimadas com o resultado final do 1º turno das eleições e desta vez, conseguiu retratar fielmente os percentuais de voto dos dois primeiros candidatos, Jair Bolsonaro e Fernando Haddad.

**Palavras-chave:** Twitter. Análise de Sentimentos. Eleições Presidenciais.

# **A Study of the Impact of Electoral Debates on the 2018 Presidential Election in Twitter Users' Opinion**

## **ABSTRACT**

Brazilian society has increasingly displayed its opinions in digital media, fomenting discussions on various subjects. The volume of data generated by social networks with this type of content, such as Twitter, which are potential research opportunities, are often not yet used. This paper aims to identify and measure the impact caused by an electoral debate in the intention of votes for the presidency of the Republic, based on the opinions on candidates expressed in the social network Twitter. Based on the results obtained, the cost and time spent in an electoral survey could be replaced by the execution of the experiment proposed in the present work. The development of the work was done by the construction of a classifier to analyze the opinions generated by tweets to 3 presidential candidates monitored at the end of the first round (Ciro Gomes, Fernando Haddad and Jair Bolsonaro), obtaining the percentage of acceptance of each one for each of the debates. The percentage of positive opinion of the candidates before and after a debate was measured for each of the debates, thus analyzing the tendency of the acceptance or rejection of each one of the candidates. The classifiers of the monitored candidates obtained reasonable evaluation metrics in the training and testing phase, but the results achieved by the work, when compared to official voting intentions surveys during the debates, didn't present totally accurate values. An additional analysis was carried out in this work to compare the estimated voting intentions with the final result of the first round of the elections and this time, it was able to faithfully portray the voting percentages of the first two candidates, Jair Bolsonaro and Fernando Haddad.

**Keywords:** Twitter. Sentiment Analysis. Presidential Elections.

## LISTA DE FIGURAS

Figura 2.1	Classificação pelo algoritmo de SVM .....	19
Figura 3.1	Processo de coleta e formação do <i>dataset</i> rotulado.....	28
Figura 3.2	Exemplo da coleta de um <i>tweet</i> de apoio ao candidato Ciro Gomes.....	31
Figura 3.3	Exemplo da coleta de um <i>tweet</i> de apoio ao candidato Jair Bolsonaro.....	31
Figura 3.4	Exemplo da coleta de um <i>tweet</i> de apoio ao candidato Luiz Inácio Lula da Silva.....	32
Figura 3.5	Processo de classificação de um conjunto de <i>tweets</i> .....	35
Figura 4.1	Candidatos monitorados ao longo dos debates.....	37
Figura 4.2	Evolução da opinião sobre Ciro Gomes ao longo dos 7 debates.....	38
Figura 4.3	Evolução da opinião sobre Fernando Haddad ao longo dos 4 debates.....	39
Figura 4.4	Evolução da opinião sobre Jair Bolsonaro ao longo dos 7 debates .....	40
Figura 4.5	Evolução da opinião sobre Luiz Inácio Lula da Silva ao longo dos 3 debates.....	41
Figura 4.6	Comparativo da Evolução da intenção de votos de Ciro Gomes.....	49
Figura 4.7	Comparativo da Evolução da intenção de votos de Fernando Haddad.....	49
Figura 4.8	Comparativo da Evolução da intenção de votos de Jair Bolsonaro .....	50
Figura 4.9	Comparativo da Evolução da intenção de votos de Luiz Inácio Lula da Silva .....	51
Figura 4.10	Comparativo entre Resultado do 1º Turno, Média dos índices estimados e Pesquisas Eleitorais.....	52

## LISTA DE TABELAS

Tabela 3.1	Período de coletas de <i>tweets</i> para cada debate .....	25
Tabela 3.2	Estrutura de um <i>tweet</i> obtida pelo data scraper .....	26
Tabela 3.3	Termos procurados na coleta de <i>tweets</i> para o conjunto alvo do experimento	27
Tabela 3.4	Número de <i>tweets</i> por candidato, em cada período de cada debate .....	27
Tabela 3.5	<i>Tweets</i> Positivos: dados da coleta.....	30
Tabela 3.6	<i>Tweets</i> Negativos: dados da coleta .....	31
Tabela 3.7	Resultados do classificador <i>Naive Bayes</i> .....	33
Tabela 3.8	Resultados do classificador SVM.....	34
Tabela 3.9	Percentuais de aprovação obtidos a partir da aplicação do modelo.....	36
Tabela 4.1	Percentuais de intenção de voto a partir dos resultados do trabalho .....	42
Tabela 4.2	Pesquisas de intenção de voto utilizadas como base para comparação .....	44

## LISTA DE ABREVIATURAS E SIGLAS

NLTK	<i>Natural Language Toolkit</i>
SVM	<i>Support Vector Machines</i>
API	<i>Application Programming Interface</i>
TSE	Tribunal Superior Eleitoral
PLN	Processamento de Linguagem Natural
TP	<i>True Positive</i>
TN	<i>True Negative</i>
FP	<i>False Positive</i>
FN	<i>False Negative</i>

## SUMÁRIO

<b>1 INTRODUÇÃO</b>	<b>11</b>
<b>2 FUNDAMENTAÇÃO TEÓRICA</b>	<b>14</b>
<b>2.1 Mineração de Textos</b>	<b>14</b>
<b>2.2 Análise de Sentimentos</b>	<b>14</b>
<b>2.3 Aprendizado de Máquinas Supervisionado</b>	<b>16</b>
2.3.1 Escolha das <i>Features</i>	17
2.3.2 Construção de Modelos de Classificação	17
2.3.2.1 <i>Naive Bayes</i>	18
2.3.2.2 <i>Support Vector Machine</i>	18
2.3.3 Métricas de Avaliação de Resultados	19
2.3.3.1 Acurácia	20
2.3.3.2 Precisão	20
2.3.3.3 <i>Recall</i>	20
2.3.3.4 <i>F-Measure</i>	21
<b>2.4 Processamento de Linguagem Natural e Ferramentas</b>	<b>21</b>
2.4.1 Tokenização	21
2.4.2 Remoção de <i>Stopwords</i>	22
2.4.3 Normalização	22
<b>2.5 Trabalhos Relacionados</b>	<b>23</b>
<b>3 PROCESSO DE COLETA E CLASSIFICAÇÃO DE DADOS SOBRE DEBATES ELEITORAIS</b>	<b>24</b>
<b>3.1 Candidatos e Períodos das Coletas</b>	<b>25</b>
<b>3.2 Coleta de Dados</b>	<b>26</b>
<b>3.3 Composição Automática de uma Base de Treino</b>	<b>27</b>
3.3.1 <i>Tweets</i> Positivos	28
3.3.2 <i>Tweets</i> Negativos	29
3.3.3 Exemplos Obtidos na Coleta para o <i>Dataset</i> Rotulado	29
<b>3.4 Pré-processamento e Definição de <i>Features</i></b>	<b>32</b>
<b>3.5 Treinamento dos Classificadores</b>	<b>33</b>
<b>3.6 Aplicação do Classificador aos Dados do Experimento</b>	<b>35</b>
<b>4 AVALIAÇÃO DOS RESULTADOS</b>	<b>37</b>
<b>4.1 Aceitação e Rejeição dos Candidatos em Relação a um Debate</b>	<b>37</b>
4.1.1 <i>Ciro Gomes</i>	38
4.1.2 <i>Fernando Haddad</i>	38
4.1.3 <i>Jair Bolsonaro</i>	39
4.1.4 <i>Luiz Inácio Lula da Silva</i>	40
<b>4.2 Comparações com as Pesquisas Eleitorais</b>	<b>41</b>
4.2.1 Comparações por Debate	43
4.2.1.1 Debate 1: Band	43
4.2.1.2 Debate 2: RedeTV!	45
4.2.1.3 Debate 4: TV Aparecida	45
4.2.1.4 Debate 5: SBT	46
4.2.1.5 Debate 6: Record	47
4.2.1.6 Debate 7: Globo	47
4.2.2 Comparativo das Evoluções de Intenção de Voto: Estimado x Pesquisa Oficial	48
4.2.2.1 <i>Ciro Gomes</i>	48
4.2.2.2 <i>Fernando Haddad</i>	49
4.2.2.3 <i>Jair Bolsonaro</i>	50

4.2.2.4 Luiz Inácio Lula da Silva.....	50
<b>4.3 Predição do Resultado do 1º Turno.....</b>	<b>51</b>
<b>5 CONCLUSÕES.....</b>	<b>53</b>
<b>REFERÊNCIAS.....</b>	<b>55</b>

## 1 INTRODUÇÃO

Observa-se nos dias atuais a presença da sociedade no ambiente digital, fornecendo voluntariamente informações pessoais, rotina, fotos e gostos. Como em um efeito dominó, as pessoas são influenciadas a ingressar em redes sociais por ver seu círculo de amizades presente nas mesmas, então buscando uma experiência semelhante. À medida que as redes sociais vêm ganhando importância e notoriedade, qualquer notícia ou fato do mundo real reverbera pela internet e ocasiona uma quantidade notável de opiniões de usuários, resultando em ricas fontes de dados.

De janeiro de 2017 a janeiro de 2018 ocorreu um aumento de 13% no número de usuários ativos nas redes sociais, alcançando 3.2 bilhões<sup>1</sup> e representando 42% da população total do globo. Ainda, 3 em cada 4 usuários de internet possuem contas em redes sociais, e quase a totalidade dos usuários ativos, 92%, o faz via dispositivos móveis. Esses números evidenciam o grande potencial da geração de conteúdo e opiniões que as redes sociais apresenta, permitindo inclusive a segmentação dos usuários em níveis e nichos de opiniões. O LinkedIn, por exemplo, segmenta opiniões e dados profissionais<sup>2</sup> e o Youtube apresenta opiniões sobre conteúdos multimídia. Já o Facebook e o Twitter resumizam opiniões dos usuários sobre diversos assuntos, sendo que o primeiro tende a apresentar textos maiores e melhor estruturados, enquanto o segundo apresenta uma opinião mais pontual devido à limitação de caracteres imposta pela plataforma.

O Twitter é uma rede social que inicialmente foi concebida para operar com mensagens de 140 caracteres, hoje estendido para 280, funcionando como um microblog dos usuários. Cada perfil pode seguir outras contas para receber suas mensagens, ser seguido e obviamente postar mensagens. Atualmente, o Twitter conta com 332.5 milhões de usuários ativos no mundo<sup>3</sup>. Ao final de 2016, o Brasil contava com cerca de 30 milhões de usuários<sup>4</sup> <sup>5</sup>, configurando umas das redes sociais mais populares do país. Sendo o Twitter uma das grandes plataformas de comunicação instantânea no Brasil e possibilitando a opinião direta dos usuários sobre uma miscelânea de assuntos, ela constitui uma fonte rica de dados sobre os quais análises interessantes podem ser feitas, incluindo a análise de opiniões.

---

<sup>1</sup><<https://wearesocial.com/blog/2018/01/global-digital-report-2018>> Acesso em Agosto de 2018

<sup>2</sup><<http://folha.com/no1781154>> Acesso em Novembro de 2018

<sup>3</sup><<https://folha.com/rwbh2ue8>> Acesso em Agosto de 2018

<sup>4</sup><<http://folha.com/no1861175>> Acesso em Setembro de 2018

<sup>5</sup><<https://www.statista.com/statistics/303931/twitter-users-latin-american-countries/>> Acesso em Setembro de 2018

A análise de sentimentos, ou mineração de opiniões (LIU, 2012) (TSYTSARAU; PALPANAS, 2012), é um campo de estudo que analisa opiniões, sentimentos e emoções em relação a entidades como eventos, produtos, organizações, indivíduos, entre outros, avaliando a polaridade desta relação. Sua aplicação em eventos como uma eleição traz o benefício de permitir a observação do desempenho de candidatos rapidamente, inclusive em tempo real, isto é, compreender e analisar os sentimentos do público-alvo com maior celeridade que o habitual, uma vez que não necessitaria uma equipe de pesquisa para realizar entrevistas pessoalmente com parcelas da população.

Em épocas de eleições, a realização de pesquisas de intenção de voto com a população para acompanhar a opinião popular sobre um candidato específico é uma das atividades chave para elaboração das campanhas. O valor dessa informação é particularmente crucial na reta final do período que antecede a votação. Quanto mais pessoas são entrevistadas, maior é o nível de confiança da pesquisa, mas maior também é o custo envolvido neste levantamento. Hoje as pesquisas tipicamente costumam estimar o tamanho de amostra que seja representativo em relação à população brasileira por meio de técnicas estatísticas<sup>6</sup>. Tendo em mente a importância das pesquisas eleitorais para definição das estratégias dos candidatos, o Twitter pode ser uma boa fonte de dados para se investir e contornar o viés do custo, através da mineração de opiniões.

As pesquisas eleitorais refletem a opinião popular acerca de diversos fatos e variáveis ocorridos a longo e principalmente a curto prazo sobre um candidato. Dentre esses eventos estão os debates televisionados entre os concorrentes. Em 2018, no período de realização deste trabalho, existiu uma atenção especial aos debates eleitorais devido ao menor tempo de campanha nas ruas e do horário político na grade televisiva em relação a anos anteriores, incentivando então que os candidatos expusessem propostas e planos de governo para ampla divulgação especialmente nesses eventos. Historicamente os debates sempre tiveram grande influência em boa parte dos eleitores brasileiros, seja por refletirem de forma destacada na mídia no dia seguinte e então atingindo mais pessoas, seja por dar insumos à formação de opinião diretamente pelo confronto de ideias entre os candidatos<sup>7</sup>.

Valendo-se do impacto que os debates televisionados proporcionam, um dos grandes trunfos dos candidatos é saber como mensurar a mudança de opiniões e a partir disto mudar os rumos de suas campanhas a partir destes marcos. Hoje, a constatação de uma mudança de opinião da população é obtida por meio de uma pesquisa eleitoral, que

---

<sup>6</sup><<https://www.bbc.com/portuguese/brasil-43845326>> Acesso em Agosto de 2018

<sup>7</sup><<https://www.bbc.com/portuguese/internacional-45216005>> Acesso em Agosto de 2018

demora 5 dias entre seu registro na Justiça Eleitoral e a divulgação dos seus resultados<sup>8</sup>. O tempo entre um evento específico, como um debate, e a observação da mudança de opinião ocasionada pelo mesmo, obtida pelos resultados da pesquisa eleitoral, pode apresentar um grande período de ação da campanha desperdiçado. Com a alteração do tempo da campanha eleitoral em 2018, o grande problema de uma equipe de um candidato passou a ser como obter uma percepção mais instantânea dos sentimentos expressos pela população sobre um dos concorrentes da eleição, acerca de fatos abordados em um debate, para então poder tratar a mudança do rumo da campanha visando sanar ou mesmo potencializar tal impacto.

O objetivo deste trabalho é realizar um estudo sobre o impacto dos debates eleitorais televisionados nas opiniões dos usuários do Twitter sobre os candidatos à presidência do Brasil de 2018. Esse impacto será mensurado comparando a polaridade das opiniões sobre os mesmos antes e após cada debate. Para cada debate pretende-se:

- construir e analisar comparativamente métricas que caracterizem a aceitação ou rejeição do candidato, tais como volume de mensagens positivas/negativas;
- analisar face às pesquisas eleitorais oficiais se este comportamento da rede social é também capturado pelas pesquisas eleitorais.

Este trabalho, através de um estudo de caso, investiga o impacto dos debates eleitorais televisionados nos sentimentos expressos por usuários na rede social Twitter. A contribuição mais significativa do presente trabalho é a aplicação da mineração de opiniões sobre eventos específicos, os debates ao longo das eleições brasileiras, gerando métricas comparativas temporais referentes à ocorrência do mesmo.

O restante do trabalho está organizado da seguinte forma. No capítulo 2 são apresentados os fundamentos teóricos relacionados à análise de sentimentos, ao processamento dos dados obtidos no Twitter e aos algoritmos de classificação de sentimentos. No capítulo 3 é apresentada a metodologia idealizada na concepção e aplicada no decorrer deste trabalho, além da construção e aplicação do classificador nos dados coletados. No capítulo 4 são discutidos os resultados provenientes da análise proposta. No capítulo 5 são apresentadas as conclusões sobre o estudo realizado neste trabalho.

---

<sup>8</sup><<http://www.tse.jus.br/legislacao/codigo-eleitoral/lei-das-eleicoes/lei-das-eleicoes-lei-nb0-9.504-de-30-de-setembro-de-1997#art33-35>> Acesso em Setembro de 2018

## **2 FUNDAMENTAÇÃO TEÓRICA**

Este capítulo tem propósito de apresentar conceitos teóricos sobre a análise de sentimentos e algoritmos para sua classificação, que foram utilizados no desenvolvimento do presente trabalho.

### **2.1 Mineração de Textos**

Segundo Witten and Frank (2005), a informação presente em um texto é clara, porém não é expressa de forma que se possa processá-la facilmente. A área de Mineração de Textos, que é aplicada na resolução desses problemas, pode ser definida como a extração de padrões, conceitos ou regras a de uma grande quantidade de dados não estruturados. As técnicas utilizadas para a extração do conhecimento em textos são as mesmas utilizadas na área de mineração de dados, adaptadas a dados textuais (GOMES, 2012). Além da mineração de dados, a mineração de texto envolve técnicas de linguística computacional e recuperação da informação, entre outras (HOTH0 et al., 2005).

De acordo com ARANHA (2007), o processo da mineração de textos pode ser dividido em 5 etapas distintas e sequenciais. A 1ª etapa, de Coleta, consiste na busca por documentos relevantes do universo dos quais deseja-se realizar a extração de dados. Na 2ª etapa é realizado o pré-processamento dos textos obtidos da etapa anterior, objetivando estruturar, eliminar dados irrelevantes e normalizar a informação morfológica de todos os documentos. A etapa 3, de indexação, busca a organização dos documentos normalizados na etapa 2, visando facilitar seu posterior acesso. Na etapa 4 realiza-se a atividade de mineração de texto propriamente dita, usando técnicas de classificação, agrupamento ou associação, por exemplo. Na 5ª e última etapa, análise, ocorre a interpretação dos padrões encontrados e seu significado no contexto analisado.

### **2.2 Análise de Sentimentos**

Dentro da área de Mineração de Textos encontra-se a análise de sentimentos, ou mineração de opiniões. O termo diz respeito a análises de opiniões, emoções, sentimentos, e outros fatores comportamentais de pessoas sobre entidades como

produtos, serviços e demais eventos, desde que representados de forma textual (LIU, 2012). Ainda, de forma mais geral, o termo pode ser empregado para definir o tratamento computacional de opinião, subjetividade e sentimento em textos (PANG; LEE, 2008).

Dentre os conceitos e termos da área de mineração de opiniões que vêm sendo descritos para tarefas associadas a detecção de sentimentos, encontram-se:

- Polaridade do sentimento: apresenta o sentimento expresso em um documento ou parte dele como resultado binário (positivo ou negativo) ou ternário (positivo, negativo ou neutro).
- Força representada pelo sentimento em uma escala (REIS et al., 2014): apresenta a intensidade da polaridade de um sentimento presente em um texto, quer de forma contínua (e.g. um intervalo de valores negativos e positivos) ou discreta (e.g. muito positivo, positivo).
- Emoção (MOHAMMAD; TURNEY, 2013): apresenta um sentimento específico como alegria, tristeza, surpresa, etc.
- Estudo da subjetividade e objetividade (FELDMAN, 2013): busca-se uma estimativa da compensação de se realizar uma análise de sentimentos em um conjunto de dados, inferindo isto a partir de suas sentenças objetivas (normalmente possuindo fatos ou informações) e sentenças subjetivas (sentimentos pessoais e opiniões).

A análise de opiniões pode ocorrer em diferentes níveis de granularidade (LIU, 2012). No nível de documento assume-se que cada texto representa a opinião geral sobre uma mesma entidade, sendo a forma de classificação de opinião mais genérica. No nível de sentença, cada frase ou oração presente no documento pode representar a opinião sobre um aspecto diferente da mesma entidade, ou mesmo sobre uma entidade diferente. Por último, o nível de entidade e aspecto busca a classificação de características específicas de uma entidade geral (e.g. preço ou qualidade de um computador), podendo estar distribuída em diversas sentenças esparsas do documento.

Existem diversas formas de realizar a análise de sentimentos, com destaque para a abordagem de aprendizado de máquina supervisionado e o método léxico (KOLCHYNA et al., 2015). O método de aprendizado de máquina supervisionado realiza um treinamento utilizando um corpus de documentos rotulados, buscando por padrões presentes neste conjunto de dados que os relacione com o rótulo marcado. O

modelo de classificação resultante permite prever o rótulo de novas instâncias. O método de aprendizado de máquina é bastante dependente dos dados usados para treinamento, sendo que os melhores resultados são observados quando o corpus é de bom tamanho, e especificamente anotado para o domínio do problema (LIU, 2012). Já o método léxico é baseado em um dicionário anotado de palavras com suas respectivas polaridades, e em uma função de agregação. O método consiste em analisar os termos individuais do documento buscando no dicionário e agregar a importância de suas polaridades individuais no contexto geral do texto. O método léxico não necessita de um treinamento prévio, mas é dependente de um léxico adequado.

Neste trabalho, escolheu-se estudar o sentimento na forma de polaridade (positiva e negativa), em nível de documento, devido ao tamanho reduzido dos *tweets*. Considerando a informalidade dos textos gerados por redes sociais, os dicionários convencionais não costumam apresentar bons resultados, optou-se então por classificar os sentimentos via aprendizado de máquina.

### **2.3 Aprendizado de Máquinas Supervisionado**

O procedimento para realizar o processo de aprendizado de máquina supervisionado, segundo BENEVENUTO et al. (2015), é compreendido em 4 etapas. A 1ª etapa consiste na obtenção de um conjunto de dados rotulados que será utilizado para treinamento e teste do modelo. A 2ª etapa é definida pela escolha das *features* utilizadas para aprender padrões a partir conjunto de dados. Na etapa 3 utiliza-se um algoritmo de aprendizagem para treinar um modelo computacional, cujo desempenho pode ser verificado utilizando o conjunto de teste. Na última etapa o modelo obtido anteriormente pode ser aplicado para prever o rótulo de novas instâncias.

O aprendizado de máquinas supervisionado necessita um conjunto previamente rotulado de exemplos para aplicar o reconhecimento de padrões (NARR et al., 2012), porém tal recurso nem sempre existe. Uma alternativa à anotação de um conjunto de dados manualmente é a aplicação de heurísticas para rotulação automática. Por exemplo, para obter dados automaticamente rotulados para utilizar no treinamento do modelo classificador para polaridade, pode-se utilizar *emojicons* (READ, 2005).

### 2.3.1 Escolha das *Features*

Para o processo de aprendizado de um algoritmo supervisionado ocorrer, deve-se fornecer ao mesmo uma ou mais *features* para cada amostra do conjunto rotulado. Ao receber a amostra com sua classe correta e a sinalização das características para analisar, o algoritmo classificador busca por padrões nessas *features* que as relacionem ao rótulo.

Para construir um classificador de sentimentos em *tweets*, Mohammad et al. (2013), que desenvolveram um classificador considerado estado da arte, investigam a contribuição de diferentes tipos de *features*, entre elas

- N gramas: presença ou ausência de *n* gramas como unigramas ou bigramas;
- Palavras alongadas: presença ou ausência de palavras alongadas, como “muito00”;
- Sufixos e Prefixos: presença ou ausência de sufixos e prefixos em palavras;
- *Emoticons*: presença ou ausência de *emoticons*, que geram por si só um forte indicativo de sentimento;
- Termos de negação: se algum termo de negação estiver presente antes ou próximo do alvo, as polaridades dos *tokens* após o termo são invertidas;
- Pontuação: a presença ou ausência de sequências de pontuações;
- *Mentions*: presença de menções a outros perfis de usuários no Twitter;
- Posição: posição de um termo em um *tweet*.

Neste trabalho utilizou-se unigramas como única *feature* para o classificador.

### 2.3.2 Construção de Modelos de Classificação

Para realizar o processo de aprendizado, um algoritmo de classificação precisa de um conjunto de dados rotulados para reconhecer os padrões característicos das classes dadas. A partir deste processo, o modelo construído torna-se apto a aplicar os mesmos padrões encontrados nos dados a classificar futuramente.

Dentre os algoritmos de aprendizagem supervisionados, os algoritmos classificadores *Naive Bayes* e *Support Vector Machine* (SVM) são muito populares para classificação de documentos.

### 2.3.2.1 Naive Bayes

Naive Bayes é um algoritmo classificador probabilístico baseado no Teorema de Bayes, conforme mostrado na Equação 2.1. O nome do algoritmo, "*naive*" (ingênuo), é baseado na característica de que o mesmo desconsidera a correlação das variáveis utilizadas como critério para classificar os textos.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (2.1)$$

O algoritmo de Naive Bayes calcula para cada classe possível a probabilidade de um documento pertencer a ela, analisando individualmente a probabilidade de cada atributo do documento pertencer à mesma. Para cada documento  $d$ , tem-se que  $X$  é seu conjunto de atributos, então a probabilidade de pertencer a uma dada classe  $C1$  é expressa como:  $P(d|C1) = P(x1|C1)*P(x2|C1)*...*P(xn|C1)*P(C1)$ , onde  $x1, \dots, xn$  são atributos pertencentes a  $X$  e  $P(C1)$  é a probabilidade à priori de um documento qualquer pertencer à classe  $C1$ . O algoritmo classificador rotula o documento  $d$  com a classe  $C$  do conjunto de classes que resultar na maior probabilidade a partir da equação.

### 2.3.2.2 Support Vector Machine

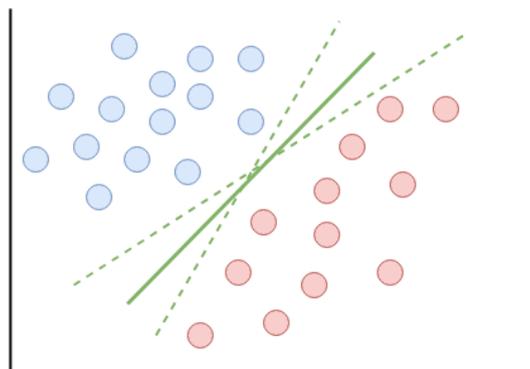
*Support Vector Machine* é um algoritmo de classificação que aprende o padrão de um conjunto previamente rotulado construindo hiperplanos com a maior distância possível para separação das classes de elementos. A representação de elementos a serem processados por SVM é vetorial, sendo que na mineração de textos utiliza-se *bag of words*.

Conforme mostrado na Figura 2.1, em um exemplo com 2 possíveis classes para rotulação, define-se uma reta que melhor separe o espaço vetorial em duas partes, repartindo os pontos vermelhos de um lado e os pontos azuis do outro. Para problemas com diversas classes possíveis para rotulamento, a relação vetorial é mantida, porém utiliza-se uma função de mapeamento para representar os vetores em uma dimensão de maior ordem.

Neste trabalho optou-se pela comparação dos algoritmos de *Naive Bayes* e SVM na etapa de treinamento e teste do modelo. O classificador que obter melhores métricas na avaliação será utilizado para analisar os *tweets* coletados dos candidatos. Para a construção do conjunto de dados rotulados que será utilizado para o treinamento dos modelos, buscou-se *tweets* com *hashtags* de apoio e rejeição, de maneira similar ao

método de busca por *emojicons* utilizado em (READ, 2005).

Figura 2.1: Classificação pelo algoritmo de SVM



Fonte: O Autor

### 2.3.3 Métricas de Avaliação de Resultados

Para obter um parâmetro da qualidade do classificador construído, deve-se realizar testes a partir de amostras do conjunto rotulado. Para realizar a avaliação do modelo existem diferentes técnicas, como *hold out*, *cross validation* e *bootstrap*.

No método de *hold out*, separa-se o conjunto de dados rotulados em subconjuntos de treinamento e teste, treinando o modelo com o subconjunto de treinamento e realizando a avaliação com o subconjunto de testes. O método de *cross validation* consiste na divisão do conjunto rotulado em  $k$  subconjuntos, denominados *folds*, então em  $k$  etapas,  $k-1$  *folds* são utilizados no treinamento e o subconjunto restante é utilizado para o teste. A capacidade de previsão a partir da técnica de *cross validation* é obtida a partir da média das avaliações das  $k$  etapas. Na técnica de *bootstrap*, para o conjunto de dados rotulados com  $n$  amostras, constrói-se um conjunto de treinamento com  $n$  amostras podendo ou não ocorrer amostras repetidas. Um subconjunto de testes obtido por *bootstrap* é formado pelas amostras que não estiverem presentes no subconjunto de treinamento, e, a avaliação calculada pelo método, é a média das avaliações de cada uma das repetições do processo (variando as amostras dos subconjuntos).

A tarefa de avaliar o desempenho de um modelo classificador está relacionada com a observância de alguns valores gerados, demonstrados a seguir.

- TP: indica que o dado foi classificado corretamente de acordo com o rótulo pré

definido;

- FP: o dado foi classificado como pertencente à classe em análise, porém é incorreto;
- TN: caso onde o dado é rotulado corretamente com uma classe diferente da classe analisada;
- FN: a classe de um dado é predita como diferente da classe analisada, porém deveria ser igual.

### 2.3.3.1 Acurácia

A acurácia é uma métrica avaliativa simples que aponta a porcentagem de dados que foi classificada corretamente dentre todo o conjunto de dados. A acurácia, por ser simples, pode ser imprecisa em seu resultado, podendo rotular todos os dados com a mesma classe e então seria avaliada de acordo com o tamanho da classe perante ao conjunto completo. Esta métrica é descrita pela Equação 2.2.

$$A = \frac{TP+TN}{TP+FP+TN+FN} \quad (2.2)$$

### 2.3.3.2 Precisão

O cálculo da precisão de uma classe se dá pela razão do número de casos classificados corretamente como ela pela soma de todos os casos classificados com esta mesma classe. Esta métrica calcula a probabilidade da predição de uma dada classe estar correta em relação a todos os casos. A precisão de uma classe é definida pela Equação 2.3.

$$P = \frac{TP}{TP+FP} \quad (2.3)$$

### 2.3.3.3 Recall

O *recall* de uma classe consiste no número de vezes que o classificador rotulou corretamente uma amostra com a classe em análise, dividido pelo número total de casos pertencentes à classe no conjunto. Tem-se então uma medida do número de documentos que deveriam ter sido classificados como a classe analisada e realmente foram. O *recall* de uma classe é definido pela Equação 2.4.

$$R = \frac{TP}{TP+FN} \quad (2.4)$$

### 2.3.3.4 *F-Measure*

A medida *F-Measure*, também referida por *F1-score* e *F-score*, é uma métrica mais abrangente que calcula a média ponderada dos valores de *recall* e precisão de uma classe. O cálculo é dado conforme a Equação 2.5.

$$F1 = \frac{2 * P * R}{P + R} \quad (2.5)$$

## 2.4 Processamento de Linguagem Natural e Ferramentas

Segundo Correa (2013), o Processamento de Linguagem Natural (PLN) consiste na aplicação de métodos e técnicas que possibilitam a extração de semântica da linguagem humana expressa em textos e voz. Já para Chowdhury (2005), PLN é uma área de pesquisa e aplicação que explora como computadores podem ser usados para entender e manipular texto ou discurso para fazer coisas úteis.

As principais tarefas de PLN são relacionadas com a etapa de pré-processamento dos documentos, em um contexto de mineração de texto, visando uma melhor identificação e preparação das *features* a serem utilizadas no modelo de classificação.

No presente trabalho, serão utilizados alguns recursos ligados ao PLN para a preparação dos *tweets*, os quais serão discutidos a seguir.

### 2.4.1 Tokenização

O processo de tokenização realizado na etapa de Pré-processamento busca dividir o corpus em unidades mínimas denominadas *tokens*. Um *token* representa geralmente uma palavra, podendo ser também uma pontuação ou símbolo. É obtido quebrando um texto quando encontra um caractere com função de delimitador, tal como um espaço em branco ou um símbolo de pontuação. Há diferentes estratégias de separação. Um exemplo do processo de tokenização pode ser visto a seguir:

"O primeiro turno das eleições ocorre em 8 de outubro de 2018!"

[O] [primeiro] [turno] [das] [eleições] [ocorre] [em] [8] [de] [outubro] [de] [2018]

[!]

### 2.4.2 Remoção de *Stopwords*

Os dados em formatos textuais são não-estruturados, então o pré-processamento em tarefas de Mineração de Textos deve padronizar as informações para a etapa seguinte, diminuindo o conteúdo léxico irrelevante do documento. Uma das formas de reduzir o léxico de um texto é a remoção de *stopwords*. *Stopwords* são palavras em um texto que possuem valor irrelevante para a análise, pois são utilizadas em todo o tipo de texto. Exemplos de *stopwords* são conjunções, pronomes, preposições e artigos (INDURKHYA; DAMERAU, 2010). Existem, ainda, listas prontas que organizam *stopwords* em diferentes idiomas.

### 2.4.3 Normalização

O processo de normalização de um texto consiste em reduzir as palavras em uma representação única, podendo agrupar várias de suas derivações em um único termo que representa o valor estrutural único do mesmo para a análise. Em geral, a aplicação de técnicas de normalização introduz uma melhora significativa nos sistemas de Mineração de Texto. Segundo CARRILHO JUNIOR (2007) essa melhora varia de acordo com o escopo, o tamanho da massa textual e o que se pretende obter como saída do sistema. Exemplos de normalização podem ser a lematização, *stemming*, e agrupamento de sinônimos. A seguir uma breve descrição dos 3 métodos.

O processo de *stemming* converte as palavras do texto em um radical, removendo seus afixos conforme o critério do algoritmo que realize a tarefa. Um algoritmo de *stemming* processa cada *token* do texto para reduzi-lo a uma forma comum de representação a todas variantes da palavra, e assim, todas as palavras de mesmo radical são representadas por um único termo.

Na lematização as palavras são transformadas na sua forma canônica, isto é, forma de dicionário. Por exemplo, as palavras "música", "músicas", "musical" e "musicalidade" possuem a mesma palavra originária: "música".

O agrupamento de sinônimos faz uso de um dicionário de sinônimos e faz a busca desses termos no texto para substituir todas as palavras com significado semelhante por um único termo representativo.

## 2.5 Trabalhos Relacionados

O alvo principal do presente trabalho é a análise de sentimentos em uma fonte de opinião fracamente estruturada, o Twitter, coletando dados relacionados aos candidatos da eleição presidencial brasileira de 2018. Neste capítulo são descritos alguns trabalhos onde foram utilizadas técnicas de análise de sentimentos sobre textos gerados por usuários (*tweets*, comentários de jornais) para realizar previsões sobre eleições.

No contexto da eleição federal alemã de 2009, um estudo (TUMASJAN et al., 2010) concluiu que o Twitter é usado extensivamente para deliberação política. As conclusões vão além: os conteúdos dos *tweets* são indicadores válidos para mensurar o sentimento político, e há uma forte correlação entre o resultado das eleições com o volume de *tweets* sobre um partido. O estudo foi conduzido analisando o texto de mais de 100 mil *tweets* com menções a candidatos ou partidos. A análise textual efetuada para extrair os sentimentos dos *tweets* foi realizada utilizando LIWC<sup>1</sup>, um *software* desenvolvido justamente para propósitos como este.

Um modelo classificador de análise em tempo real para os sentimentos sobre os candidatos à eleição presidencial norte americana de 2012, foi descrito em (WANG et al., 2012). Foi utilizado o algoritmo *Naive Bayes*, utilizando unigrama como *feature*, e foram consideradas 4 polaridades do sentimento: positiva, negativa, neutra e incerta. A acurácia do classificador construído foi de 59%.

O emprego de técnicas de análise de sentimentos sobre comentários das notícias sobre os cenários eleitorais em um site de determinado jornal brasileiro, abrangendo três eleições, foi relatado em (TUMITAN; BECKER, 2014). A proposta do trabalho foi investigar variações nas intenções de voto a partir dos dados coletados. Foram utilizadas duas abordagens para desenvolver o classificador: baseada em dicionários e baseada em aprendizado de máquina. Para a abordagem de aprendizado de máquina, foi utilizado o algoritmo de SVM com unigramas como *feature*. Uma das contribuições do trabalho, ainda, foi propor diferentes métricas para definir o sentimento captado em relação a um candidato.

---

<sup>1</sup><<http://www.liwc.net/>> Acessado em Outubro de 2018

### 3 PROCESSO DE COLETA E CLASSIFICAÇÃO DE DADOS SOBRE DEBATES ELEITORAIS

O cenário de uma eleição pode ser posto em números com ajuda de redes sociais, em especial o Twitter, através de processos de análise de sentimento. No contexto da eleição de 2018, o experimento retratado por este trabalho objetiva representar as opiniões populares sobre os candidatos nos debates eleitorais através de métricas de sentimento extraídas das redes sociais, mais especificamente, o Twitter. Espera-se que este sentimento possa ser utilizado para predições com um certo grau de fidelidade do sentimento geral da população.

As análises propostas no presente trabalho possuem dois objetivos:

- mensurar se um debate altera a aceitação e rejeição de um candidato;
- comparar se as opiniões sobre um candidato na rede social Twitter correspondem aos resultados das pesquisas eleitorais do período do estudo.

A primeira análise proposta é motivada pela hipótese de que o comportamento em uma rede social reflete o do mundo real, especificamente no escopo de uma eleição. Para analisar os sentimentos presentes nos *tweets* de cada candidato antes e após cada um dos debates, desenvolveu-se um processo de coleta de *tweets*, pré-processamento e construção de um classificador de polaridade.

O segundo objetivo tem um propósito comparativo motivado pela hipótese de que o comportamento de uma rede social pode ser representativo o bastante para ser um bom substituto ao corrente método de mensuração de opinião pública, as pesquisas eleitorais.

Ambos objetivos são baseados na classificação automática dos *tweets* quanto à polaridade. Para este fim, desenvolvemos como parte deste trabalho um classificador de sentimentos de acordo com a abordagem de aprendizado supervisionada. Para evitar a anotação manual da polaridade dos *tweets*, que além de ser trabalhosa requer mais de uma pessoa para ser validada, utilizamos um método de composição automática de um conjunto de treino/teste tomando como ponto de partida os perfis oficiais dos candidatos.

No restante deste capítulo são apresentadas as etapas do processo de classificação de sentimentos nos *tweets*, incluindo os candidatos monitorados, coleta de dados, pré-processamento e desenvolvimento de um modelo de classificação, junto com as decisões de projeto e suas motivações.

### 3.1 Candidatos e Períodos das Coletas

A análise proposta no trabalho visa monitorar os candidatos das três chapas que lideravam as intenções de voto na pesquisa divulgada pela BTG Pactual no dia 02 de setembro de 2018. A chapa PT / PCdoB liderava com o candidato Luiz Inácio Lula da Silva, seguida pela chapa PSL / PRTB do candidato Jair Bolsonaro e do PDT com Ciro Gomes.

Contudo, em um ano eleitoral atípico, não houve debates onde os candidatos das três chapas eleitorais analisadas participaram simultaneamente, devido ao indeferimento da candidatura de Lula, e à agressão ao candidato Jair Bolsonaro. Jair Bolsonaro participou dos dois primeiros debates (Band e RedeTV), Fernando Haddad dos quatro últimos (TV Aparecida, SBT, Record e Globo), Ciro Gomes compareceu em todos e Luiz Inácio Lula da Silva em nenhum dos eventos. Mesmo nunca acontecendo o embate direto, decidiu-se manter a análise de todos os candidatos para cada um dos debates por entendermos que mesmo ausentes, aprovações e reprovações continuariam sendo observadas ao comparar com os candidatos presentes e assim, a intenção de votos de qualquer um poderia ser afetada.

Decidiu-se que a coleta total dos *tweets* englobaria os 7 debates oficiais transmitidos via televisão ou internet, ocorridos no período de 07/08 a 06/10, conforme a Tabela 3.1.

Tabela 3.1: Período de coletas de *tweets* para cada debate

	Debate Band	Debate RedeTV	Debate TV Gazeta	Debate TV Aparecida	Debate SBT	Debate Record	Debate Globo
<b>Data do Debate</b>	09/08	17/08	09/09	20/09	26/09	30/09	04/10
<b>Início da Coleta</b>	07/08	15/08	07/09	18/09	24/09	28/09	02/10
<b>Final da Coleta</b>	11/08	19/08	11/09	22/09	28/09	02/10	06/10

### 3.2 Coleta de Dados

Devido às mudanças no cenário político nacional que afetaram os candidatos monitorados neste trabalho, ocorreu a necessidade de coletar *tweets* mais antigos referentes aos eventos. Com a versão gratuita da API oficial do Twitter é possível coletar somente mensagens dos últimos 7 dias, além de ela não mostrar todos os *tweets* da busca seguindo um critério de relevância.

Devido às limitações encontradas na API oficial, foi utilizado um *web scraper* de código aberto desenvolvido em Python, o GetOldTweets<sup>1</sup>. Com esta ferramenta foi possível buscar *tweets* do passado sem restrições, podendo então realizar a etapa de coleta dos dados de qualquer período da campanha eleitoral em qualquer momento da pesquisa.

A estrutura dos *tweets* coletados pelo *web scraper*, mostrada na Tabela 3.2, possui informações básicas que são mostradas para cada *tweet* na rede social. Dos atributos descritos, os únicos utilizados ao decorrer do desenvolvimento do classificador foram os campos “*text*” e “*id*”, além do campo “*date*” que foi um dos critérios da pesquisa no *web scraper*. Enquanto que o campo “*text*” foi o grande protagonista da estrutura de um *tweet* coletado, o campo “*id*” foi utilizado para eliminar mensagens que estivessem duplicadas tanto no conjunto rotulado para treinamento e teste, quanto no conjunto que seria classificado.

Tabela 3.2: Estrutura de um *tweet* obtida pelo data scraper

Informação	Tipo de Dado	Descrição
username	string	Usuário que escreveu o <i>tweet</i>
date	date	Data da publicação da mensagem
retweets	int	Número de <i>retweets</i>
favorites	int	Número de favoritações da mensagem
text	string	O <i>tweet</i> propriamente dito
geo	string	Localização geográfica na emissão do <i>tweet</i>
mentions	string	Usuários marcados na mensagem
hashtags	string	<i>Hashtags</i> utilizadas na mensagem
id	string	Identificador único da mensagem
permalink	string	URL do <i>tweet</i>

<sup>1</sup><<https://github.com/Jefferson-Henrique/GetOldTweets-python>> Acesso em Agosto de 2018

A partir da definição dos candidatos a serem monitorados ao longo do período eleitoral, foi necessário estabelecer o processo de coleta dos dados a serem avaliados para o experimento. Para formar os conjuntos de dados que seriam alvos do presente trabalho, optou-se por coletar *tweets* contendo o nome do candidato, sua *hashtag* ou a menção ao seu perfil na rede social, conforme explicitado na Tabela 3.3. Para cada um dos debates foram coletados *tweets* dois dias antes da hora de início do evento do debate, e dois dias após ele, resultando nos volumes informados na Tabela 3.4.

Tabela 3.3: Termos procurados na coleta de *tweets* para o conjunto alvo do experimento

Candidato	Termos de busca
<b>Ciro Gomes</b>	“ciro”, “#ciro”, “@cirogomes”
<b>Fernando Haddad</b>	“haddad”, “#haddad”, “@haddad_fernando”
<b>Jair Bolsonaro</b>	“bolsonaro”, “#bolsonaro”, “@jairbolsonaro”
<b>Luiz Inácio Lula da Silva</b>	“lula”, “#lula”, “@lulaoficial”

Tabela 3.4: Número de *tweets* por candidato, em cada período de cada debate

	<b>Gomes</b>		<b>Haddad</b>		<b>Bolsonaro</b>		<b>Lula da Silva</b>	
	Antes	Depois	Antes	Depois	Antes	Depois	Antes	Depois
<b>Debate 1</b>	30258	128803	X	X	349451	222597	95944	83005
<b>Debate 2</b>	25790	46944	X	X	116612	168399	200943	79077
<b>Debate 3</b>	91466	126933	X	X	370966	192882	60551	83758
<b>Debate 4</b>	70577	82872	112420	134955	220373	189402	X	X
<b>Debate 5</b>	89854	112041	145321	101911	255761	273112	X	X
<b>Debate 6</b>	83762	97017	235658	198776	230458	211089	X	X
<b>Debate 7</b>	112672	152075	178433	165285	319877	292322	X	X

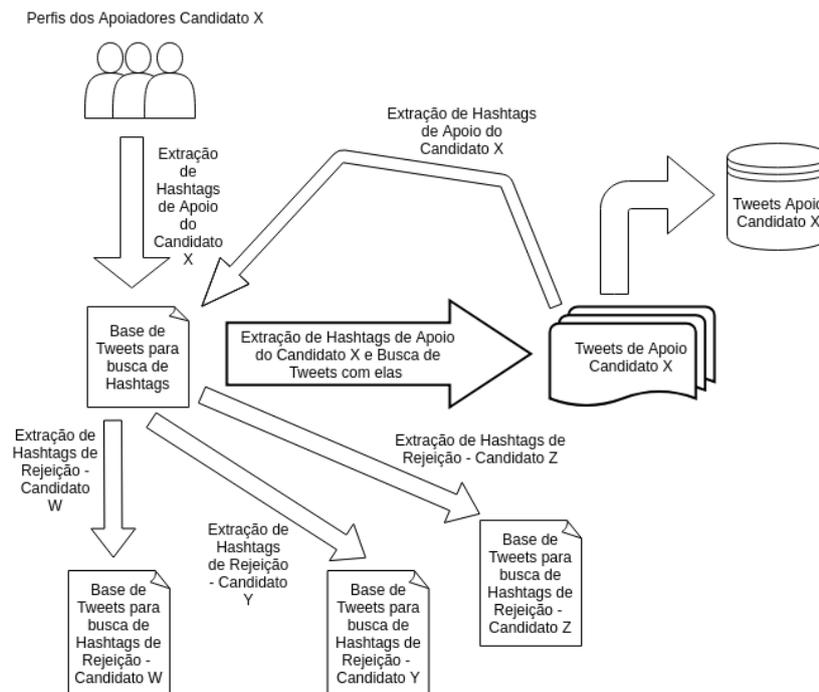
### 3.3 Composição Automática de uma Base de Treino

Para a etapa de aprendizado de um algoritmo classificador supervisionado faz-se necessária a presença de um *dataset* rotulado. Uma das questões a ser pontuada sobre um algoritmo de aprendizado supervisionado é a forma como este *dataset* rotulado é

obtido, sendo geralmente anotado manualmente por mais de uma pessoa. A partir de múltiplas anotações, extrai-se um consenso sobre a polaridade do sentimento expresso na mensagem. Este *dataset* deve ser também de um tamanho considerável para que seja eficaz, e ligado ao domínio. Como o presente trabalho não pôde contar com um segundo anotador para o trabalho manual, e considerando o volume de dados rotulados necessários, decidiu-se utilizar uma estratégia automática a partir de *hashtags* de apoio e rejeição para cada candidato, resumida na Figura 3.1.

Pela dificuldade em aferir o sentimento de neutralidade a uma *hashtag* de forma automática, optou-se por não manipular esta categoria, somente dividindo as mensagens entre positivas e negativas.

Figura 3.1: Processo de coleta e formação do *dataset* rotulado



Fonte: O Autor

### 3.3.1 Tweets Positivos

O método da busca, ilustrado pela Figura 3.1, consistiu em coletar todos os *tweets* das contas oficiais dos candidatos e apoiadores diretos de cada um deles, em uma janela de tempo de dois dias antes e dois após cada debate, obtendo uma base textual inicial para procura de *tweets* positivos de cada candidato. Nesta base procuraram-se *hashtags*

semente de apoio, e estas *hashtags* semente foram utilizadas para buscar mais *tweets* com o *web scraper*, armazenando-os no *dataset* rotulado de apoio, ao mesmo tempo que ia obtendo-se bases textuais adicionais para outras buscas. O processo resulta iterativamente na composição do conjunto de dados rotulados de cada candidato adicionando as novas bases textuais, além da descoberta de novas *hashtags* de apoio. O número de iteração do processo dependeu do tamanho desejado do conjunto rotulado. A Tabela 3.5 descreve os perfis utilizados na busca inicial de *hashtags*, *hashtags* mais frequentemente utilizadas e o tamanho do conjunto de *tweets* de apoio, por candidato, ao longo do período de monitoramento.

### 3.3.2 *Tweets* Negativos

A busca pelas mensagens negativas teve um comportamento levemente diferente. A coleta de *hashtags* de rejeição foi realizada paralelamente à busca de mensagens positivas dos candidatos concorrentes, onde foi observado que a defesa de um muitas vezes estava relacionada ao ataque a outro. A Figura 3.1 mostra o processo de formação da base de *tweets* para busca de *hashtags* de rejeição de candidatos a partir da busca de *hashtags* de apoio dos outros. Assim como na busca aos *tweets* com *hashtags* de apoio, ocorre um processo iterativo, onde a cada pesquisa de *hashtags* forma-se uma nova base de *tweets* para possibilitar novas buscas. A Tabela 3.6 descreve os dados obtidos na coleta de *tweets* negativos, informando exemplos de *hashtags* de rejeição e o tamanho total do conjunto de *tweets* de rejeição para cada candidato ao longo dos 7 debates.

### 3.3.3 Exemplos Obtidos na Coleta para o *Dataset* Rotulado

A seguir, exemplos de coletas *tweets* rotulados como positivos a partir das *hashtags* para cada um dos três candidatos acompanhados inicialmente pela análise. Na Figura 3.2 é apresentado o caminho percorrido do perfil @TeamCiroGomes, de apoio ao candidato Ciro Gomes, até o *tweet* de ID 1027640912585912320; na Figura 3.3 é mostrada a forma com que se captura o *tweet* de ID 1027623833136553984 a partir do perfil @BolsonaroSP, de apoio ao candidato Jair Bolsonaro; finalmente a Figura 3.4 mostra como chega-se ao *tweet* de ID 1027654576789446657 partindo do perfil @LulaOficial, do próprio candidato Luiz Inácio Lula da Silva. Os três exemplos foram

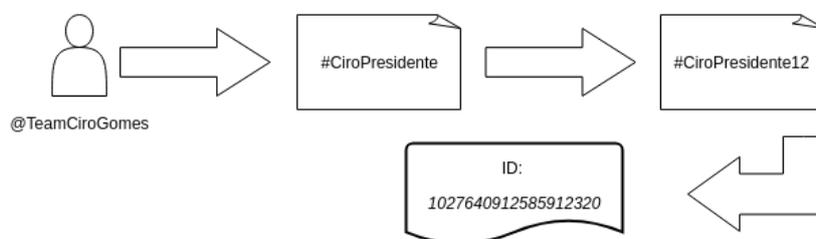
coletados na tarde da véspera do primeiro debate transmitido por televisão entre os presidenciáveis.

Tabela 3.5: *Tweets* Positivos: dados da coleta

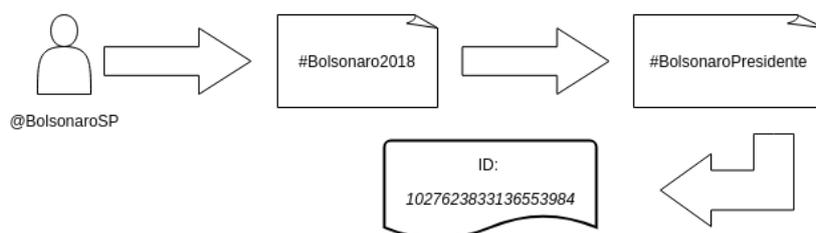
<b>Candidato</b>	<b>Perfis Pesquisados</b>	<b>Hashtags Mais Frequentes</b>	<b>Número de Tweets</b>
<b>Gomes</b>	@cirogomes, @KatiaAbreu, @TeamCiroGomes, @todoscomciro	#viraviraciro, #tsunamiciro, #cirogomespresidente, #ciropresidente12	60904
<b>Haddad</b>	@Haddad_Fernando, @ManuelaDavila, @LulaOficial, @ptbrasil	#obrasilfelizdenovo, #haddadpresidente, #haddadpresidente13, #eusouhaddad13	89231
<b>Bolsonaro</b>	@JairBolsonaro, @BolsonaroSP, @CarlosBolsonaro, @FlavioBolsonaro	#euvotobolsonaro, #euvotobolsonaro17, #bolsonaropresidente17, #estoucombolsonaro	218395
<b>Lula da Silva</b>	@Haddad_Fernando, @ManuelaDavila, @LulaOficial, @ptbrasil	#obrasilfelizdenovo, #lulaecandidato, #lulapresidente, #somostodoslula	26020

Tabela 3.6: *Tweets* Negativos: dados da coleta

<b>Candidato</b>	<b>Hashtags Mais Frequentes</b>	<b>Número de Tweets</b>
<b>Gomes</b>	#cironão, #ciromente, #cirojumento, #cirocoronel	60727
<b>Haddad</b>	#haddadnãõ, #haddadcorrupto, #haddadnunca, #haddadjamais	89131
<b>Bolsonaro</b>	#elenao, #bolsonaronãõ, #bolsonaronunca, #bolsonarojamais, #bolsomico	218284
<b>Lula da Silva</b>	#luladrão, #lulafichasuja, #lulapresosim, #lulaéladrão	25928

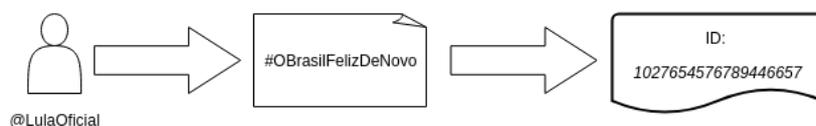
Figura 3.2: Exemplo da coleta de um *tweet* de apoio ao candidato Ciro Gomes

Fonte: O Autor

Figura 3.3: Exemplo da coleta de um *tweet* de apoio ao candidato Jair Bolsonaro

Fonte: O Autor

Figura 3.4: Exemplo da coleta de um *tweet* de apoio ao candidato Luiz Inácio Lula da Silva



Fonte: O Autor

### 3.4 Pré-processamento e Definição de *Features*

Como mencionado na seção 3.2, a maior parte dos atributos de um *tweet* coletado não seria utilizada no trabalho, mantendo-se apenas o campo de texto e da identificação de um *tweet*. O campo “*id*” foi utilizado apenas para a filtragem dos *tweets* na procura por duplicatas nas coletas. Já campo o “*text*” foi explorado para a extração de *features* para o treino do classificador.

Como pré-processamento, foram eliminados termos dentro das mensagens como os perfis de usuários mencionados, pontuação, *links*, caracteres especiais (exceto o #) e *stopwords*. Gírias foram traduzidas para os termos correspondentes através de um dicionário formulado pelo autor, como “vc” passando a ser “você”, e palavras com repetição de letras para demonstrar ênfase foram normalizadas através de expressões regulares. Foi aplicado o processo de lematização em cada termo presente nos *tweets*, através da biblioteca spaCy<sup>2</sup>. Além disso, foram eliminados os *tweets* que continham menos de 4 termos por entendermos que ofereceriam pouca variabilidade de informação ao processo de aprendizado dos classificadores.

Utilizou-se somente unigramas para o aprendizado do modelo classificador, baseado em testes preliminares com outros tipos de *features*. Os resultados das métricas obtidas somente com o unigrama foram superiores a testes com bigramas e prefixos, únicos e combinados.

<sup>2</sup><<https://spacy.io/>> Acesso em Outubro de 2018

### 3.5 Treinamento dos Classificadores

Para obter um melhor desempenho na classificação de sentimentos dos *tweets* coletados para o estudo, foram testados dois algoritmos de classificação no treinamento: SVM e *Naive Bayes*. Para cada candidato monitorado, foram construídos dois classificadores, todos eles treinados e testados utilizando o mesmo conjunto de instâncias rotuladas como positivo e negativo de cada presidenciável. A quantidade de *tweets* rotulados para cada um dos candidatos encontra-se nas Tabelas 3.5 e 3.6, respectivamente de apoio e rejeição, sendo que cada um dos conjuntos abrange os debates do primeiro turno em que presidenciável monitorado era de fato o candidato oficial registrado junto ao Tribunal Superior Eleitoral no momento do debate.

Os experimentos com os classificadores foram conduzidos utilizando a linguagem de programação Python e as bibliotecas NLTK<sup>3</sup> e Scikit Learn<sup>4</sup>.

Para o processo de avaliação dos modelos utilizou-se a técnica de *k-fold cross validation*, com  $k = 10$ . A cada iteração eram armazenados os TP, FP, TN e FN obtidos e ao final foram calculadas suas médias, sendo possível após isso obter as métricas de acurácia, precisão, *recall* e *F-measure*. Os resultados da avaliação para ambos os algoritmos classificadores *Naive Bayes* e SVM estão descritos respectivamente nas Tabelas 3.7 e 3.8.

Tabela 3.7: Resultados do classificador *Naive Bayes*

	Acurácia	Precisão		Recall		F-measure	
		POS	NEG	POS	NEG	POS	NEG
<b>Gomes</b>	0.702	0.671	0.745	<b>0.792</b>	0.611	0.726	0.671
<b>Haddad</b>	0.603	0.595	0.613	0.647	0.559	0.619	0.584
<b>Bolsonaro</b>	0.736	0.719	<b>0.756</b>	0.729	0.696	0.723	0.724
<b>Lula da Silva</b>	<b>0.743</b>	<b>0.734</b>	<b>0.751</b>	0.761	<b>0.724</b>	<b>0.747</b>	<b>0.737</b>
<b>Médias</b>	0.696	0.679	0.716	0.732	0.647	0.703	0.679

Os dois algoritmos classificadores utilizados no presente trabalho obtiveram bons resultados, considerando as métricas utilizadas na avaliação de seu desempenho. Em ambos, as métricas calculadas para o candidato Fernando Haddad foram as piores entre os 4 candidatos, enquanto os resultados de Lula foram os melhores.

<sup>3</sup><<https://www.nltk.org/>> Acesso em Agosto de 2018

<sup>4</sup><<https://scikit-learn.org/>> Acesso em Setembro de 2018

Tabela 3.8: Resultados do classificador SVM

	Acurácia	Precisão		Recall		F-measure	
		POS	NEG	POS	NEG	POS	NEG
<b>Gomes</b>	0.677	0.645	<b>0.728</b>	<b>0.790</b>	0.565	<b>0.710</b>	0.636
<b>Haddad</b>	0.547	0.546	0.548	0.564	0.530	0.554	0.538
<b>Bolsonaro</b>	<b>0.690</b>	0.678	0.702	0.722	0.657	0.699	0.678
<b>Lula da Silva</b>	<b>0.691</b>	<b>0.693</b>	0.689	0.688	<b>0.694</b>	0.690	<b>0.691</b>
<b>Médias</b>	0.651	0.640	0.666	0.691	0.611	0.663	0.635

Na classificação com SVM os candidatos Fernando Haddad e Luis Inácio Lula da Silva obtiveram um desempenho semelhante para ambas as classes tanto para precisão quanto para *recall*. Portanto, a *F-measure* de ambos seguiu a mesma proporção. Para o candidato Jair Bolsonaro, embora a *F-measure* seja semelhante para ambas as classes, existem diferenças de 2 e 7 pontos percentuais (pp) entre as classes positiva e negativa no que tange precisão e *recall*, respectivamente. Em termos de *F-measure* o resultado para Lula e Bolsonaro são parecidos. Já o classificador para o candidato Ciro Gomes obteve as métricas mais desproporcionais, mostrando uma baixa capacidade de reconhecer a classe negativa (56.5%), ao mesmo tempo que a classe positiva é bem reconhecida (79%).

Utilizando o classificador de *Naive Bayes* observa-se que as métricas foram levemente superiores. De mesma forma como SVM, os candidatos Luiz Inácio Lula da Silva e Fernando Haddad obtiveram métricas de *recall* e precisão balanceadas para ambas as classes, sendo os resultados de Lula os melhores no geral, e os de Haddad os piores. O candidato Jair Bolsonaro obteve um bom *F-measure*, equilibrado para ambas as classes, mas estas tiveram leves diferenças entre precisão (melhor para a classe negativa) e *recall* (melhor para a classe positiva). Mesmo assim, são resultados comparáveis aos obtidos para o candidato Lula em termos de *F-measure*. Novamente o candidato Ciro Gomes apresentou o desempenho mais desequilibrado em termos das classes positiva e negativa, sendo o mesmo melhor para a classe positiva em termos de *recall* e *F-measure*, e melhor em termos de precisão para a classe negativa.

Como as métricas dos classificadores gerados com o *Naive Bayes* apresentaram melhores resultados comparados ao classificador SVM, decidiu-se desenvolver a análise dos *tweets* coletados de cada candidato utilizando os classificadores desenvolvidos com este algoritmo.

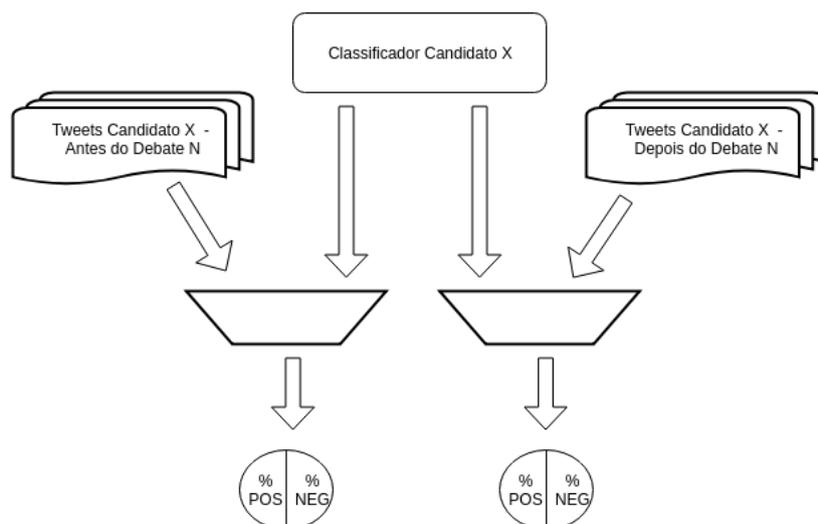
### 3.6 Aplicação do Classificador aos Dados do Experimento

A partir do treinamento do classificador utilizando o algoritmo de *Naive Bayes* e unigramas como único tipo de *feature*, previu-se a polaridade dos *tweets* coletados sobre os candidatos, descritos na Tabela 3.4.

O processo que aplica o modelo classificador nos dados não rotulados, para o experimento, é ilustrado na Figura 3.5. Na imagem, observa-se a aplicação do classificador do candidato X, em um debate N, juntamente com o conjunto de *tweets* deste candidato X, antes e depois do evento. Para cada conjunto de *tweets*, de um candidato para uma fase do debate, o classificador rotula uma amostra por vez e o resultado é a porcentagem de *tweets* rotulados como positivos no conjunto e a porcentagem de negativos. A métrica representada pela equação 3.1, utilizada para sumarizar os sentimentos sobre cada candidato e traçar a evolução da sua popularidade ao longo dos debates, foi retirada de (TUMITAN; BECKER, 2014). Na equação, Q representa o conjunto de *tweets* de um período de um debate do candidato j.

$$sentimento_{jQ} = \frac{\sum_{t \in Q} pos_{jt}}{\sum_{t \in Q} (pos_{jt} + neg_{jt})} \quad (3.1)$$

Figura 3.5: Processo de classificação de um conjunto de *tweets*



Fonte: O Autor

A Tabela 3.9 mostra os sentimentos obtidos com os classificadores para cada candidato em cada momento, anterior e posterior, de cada debate. Cada sentimento na Tabela 3.9 é demonstrado como o percentual de *tweets* rotulados como positivos do

conjunto de dados coletados de cada candidato em um dos períodos, antes ou depois, de cada debate. Uma análise sobre o comportamento dos valores mostrados na Tabela 3.9 será apresentada no Capítulo 4.

Tabela 3.9: Percentuais de aprovação obtidos a partir da aplicação do modelo

	<b>Gomes</b>		<b>Haddad</b>		<b>Bolsonaro</b>		<b>Lula da Silva</b>	
	<b>Antes</b>	<b>Depois</b>	<b>Antes</b>	<b>Depois</b>	<b>Antes</b>	<b>Depois</b>	<b>Antes</b>	<b>Depois</b>
<b>Debate 1</b>	68.2%	74.7%	X	X	65.3%	72.8%	56.6%	47.1%
<b>Debate 2</b>	64.4%	62.6%	X	X	71.4%	75.2%	46.5%	52.1%
<b>Debate 3</b>	66.9%	61.8%	X	X	81.2%	76.6%	67.9%	69.9%
<b>Debate 4</b>	58.1%	62.4%	71.8%	66.5%	58.8%	48.8%	X	X
<b>Debate 5</b>	71.1%	65.5%	62.3%	67.4%	47.3%	48.9%	X	X
<b>Debate 6</b>	63.7%	67.4%	59.1%	56.9%	56.1%	63.6%	X	X
<b>Debate 7</b>	72.6%	76.3%	55.7%	64.0%	72.9%	69.4%	X	X

## 4 AVALIAÇÃO DOS RESULTADOS

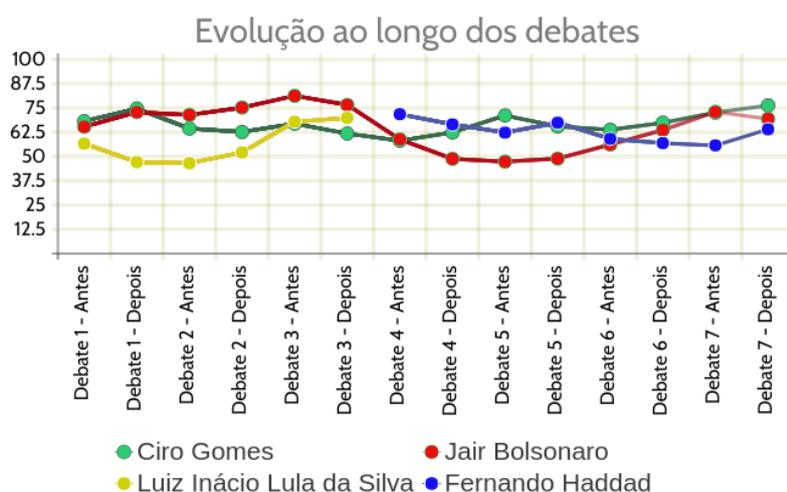
Este capítulo é dedicado à discussão dos resultados do experimento, que são mostrados na Tabela 3.9. A partir destes resultados pôde-se mensurar e sumarizar a mudança de opinião sobre os candidatos em função dos debates a partir dos *tweets* coletados de usuários da rede social Twitter.

Ao construir o classificador de sentimentos e aplicá-lo aos conjuntos de *tweets* coletados sobre os candidatos foi possível alcançar os meios necessários para atender os objetivos propostos para o presente trabalho. As análises propostas no Capítulo 3 podem ser executadas uma vez que já se têm subsídios para tal.

### 4.1 Aceitação e Rejeição dos Candidatos em Relação a um Debate

A primeira análise proposta visa mensurar o impacto na opinião pública sobre um candidato em relação a um debate, na visão dos usuários do Twitter. A Figura 4.1 ilustra o monitoramento dos 4 candidatos ao longo dos 7 debates, utilizando o percentual de mensagens classificadas como positivas de um candidato sobre o total das suas mensagens para cada respectivo evento. Este comportamento será detalhado ao longo desta seção.

Figura 4.1: Candidatos monitorados ao longo dos debates

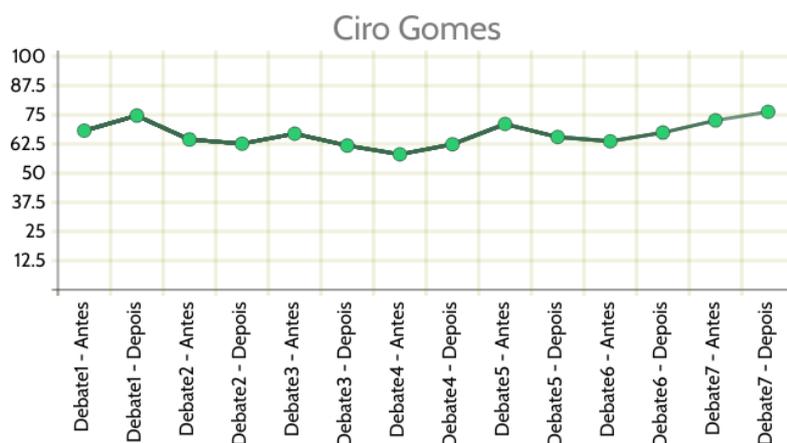


Fonte: O Autor

### 4.1.1 Ciro Gomes

O candidato Ciro Gomes foi o teve a opinião classificada como positiva bastante estável ao longo dos diferentes debates, mantendo uma porcentagem de aprovação entre 58.1% e 76.3%, com 2 picos ao longo do período. O primeiro pico aconteceu após o primeiro debate e o segundo, ocorreu após o sétimo evento. Uma possível explicação do último debate apresentar este pico de opinião positiva, considerando o período monitorado, é a impulsão da *hashtag* #ViraViraCiro, em uma tentativa de engajamento para levar o candidato ao segundo turno das eleições. A Figura 4.2 apresenta a porcentagem de opiniões classificadas como positivas ao candidato Ciro Gomes ao longo dos 7 debates.

Figura 4.2: Evolução da opinião sobre Ciro Gomes ao longo dos 7 debates



Fonte: O Autor

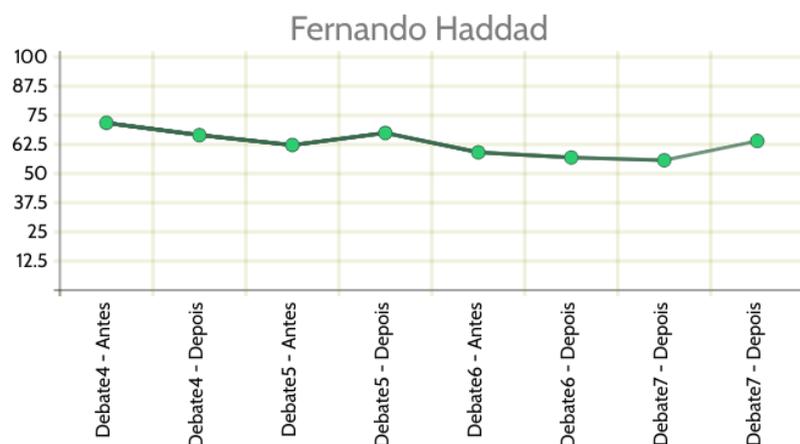
### 4.1.2 Fernando Haddad

A Figura 4.3 mostra que o candidato Fernando Haddad inicia sua participação somente no quarto debate dentre aqueles monitorados neste trabalho, com uma boa opinião pública (71.8% dos *tweets* a seu respeito). Entretanto, a monitoração nos debates seguintes aponta que sua aprovação caiu para 66.5%. No 5º debate o candidato consegue ganhar aprovação, porém no 6º volta a perder popularidade, terminando com 2.2% a menos de aprovação em relação ao período pré-debate 6.

No 7º debate o candidato da chapa PT / PCdoB tem seu melhor desempenho,

iniciando com 55,7% dos *tweets* classificados como positivos e terminando com 64%.

Figura 4.3: Evolução da opinião sobre Fernando Haddad ao longo dos 4 debates



Fonte: O Autor

#### 4.1.3 Jair Bolsonaro

Ao analisar a Figura 4.4 nota-se que os índices de aceitação do candidato Jair Bolsonaro foram os que tiveram maior variabilidade ao longo dos debates. Devido a uma agressão sofrida durante o período de monitoramento, o candidato acabou participando somente dos dois primeiros debates.

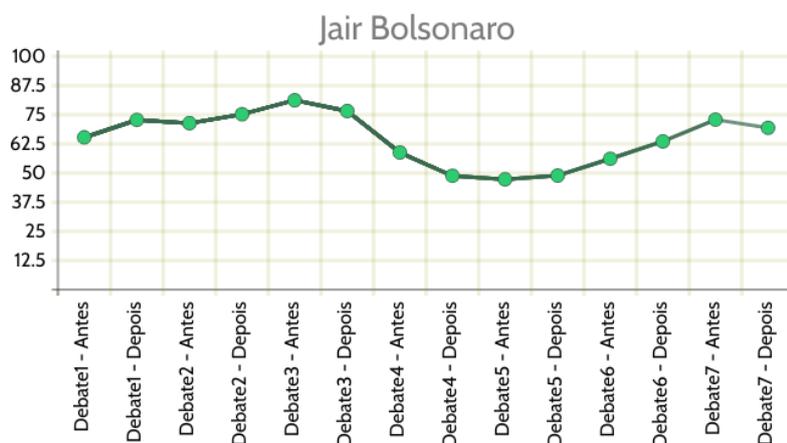
Em ambos os debates nos quais Jair Bolsonaro participara, ele obteve uma boa aceitação, aumentando respectivamente 7.5% e 3.8% da mesma em comparação ao início do evento. No período prévio ao 3º debate ocorreu o pico de popularidade do candidato com 81.2% de aprovação, o que pode ser talvez explicado por uma comoção em relação à agressão ocorrida dias antes.

O período posterior ao terceiro debate registra uma queda de 4.6% em sua popularidade, sendo seguido por uma queda vertiginosa de 10% no quarto debate, iniciando com 58.8% e registrando no período pós-debate um índice de 48.8% de aprovação (neste caso, a rejeição já era maior, com o valor de 51.2%). O debate 5 registrou um aumento na popularidade do candidato, porém os dois períodos de coleta ainda registravam uma taxa de rejeição maior que a de aprovação. Entre os fatores que podem ser usados para talvez explicar esta queda repentina de popularidade encontram-se o fato da ausência do candidato nos debates impossibilitando sua defesa e,

as declarações polêmicas sobre impostos de Paulo Guedes, assessor de Jair Bolsonaro.

O 6º debate já registrava que a popularidade do candidato já estava crescendo novamente, com início de 56.1% e término em 63.6%, enquanto o 7º debate mostrou uma pequena queda de aprovação ao final, porém mesmo assim com um índice considerável de 69.4%.

Figura 4.4: Evolução da opinião sobre Jair Bolsonaro ao longo dos 7 debates



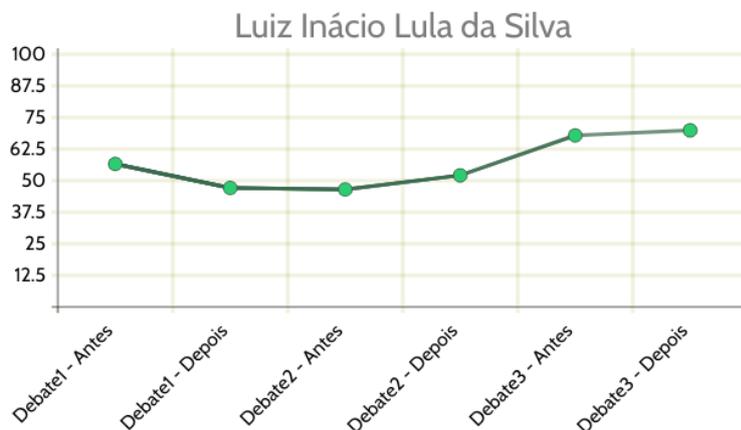
Fonte: O Autor

#### 4.1.4 Luiz Inácio Lula da Silva

O candidato Luiz Inácio Lula da Silva, primeiro postulante a presidente registrado na chapa PT / PCdoB, registrou nos dois primeiros debates uma alternância entre índices de aprovação e rejeição, com um sobrepondo-se ao outro. A Figura 4.5 ilustra tal situação, onde no período prévio ao 1º debate é registrado um índice de aprovação de 56.6%, porém o período posterior do mesmo apresenta uma queda e fechamento em 47.1% (ou 52.9% de rejeição). Já no 2º debate a situação inverte-se: o período prévio do debate marca índice de aprovação de 46.5% (e portanto índice de rejeição de 53.5%), mas o período posterior acaba obtendo um valor de 52.1% de aprovação.

O 3º debate registra um crescimento no índice de aprovação do candidato Luiz Inácio Lula da Silva, tanto em relação ao fechamento do 3º debate, quanto comparado ao índice registrado no período prévio do debate em questão. Uma possível explicação para tal acontecimento pode ser pelo fato de que o candidato estivesse prestes a ser substituído em sua chapa por Fernando Haddad, recebendo então um último apoio de seus eleitores.

Figura 4.5: Evolução da opinião sobre Luiz Inácio Lula da Silva ao longo dos 3 debates



Fonte: O Autor

## 4.2 Comparações com as Pesquisas Eleitorais

A segunda análise proposta no Capítulo 3 tem o caráter de comparação dos resultados obtidos no experimento com as pesquisas eleitorais registradas junto ao TSE. A motivação para se chegar a índices equivalentes às pesquisas de intenção de voto utilizando somente dados de uma rede social se dá pelo fato de que se for possível, a forma como a intenção de voto é mensurada hoje poderia sofrer transformações benéficas.

As mudanças introduzidas pelo sucesso da análise abrangeriam fatores econômicos, devido aos custos envolvidos com a equipe física de pesquisas, e de praticidade, já que qualquer equipe de campanha de um candidato estaria apta a mensurar as intenções de voto de forma autônoma e independente.

Para realizar a comparação dos *tweets* de apoio a um candidato com a porcentagem de intenções de voto do mesmo em uma pesquisa eleitoral, deve-se normalizar os valores obtidos no experimento. Primeiramente, definiu-se que, para calcular o percentual de intenção de voto, para cada fase, de cada um dos debates, o universo de votos seria: a soma de *tweets* coletados para tal fase, rotulados como positivos pelo classificador, de cada um dos candidatos monitorados, obtendo os valores com os percentuais da Tabela 3.9, aplicados aos tamanhos dos conjuntos na Tabela 3.4.

No presente trabalho decidiu-se analisar os candidatos das 3 chapas que concorreram à presidência da república e lideraram as intenções de voto na parte final da

campanha. Para poder comparar os índices obtidos no experimento com os resultados observados nas pesquisas eleitorais, porém, deve-se considerar todos os candidatos elegíveis. Para solucionar o problema causado pela ausência dos votos dos candidatos não monitorados, foi calculada a média de votos dos 3 primeiros colocados no primeiro turno das últimas 6 eleições (1994 a 2014), chegando a 94.16% do total de votos. Este valor de 94.16% foi utilizado como sendo o percentual que representa a soma dos *tweets* de apoio aos 3 candidatos monitorados neste trabalho, para cada fase de cada debate analisado. Tem-se, então, que a soma dos percentuais de voto dos candidatos não monitorados representa 5.84%.

A Tabela 4.1 sumariza os percentuais de opinião pública baseados da Tabela 3.9, transformados em percentuais de voto para cada fase e debate do período analisado. O processo de transformação do percentual de aprovação da opinião pública de um candidato obtido pelo classificador, para um índice de intenção de voto estimado, é realizado a partir da equação 4.1. Na equação, a intenção de votos estimada para um candidato  $j$  a partir de uma fase e debate  $Q$ , é obtida calculando a soma dos *tweets* do conjunto  $Q$  que pertencem ao candidato  $j$ , pelo tamanho total de  $Q$ . Ainda, o resultado é multiplicado por 0.9416, o percentual que corresponde à soma dos votos dos 3 primeiros colocados em uma eleição.

$$intencaoVoto_{jQ} = \frac{\sum_{t \in Q} (t \in j) * 0.9416}{size(Q)} \quad (4.1)$$

Tabela 4.1: Percentuais de intenção de voto a partir dos resultados do trabalho

	<b>Gomes</b>		<b>Haddad</b>		<b>Bolsonaro</b>		<b>Lula da Silva</b>	
	<b>Antes</b>	<b>Depois</b>	<b>Antes</b>	<b>Depois</b>	<b>Antes</b>	<b>Depois</b>	<b>Antes</b>	<b>Depois</b>
<b>Debate 1</b>	6.4%	30.46%	X	X	70.88%	51.31%	16.86%	12.37%
<b>Debate 2</b>	8.08%	14.02%	X	X	40.55%	60.46%	45.51%	19.66%
<b>Debate 3</b>	14.27%	25.94%	X	X	70.28%	48.85%	9.59%	19.36%
<b>Debate 4</b>	15.36%	20.81%	30.24%	36.13%	48.55%	37.21%	X	X
<b>Debate 5</b>	21.90%	25.07%	31.04%	23.46%	41.21%	45.62%	X	X
<b>Debate 6</b>	15.60%	19.68%	40.73%	34.05%	37.81%	40.42%	X	X
<b>Debate 7</b>	18.58%	25.72%	22.58%	23.45%	52.98%	44.97%	X	X

### 4.2.1 Comparações por Debate

Para realizar a comparação das intenções de votos de pesquisas eleitorais registradas junto ao TSE com as estimativas calculadas para o presente trabalho foram utilizadas como base as pesquisas da Tabela 4.2. No dia 31/08, a candidatura de Luiz Inácio Lula da Silva foi indeferida pelo TSE, fazendo com que as pesquisas eleitorais envolvendo o nome do candidato fossem cessadas antes do 3º debate acontecer. Por observar que a distribuição das intenções de voto das pesquisas oficiais mudou bruscamente com a troca de Fernando Haddad por Luiz Inácio Lula da Silva, o debate 3 não foi utilizado para motivos de comparação.

#### 4.2.1.1 Debate 1: Band

A estimativa obtida na Tabela 4.1 para o período pré-debate do candidato Ciro Gome obteve uma grande semelhança, marcando 6.4% contra 6% da pesquisa oficial. Contando com a margem de erro dos pontos percentuais de 3.2% estimado pela pesquisa oficial, pode-se afirmar que o presente trabalho estimou o percentual de forma correta. O período pós-debate para o mesmo candidato, no entanto, mostrou uma métrica muito distante, com 7% contra 30.46% do experimento. O que se pôde observar em comum na comparação é a tendência de que Ciro Gomes causou impacto positivo na opinião pública, subindo os índices de intenção tanto na pesquisa oficial quanto no índice apresentado na Tabela 4.1.

O candidato Luiz Inácio Lula da Silva apresentou percentuais nas pesquisas eleitorais oficiais muito acima das estimadas no experimento. O candidato registrou antes e depois do debate índices de 31% e 29%, e 16.86% e 12.37%, respectivamente na pesquisa oficial e na estimativa do presente trabalho. Assim como o ocorrido com Ciro Gomes, o experimento registrou a tendência da evolução da opinião, porém, com o candidato petista foi observada a queda da intenção de voto, não a ascensão.

O candidato Jair Bolsonaro obteve o pior resultado entre os candidatos monitorados para o debate em questão. No experimento o mesmo estava sendo apontado como líder de intenções de voto e, após o debate, era registrada uma queda na popularidade. A pesquisa oficial, entretanto, mostrou que o candidato se manteve na segunda colocação das intenções de voto, não tendo sofrido alterações pelo seu desempenho no debate.

Tabela 4.2: Pesquisas de intenção de voto utilizadas como base para comparação

<b>Debate</b>	<b>Período de Execução</b>	<b>Empresa</b>	<b>Margem de Erro</b>	<b>Ciro Gomes</b>	<b>Fernando Haddad</b>	<b>Jair Bolsonaro</b>	<b>Luiz Inácio Lula da Silva</b>
<b>Debate 1 - Antes</b>	06/08 a 08/08	XP/ Ipespe	3.2%	6%	X	19%	31%
<b>Debate 1 - Depois</b>	10/08 a 12/08	RealTime Big Data	2%	7%	X	19%	29%
<b>Debate 2 - Antes</b>	13/08 a 15/08	XP/ Ipespe	3.2%	7%	X	20%	31%
<b>Debate 2 - Depois</b>	20/08 a 21/08	Datafolha	2%	5%	X	19%	39%
<b>Debate 4 - Antes</b>	18/09 a 19/09	Datafolha	2%	13%	16%	28%	X
<b>Debate 4 - Depois</b>	22/09 a 23/09	FSB Pesquisa	2%	10%	23%	33%	X
<b>Debate 5 - Antes</b>	24/09 a 26/09	XP/ Ipespe	2.2%	11%	21%	28%	X
<b>Debate 5 - Depois</b>	27/09 a 28/09	CNT/ MDA	2%	9.4%	25.2%	28.2%	X
<b>Debate 6 - Antes</b>	29/09 a 30/09	IBOPE	2%	11%	21%	31%	X
<b>Debate 6 - Depois</b>	01/10 a 02/10	IBOPE	2%	10%	23%	32%	X
<b>Debate 7 - Antes</b>	03/10 a 04/10	Datafolha	2%	11%	22%	35%	X
<b>Debate 7 - Depois</b>	05/10 a 06/10	Datafolha	2%	13%	22%	36%	X

#### *4.2.1.2 Debate 2: RedeTV!*

Para o candidato *Ciro Gomes*, como aconteceu no 1º debate, a estimativa da Tabela 4.1 obteve um grande grau de acerto, registrando 8.08% contra 7% da pesquisa oficial. O resultado do experimento pode ainda ser considerado como correto, contando novamente com os 3.2 pontos percentuais de margem de erro calculados pelo instituto que realizou a pesquisa. O período pós-debate, porém, mostrou caminhos opostos tomados pelos dois índices comparativos. Enquanto a pesquisa oficial registrava uma queda na popularidade do candidato de 2%, a estimativa do experimento deste trabalho mostrou um aumento de 5.94% nas intenções de voto.

O candidato *Luiz Inácio Lula da Silva*, desta vez, pôde ser apontado como líder das intenções de voto na fase pré-debate, com 45.51% dos votos, porém muito distante dos 31% marcados na pesquisa oficial. O período pós-debate apresentou algo semelhante ao encontrado nos dados do candidato *Ciro Gomes*: os números da pesquisa oficial e da estimativa do experimento tomaram caminhos opostos. Enquanto a pesquisa oficial mostra um aumento de 8% na popularidade, o experimento registrou uma queda de 25.85%.

Assim como no debate 1, o candidato *Jair Bolsonaro* apresentou os piores resultados em comparação à pesquisa oficial. Nem os valores dos índices foram próximos nem a tendência de evolução ou queda foi observada na comparação.

#### *4.2.1.3 Debate 4: TV Aparecida*

*Ciro Gomes* obteve mais uma vez um índice pré-debate extremamente semelhante. Enquanto que na pesquisa oficial foi registrado 13% das intenções de voto, chegando a 15% com a margem de erro dos pontos percentuais, a estimativa do experimento chegou a 15.36%. A evolução das intenções de voto observada no pós-debate na pesquisa oficial, porém, foi negativa, caindo 3%, enquanto na estimativa proposta por este trabalho a popularidade do candidato subiu em 5.45%.

O candidato *Fernando Haddad* em seu primeiro debate registrou, tanto nas pesquisas oficiais quanto na estimativa do presente trabalho, uma ascensão nas intenções de votos após o debate comparado com o período pré evento. A precisão da intenção de votos estimada, porém, mostrou um índice pré-debate 14.24 pontos percentuais a menos e um índice pós-debate 13.13 pontos percentuais inferior, ao comparar com as medidas das pesquisas oficiais.

Os índices do candidato Jair Bolsonaro tanto na pesquisa oficial quanto na estimativa calculada foram muito diferentes, mas ambos mostrando a liderança do mesmo nas intenções de voto. A pesquisa oficial pós-debate mostrou um aumento de 5 pontos percentuais na popularidade do candidato, chegando a 33%, já na estimativa deste trabalho o período após o 4º debate registra uma queda de 48.55% para 37,21%. Contando com a margem dos pontos percentuais da pesquisa oficial, o índice de intenção de votos do candidato chega a 35%, muito aproximado do valor informado na Tabela 4.1 para o candidato no pós-debate.

#### *4.2.1.4 Debate 5: SBT*

No debate realizado pelo SBT, os resultados do candidato Ciro Gomes apresentaram as piores comparações dentre todos os debates. A estimativa gerada por este trabalho para o candidato no período pré-debate SBT é de 21.90%, muito acima do registro de 11% da pesquisa oficial. Já no período pós-debate, o índice estimado foi de 25.07%, enquanto na pesquisa oficial para o mesmo período consta o valor de 9.4%. Tanto os valores precisos quanto a evolução do impacto de debate na opinião não foram observados neste caso.

O candidato Jair Bolsonaro apareceu na liderança das intenções de voto novamente nos dois ambientes da comparação. No período pré-debate a estimativa apontava 41.21% dos votos, aumentando para 45.62% pós-debate. Já na pesquisa oficial anterior ao debate marcava o candidato com 28% das intenções de voto, aumentando 0.2% após o evento. A estimativa não conseguiu chegar nos valores de intenção de voto observados na pesquisa oficial, tendo como acertos somente a liderança dos votos e a ocorrência do crescimento da positividade da opinião pública a partir do debate.

Para o candidato Fernando Haddad a estimativa do período pré-debate diferiu em 10.04% da pesquisa oficial. Enquanto que o período pós-debate retratou a ascensão da popularidade do candidato a partir da pesquisa eleitoral oficial com o valor de 25.2%, a estimativa do experimento para o mesmo período chega em 23.46% das intenções de voto, representando uma queda. Com a margem de erro de 2% da pesquisa oficial, pode-se afirmar que a popularidade pós-debate do candidato no presente debate observada na Tabela 4.1 é correta.

#### 4.2.1.5 Debate 6: Record

Como no último debate, a estimativa mostrou valores muito diferentes do candidato Ciro Gomes em comparação à pesquisa eleitoral oficial, seja para o período pré-debate, seja para o período pós-debate. A evolução da popularidade do candidato em relação ao debate também não foi observada: enquanto na pesquisa oficial houve a queda de 1 ponto percentual das intenções de voto após o debate, a estimativa apontou um aumento de 4.08%.

Para o candidato Jair Bolsonaro a pesquisa oficial e o resultado apontado pelo presente trabalho registram um crescimento das intenções de voto do candidato no período pós-debate (respectivamente, aumentos de 1% e 2.61%). A correção dos valores, porém, permanece distante, com 37.81% e 40.42% obtidos pela estimativa do experimento nos períodos pré e pós-debate, e 31% e 32% na pesquisa oficial.

As estimativas dadas pelo experimento para o candidato Fernando Haddad no presente debate não se mostraram fiéis às pesquisas oficiais. No período pré-debate, a estimativa calculou o valor de 40.73% para a intenção de votos, contra 21% da pesquisa oficial. Já para o período pós-debate, a estimativa capturou uma queda de popularidade, chegando em 34.05%, enquanto na pesquisa oficial observou-se um crescimento para 23%.

#### 4.2.1.6 Debate 7: Globo

Para o debate 7, as estimativas de pré e pós-debate do candidato Fernando Haddad foram corretas em relação à pesquisa oficial. O índice da pesquisa pré-debate apontou um valor de 22% contra 22.58% do presente trabalho, enquanto que o valor para o candidato na pesquisa eleitoral pós-debate apresenta o resultado ainda de 22%, contra 23.45% da estimativa. Contando com a margem de erro percentual da pesquisa, ambos os valores se mostram corretos no último debate do 1º turno das eleições.

A estimativa do candidato Ciro Gomes não resultou em valores semelhantes aos observados na pesquisa oficial, porém, o crescimento da aceitação do candidato no período pós-debate foi constatada em ambos os casos.

Os índices referentes ao candidato Jair Bolsonaro não puderam refletir os mesmos das duas pesquisas eleitorais do caso. Enquanto observou-se um crescimento na opinião popular positiva em relação ao candidato, a partir da pesquisa oficial, os resultados obtidos no experimento mostraram uma queda na intenção de votos em relação ao período pré-

debate.

#### **4.2.2 Comparativo das Evoluções de Intenção de Voto: Estimado x Pesquisa Oficial**

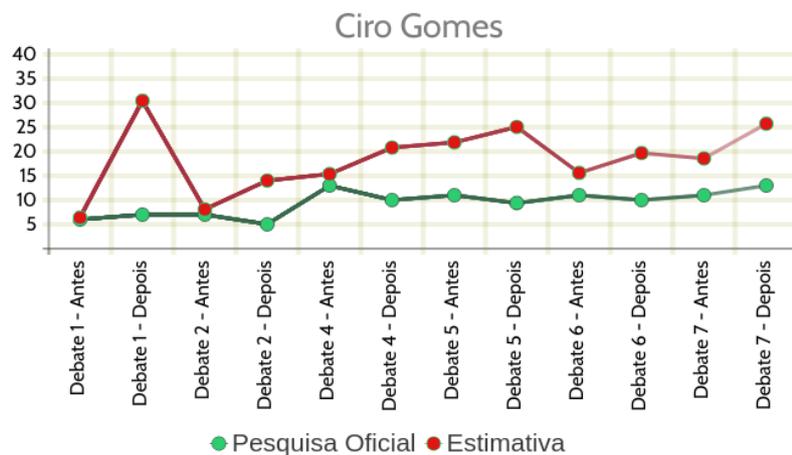
A partir das análises de cada um dos debates, para cada um dos candidatos, comparando a estimativa da intenção de votos calculada neste trabalho e os índices das pesquisas eleitorais oficiais cabíveis para cada um dos períodos, nota-se que o classificador não obteve resultados tão precisos. De modo geral, observa-se que os índices estimados não puderam ser replicados à imagem dos dados das pesquisas oficiais, porém, a evolução da opinião pública positiva de um candidato por diversas vezes teve o mesmo resultado.

##### *4.2.2.1* *Ciro Gomes*

A estimativa calculada no presente trabalho apresentou, para o candidato *Ciro Gomes*, uma média de 9.02 pontos percentuais superior em relação aos índices do candidato nas pesquisas utilizadas como base da análise, presentes na Tabela 4.2. A Figura 4.6 apresenta a evolução das intenções de voto do candidato ao longo do período monitorado (exceto em relação ao debate 3).

As intenções de voto do período pré-debate 1, pré-debate 2, pré-debate 4 e pré-debate 6 da estimativa apresentaram resultados exatos (dentro da margem de erro da pesquisa) ou muito próximos. Os índices dos debates 1 e 7 mostraram a mesma tendência de ter ganho popularidade após a realização dos respectivos eventos.

Figura 4.6: Comparativo da Evolução da intenção de votos de Ciro Gomes

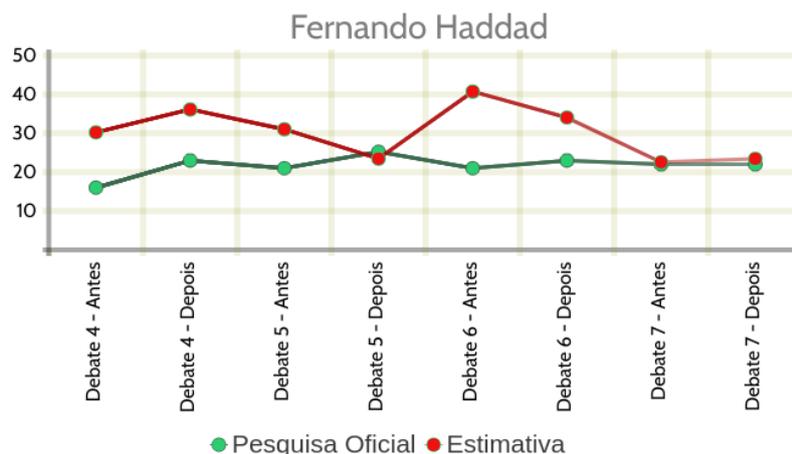


Fonte: O Autor

#### 4.2.2.2 Fernando Haddad

Para o candidato Fernando Haddad, a diferença média de pontos percentuais entre a estimativa e pesquisa oficial foi de 8.56 pontos a mais nos resultados do presente trabalho. Na Figura 4.7 é observada a comparação das intenções de voto a partir do primeiro debate em que o candidato participou.

Figura 4.7: Comparativo da Evolução da intenção de votos de Fernando Haddad



Fonte: O Autor

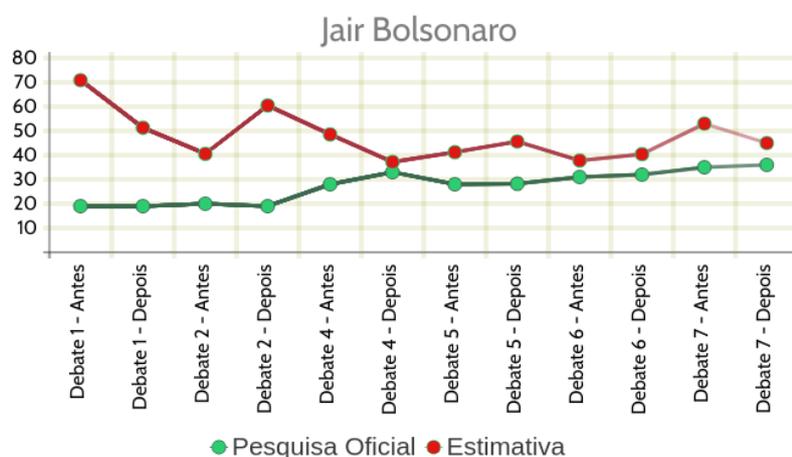
A partir da Figura 4.7 observa-se a exatidão da estimativa em relação à pesquisa oficial no período após o debate 5 e no pré e pós-debate 7. Além, no debate 4 nota-se a evolução positiva na intenção de votos ocorrida após o evento, presente tanto na pesquisa

oficial quanto na estimativa calculada.

#### 4.2.2.3 *Jair Bolsonaro*

A estimativa calculada para o candidato Jair Bolsonaro apresentou um resultado muito diferente em relação à pesquisa oficial, com média de 20.31 pontos percentuais superior a cada estimativa. A Figura 4.8 apresenta a evolução das intenções de votos, das pesquisas e estimadas, para o candidato Jair Bolsonaro.

Figura 4.8: Comparativo da Evolução da intenção de votos de Jair Bolsonaro



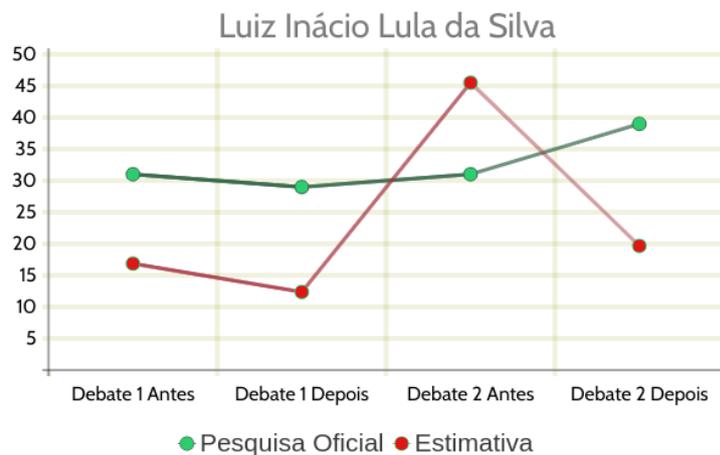
Fonte: O Autor

Não foram observadas estimativas que fossem exatas com os índices de intenção de voto do candidato nas pesquisas oficiais, apenas alguns comportamentos da evolução de popularidade referente a um debate. Nos debates 5 e 6 há a réplica do comportamento observado a partir das pesquisas oficiais, ocorrendo um crescimento das intenções de voto do candidato.

#### 4.2.2.4 *Luiz Inácio Lula da Silva*

A Figura 4.9 mostra que para o candidato Luiz Inácio Lula da Silva os índices estimados não conseguiram alcançar resultados satisfatórios e, o único aspecto semelhante aos dados das pesquisas oficiais foi a tendência da queda nas intenções de voto do candidato após o debate 1.

Figura 4.9: Comparativo da Evolução da intenção de votos de Luiz Inácio Lula da Silva



Fonte: O Autor

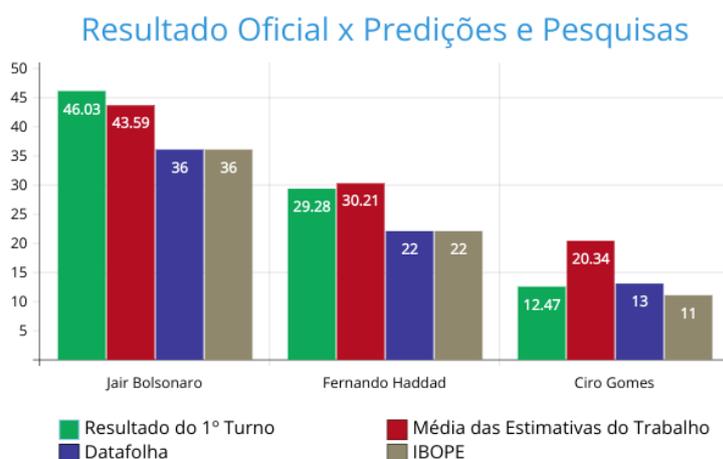
#### 4.3 Predição do Resultado do 1º Turno

A partir dos resultados obtidos com o classificador desenvolvido no Capítulo 3 e discutidos ao longo do Capítulo 4, foi observado que as estimativas não foram razoáveis ao comparar com as pesquisas eleitorais nos períodos monitorados. Entretanto, o resultado oficial do 1º turno da eleição presidencial brasileira de 2018 mostrou que as pesquisas eleitorais oficiais não retrataram fielmente os índices da votação final.

O fato descrito no parágrafo anterior fez levantar um questionamento adicional ao presente trabalho: “O resultado do 1º turno da eleição poderia ser predito com maior precisão coletando e classificando *tweets* dos candidatos com o classificador desenvolvido neste trabalho?”

Tendo o questionamento em mente, calculou-se a média das intenções de voto de cada um dos candidatos nos 8 conjuntos de *tweets* onde havia disputa de votos entre Ciro Gomes, Fernando Haddad e Jair Bolsonaro. A saber, os conjuntos de *tweets* utilizados foram dos períodos pré e pós-debates 4 a 7. A Figura 4.10 mostra um comparativo do resultado oficial de votação do 1º turno das eleições de cada um dos três candidatos monitorados, da média das estimativas comentadas anteriormente, além das últimas pesquisas IBOPE e Datafolha antes do dia da votação.

Figura 4.10: Comparativo entre Resultado do 1º Turno, Média dos índices estimados e Pesquisas Eleitorais



Fonte: O Autor

Observa-se no gráfico da Figura 4.10 que as métricas mais próximas dos resultados oficiais do 1º turno dos dois primeiros colocados são as obtidas pelo presente trabalho. Enquanto que para o candidato Jair Bolsonaro as duas pesquisas eleitorais de intenção de voto apresentaram um índice com mais de 10 pontos percentuais a menos, a métrica estimada no experimento diferenciou-se por 2.44 pontos. Já para o segundo colocado, Fernando Haddad, a diferença das pesquisas para a votação final era de 7.28 pontos percentuais, ao passo que o índice estimado diferiu em 0.93 pontos a mais. Para o candidato Ciro Gomes a métrica obtida por este trabalho não obteve o mesmo sucesso, sendo o índice analisado mais distante do valor real.

Há diversos fatores que podem explicar a maior fidelidade com o resultado final do 1º turno para as métricas obtidas do presente trabalho, em comparação às pesquisas de intenção de voto. A maior utilização das redes sociais por brasileiros, uma nova onda política no cenário brasileiro, as situações e estratégias tomadas pelos partidos refletirem de modo diferente nas redes sociais, divulgação excessiva de fake news sobre um candidato e a construção de um candidato para um público em específico são alguns dos exemplos que podem ser discutidos para explicar os resultados do presente trabalho.

## 5 CONCLUSÕES

Neste trabalho foi apresentada a construção de um modelo classificador cujo objetivo é mensurar a popularidade de um candidato a partir das opiniões expressas sobre ele, e, particularmente, observar o impacto que um debate ocasiona nos índices de aprovação de um presidencial.

O conjunto de dados rotulados necessário para o treinamento do modelo classificador foi construído de forma automática, sem a presença explícita de anotadores. Cada amostra do conjunto rotulado, tanto para a classe de sentimento positivo, quanto para a classe de negativo, foi obtida pela busca de *tweets* com alguma *hashtag* que caracterizasse o sentimento sobre o candidato. O método utilizado mostrou-se uma grande possibilidade a ser explorada em cenários onde não há disponibilidade de anotador algum ou um número inferior ao que é preciso para se obter um grande volume de dados.

A partir dos resultados de desempenho do classificador, observou-se que as métricas de avaliação são bastante promissoras, embora ainda passíveis de melhoria. O estudo de mais tipos de *features* a serem aplicadas ao modelo, além de mais dados rotulados para o treinamento podem aumentar as métricas dos resultados de forma considerável. A relação dos resultados obtidos de sentimento com os respectivos índices de intenção de voto das pesquisas oficiais puderam ser observados também pelo presente experimento, assim como algumas tendências na evolução da popularidade de um candidato. Outro dado relevante oriundo deste trabalho foi a predição do resultado final do 1º turno da eleição a partir das médias das intenções de voto calculadas pelo classificador.

Um dos principais questionamentos presentes no desenvolvimento do trabalho foi a influência de não poder afirmar que a população representada pelo *dataset* dos *tweets* represente uma amostra representativa da população brasileira. Este fato torna o resultado estatisticamente duvidoso por não poder inferir-se nenhum fato com algum grau de certeza.

Deve-se considerar também que o cenário eleitoral não foi totalmente representado. Para o experimento proposto no trabalho foram monitorados somente os três candidatos que lideraram as pesquisas eleitorais na data de 02/09, e posteriormente terminaram com as três maiores votações ao final do 1º turno. Uma análise completa compreenderia tantos candidatos quanto fossem possíveis de se monitorar.

Outro fator que influenciou os resultados do trabalho foi a restrição de um *tweet* somente poder ser classificado em positivo ou negativo, não neutro. Esta limitação deu-se pelo método da construção do conjunto rotulado para a fase de treinamento e teste, onde foram utilizadas *hashtags* de apoio e rejeição de cada candidato para coletar os *tweets* rotulados. Não foi planejado um método adequado para representar opiniões neutras para um candidato.

Como trabalhos futuros, o primeiro passo a ser dado neste sentido é a coleta de diversos outros conjuntos de *tweets* mencionando os candidatos durante o período de campanha eleitoral, possibilitando obter maior precisão na evolução da aceitação de um candidato. Além disso, deve-se realizar o monitoramento do maior número de candidatos possíveis para representar mais fielmente o cenário eleitoral e elaborar uma estratégia para poder classificar *tweets* neutros sobre um candidato como realmente neutros.

## REFERÊNCIAS

ARANHA, C. N. **Uma Abordagem de PréProcessamento Automático para Mineração de Textos em Português: Sob o Enfoque da Inteligência Computacional**. Dissertation (Master) — Pontifícia Universidade Católica do Rio de Janeiro, 2007.

BENEVENUTO, F.; RIBEIRO, F.; ARAÚJO, M. **Métodos para Análise de Sentimentos em mídias sociais**. 2015. Disponível na Internet: <<http://db.csail.mit.edu/labdata/labdata.html>>.

CARRILHO JUNIOR, J. **DESENVOLVIMENTO DE UMA METODOLOGIA PARA MINERAÇÃO DE TEXTOS**. Dissertation (Master) — PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO, 2007. Disponível na Internet: <[http://www.maxwell.vrac.puc-rio.br/Busca\\_etds.php?strSecao=resultado&nrSeq=11675@1](http://www.maxwell.vrac.puc-rio.br/Busca_etds.php?strSecao=resultado&nrSeq=11675@1)>.

CHOWDHURY, G. G. Natural language processing. **Annual Review of Information Science and Technology**, Association for Information Science and Technology, v. 37, n. 1, p. 51–89, 2005. Disponível na Internet: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/aris.1440370103>>.

CORREA, R. F. **Processamento de Linguagem Natural**. 2013. Disponível na Internet: <<https://sites.google.com/site/renatocorrea/temas-de-interesse/processamento-de-linguagem-natural>>.

FELDMAN, R. Techniques and applications for sentiment analysis. **Commun. ACM**, ACM, New York, NY, USA, v. 56, n. 4, p. 82–89, abr. 2013. ISSN 0001-0782. Disponível na Internet: <<http://doi.acm.org/10.1145/2436256.2436274>>.

GOMES, H. J. **Text Mining: Análise de Sentimentos na classificação de notícias**. Dissertation (Master) — Universidade Nova de Lisboa, 2012.

HOTH, A.; NURNBERGER, A.; PAASS, G. A brief survey of text mining. **LDV Forum - GLDV Journal for Computational Linguistics and Language Technology**, v. 20, p. 19–62, 2005.

INDURKHIA, N.; DAMERAU, F. J. **Handbook of Natural Language Processing**. 2nd. ed. [S.l.]: Chapman & Hall/CRC, 2010. ISBN 1420085921, 9781420085921.

KOLCHYNA, O.; SOUZA, T. T. P.; TRELEAVEN, P. C.; ASTE, T. Twitter sentiment analysis. **CoRR**, abs/1507.00955, 2015.

LIU, B. **Sentiment Analysis and Opinion Mining**. [S.l.]: Morgan & Claypool Publishers, 2012. ISBN 1608458849, 9781608458844.

MOHAMMAD, S.; KIRITCHENKO, S.; ZHU, X. Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets. Em: **Second Joint Conference on Lexical and Computational Semantics (\*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)**. Association for Computational Linguistics, 2013. p. 321–327. Disponível na Internet: <<http://aclweb.org/anthology/S13-2053>>.

MOHAMMAD, S. M.; TURNEY, P. D. Crowdsourcing a word–emotion association lexicon. **Computational Intelligence**, National Research Council Canada, v. 29, n. 3, p. 436–465, 2013.

NARR, S.; HULFENHAUS, M.; ALBAYRAK, S. Language-independent twitter sentiment analysis. **Workshop on Knowledge Discovery, Data Mining and Machine Learning**, LWA, p. 12–14, 2012.

PANG, B.; LEE, L. Opinion mining and sentiment analysis. **Foundations and Trends in Information Retrieval**, Now, v. 2, p. 1–135, 2008.

READ, J. Using emoticons to reduce dependency in machine learning techniques for sentiment classification. **Proceedings of the ACL Student Research Workshop**, ACL, p. 43–48, 2005. Disponível na Internet: <<http://dl.acm.org/citation.cfm?id=1628960>. 1628969>.

REIS, J.; GONÇALVES, P.; OLMO, P.; PRATES, R.; BENEVENUTO. Magnet news: You choose the polarity of what you read. **Proceedings of the 8th International AAI Conference on Weblogs and Social Media (ICWSM)**, p. 652–653, 2014. Disponível na Internet: <<https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8105>>.

TSYTSARAU, M.; PALPANAS, T. Survey on mining subjective data on the web. **Data Mining and Knowledge Discovery**, Kluwer Academic Publishers, Hingham, MA, USA, v. 24, n. 3, p. 478–514, 2012. ISSN 1384-5810. Disponível na Internet: <<http://dx.doi.org/10.1007/s10618-011-0238-6>>.

TUMASJAN, A.; SPRENGER, T.; SANDNER, P.; WELPE, I. Predicting elections with twitter: What 140 characters reveal about political sentiment. **Proceedings of the 4th International AAI Conference on Weblogs and Social Media (ICWSM)**, 2010. Disponível na Internet: <<https://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/view/1441>>.

TUMITAN, D.; BECKER, K. Sentiment-based features for predicting election polls: A case study on the brazilian scenario. **2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)**, v. 2, p. 126–133, 2014.

WANG, H.; CAN, D.; KAZEMZADEH, A.; BAR, F.; NARAYANAN, S. A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. **50th Annual Meeting of the Association for Computational Linguistics**, ACL, p. 115–120, 2012. Disponível na Internet: <<http://aclweb.org/anthology/P12-3>>.

WITTEN, I. H.; FRANK, E. **Data Mining Practical Machine Learning Tools and Techniques**. [S.l.]: Morgan Kaufmann, 2005.