

RECUPERAÇÃO E CLASSIFICAÇÃO DE SENTIMENTOS DE USUÁRIOS DO TWITTER EM PERÍODO ELEITORAL

RECOVERY AND CLASSIFICATION OF TWITTER USER FEELINGS IN ELECTORAL PERIOD

Fernanda Fernandes Matos^a

Lúcia Helena de Magalhães^b

Renato Rocha Souza^c

RESUMO

Introdução: As redes sociais tornaram-se um espaço importante para usuários expressarem seus sentimentos. Esses comentários são valiosos para governantes saberem o ponto de vista dos cidadãos sobre as suas propostas políticas e candidatos perceberem a reação dos eleitores a respeito da campanha eleitoral. **Objetivos:** Analisar os sentimentos expressos pelos usuários no Twitter, referentes aos candidatos que concorreram à presidência do Brasil no ano de 2018, e prever o resultado das eleições com base nessas postagens. **Metodologia:** Os posts sobre os candidatos que disputaram o segundo turno das eleições foram o objeto de estudo. Usou-se o software Orange Canvas, uma ferramenta de aprendizado de máquina livre e de código aberto, para a coleta da amostra e para a extração das informações relevantes. A técnica de análise de opinião foi aplicada para classificação automática dos sentimentos em positivos, negativos e neutros. Para melhor análise e interpretação dos resultados, exibiram-se as palavras mais importantes dos comentários em nuvens de palavras e as emoções, em gráficos de distribuição de frequência. **Resultados:** Detectaram-se muitos sentimentos negativos nas postagens e a emoção de surpresa foi a que mais se destacou para ambos os concorrentes. **Conclusões:** O estudo mostrou que o Twitter é um local interessante para usuários expressarem seus sentimentos no período eleitoral. Porém, o trabalho não foi capaz de prever o resultado das eleições com base nas emoções. Acredita-se que isso se deve às altas taxas de rejeição dos eleitores quanto aos candidatos e a polarização que tem caracterizado a política brasileira nos últimos tempos.

Descritores: Análise de redes sociais. Classificação automática. Recuperação da Informação. Extração da Informação. Mineração de texto.

^a Mestre em Ciência da Informação pelo Programa de Gestão e Organização do Conhecimento na Universidade Federal de Minas Gerais. Email: matos.fernandafernandes@gmail.com

^b Doutoranda em Gestão e Organização do Conhecimento na Universidade Federal de Minas Gerais. Email: lucia.magalhaes@ifsudestemg.edu.br

^c Pós-doutorado em Tecnologias Semânticas para Recuperação de Informação. Email: rsouzaufmg@gmail.com

1 INTRODUÇÃO

Com a expansão da grande rede mundial de computadores e o crescimento da sua popularidade, um amplo volume de dados e informação tem sido gerado e publicado em inúmeras páginas da internet nas quais se encontram dados das mais diversas áreas do saber. Além disso, a web tornou-se um espaço para usuários expressar e compartilhar sentimentos em resenhas, blogs, fórum de discussão, noticiário *on-line* e redes sociais. Esses comentários são importantes para as empresas e organizações conhecerem a opinião pública sobre os seus produtos e serviços, governos locais e federais saberem o ponto de vista dos cidadãos sobre as suas propostas políticas e candidatos perceberem a reação dos possíveis eleitores a respeito da campanha eleitoral.

Com esse volume de informações disponível e as taxas diárias em crescimento, são indispensáveis técnicas eficientes para extração de conhecimentos oriundos da participação dos usuários nos conteúdos de páginas web e comunidades virtuais, pois a busca por informação relevante e em tempo real torna-se cada vez mais necessária. Além do mais, com a abundância de opiniões publicamente disponíveis nas redes, encontrar informações relevantes e extraí-las de forma resumida e organizada não é tarefa fácil. Mas saber se os planos de governo de um candidato foram elogiados ou criticados pelos eleitores e identificar palavras e frases de opinião (por exemplo, bom, ruim, fraco) são questões fundamentais para análise de sentimentos.

Segundo Sousa (2012), é possível extrair dados importantes nos meios socioculturais e políticos de uma região, bem como obter os comentários dos usuários sobre política e governo, cultura e outros aspectos de uma localidade. Através da análise de sentimentos, é possível também conseguir informações significativas relacionadas às campanhas eleitorais. Além disso, as opiniões são as principais influenciadoras de comportamentos. As crenças e percepções da realidade são amplamente condicionadas a como os outros enxergam o mundo. Porém, saber o que apenas uma pessoa pensa, frequentemente, não é suficiente para influenciar outras, por isso, a necessidade dos palpites de muitas. Logo, a necessidade de sintetizar. Para isso, ferramentas de tratamento de dados são

imprescindíveis para análise dos pareceres dos usuários da internet e para a visualização dos resultados.

Por isso, a importância de um estudo na área de recuperação, extração e classificação de opiniões, com o objetivo de desenvolver estratégias para extrair, automaticamente e de forma inteligente, características importantes contidas em postagem de usuários e armazená-las em alguma estrutura representativa ou semântica, facilitando, deste modo, a leitura e compreensão dessas informações.

Nesse contexto, este trabalho tem como objetivo analisar os sentimentos expressos pelos eleitores, na rede social Twitter, sobre os candidatos que disputaram o segundo turno das eleições para presidente do Brasil no ano de 2018. E, desse modo, emitir de forma prática e condensada as emoções dos usuários, além de detectar sentimentos positivos e negativos relacionados aos políticos Bolsonaro e Haddad. Acredita-se que esta pesquisa é de extrema relevância para analisar as opiniões das pessoas a respeito de concorrentes a cargos do governo, questões políticas, planos administrativos e assim por diante. Por conseguinte, nesta pesquisa, utilizou-se o software Orange Canvas, uma ferramenta de aprendizado de máquina livre e de código aberto, para coleta, pré-processamento dos *tweets*, análise e visualização dos resultados.

2 ANÁLISE DE SENTIMENTOS

Análise de Sentimento ou Mineração de Opinião é uma área da Mineração de Textos cujo objetivo não é categorizar os textos por tópicos, e sim classificá-los com base na emoção contido em determinado documento. Normalmente, a classificação é baseada em sentimentos positivos e negativos. Segundo Pang e Lee (2008), a análise, de uma forma mais abrangente, é usada para o tratamento computacional de opinião, sentimento e subjetividade em textos. Para esses autores, o objetivo é usar Processamento de Linguagem Natural (PNL), estatísticas ou métodos de Aprendizado de Máquina para extrair, identificar ou caracterizar esse tipo de comentários em uma unidade de texto.

Segundo Esuli (2006), Análise de Sentimento é um campo da linguística computacional em que a preocupação é descobrir a opinião que um documento

expressa sobre tal tópico ou produto ao qual se refere. Para Liu (2010), essa área é útil tanto para aplicações voltadas para empresas e organizações, quanto para indivíduos. Pang e Lee (2008) alegam que em outros países, principalmente nos Estados Unidos, já tem muitas empresas que se dedicam a oferecerem serviços de Análise de Sentimento. Os autores também destacam que o número de aplicações possíveis nessa área é grande e apresentam algumas, tais como: Aplicações para sites de resenhas; aplicações com tecnologia de base; aplicações ligadas ao mercado e ao governo e aplicações através de diferentes domínios, como, por exemplo, Sociologia, Administração ou Direito.

Serviços de monitoramento de redes sociais também estão disponíveis para análise de comportamentos e sentimentos. Para Oliveira (2014), o monitoramento de mídias sociais tem diversas finalidades, a depender do objetivo a ser alcançado. Entre os objetivos do monitoramento, estão a Análise de reações, sentimentos e desejos relativos a produtos, marcas, pessoas, campanhas e assuntos; SAC 2.0 que tem como finalidade identificar demandas para o atendimento, dúvidas, críticas ou sugestões; Análise de tendências e comportamentos de consumo do público-alvo; Compreensão de quais são os assuntos e o que motiva as conversações em torno do objeto; Identificação de crises/ameaças; Identificação de oportunidades de ações reativas e proativas (real-time marketing); Compreensão dos públicos que falam do objeto monitorado; Levantamento de influenciadores, detratores ou outros públicos; Análise de territórios de conteúdo explorados pelo objeto monitorado; análise de reputação e marketing político.

Com o objetivo de descobrir informações relevantes a partir das interações dos usuários nas redes sociais, ferramentas como Web Analytics, como Facebook Insights e Google Analytics, que são focadas em apresentar dados quantitativos e métricas de canais das mídias sociais e proprietários das marcas, dominaram a categoria de ferramentas mais utilizadas pelos profissionais de inteligência em 2013 (ZANDAVALLE, 2016). Já o Twitter Analytics, ferramenta de análise de atividades na rede social, teve sua primeira aparição no estudo de Junior Siri (2014).

O IBM Watson¹ também oferece um módulo de análise de sentimentos dos serviços cognitivos da Microsoft². De acordo com informações contidas nos sites da Microsoft Azure, a API (*Application Programming Interface*) de Análise de Texto é um serviço baseado em nuvem que fornece um processamento de idioma natural avançado sobre texto bruto e inclui quatro funções principais: análise de sentimento, extração de frases-chave, detecção de idioma e reconhecimento de entidade. A ferramenta de análise de sentimento possui uma API que é utilizada para automatizar o processo de classificação de um texto. A API avalia cada texto de entrada fornecido e retorna uma pontuação para cada texto que varia de 0 (muito negativo) até 1 (muito positivo) e pontuações próximas de 0,5 são consideradas neutras. (MICROSOFT, 2019).

Com base no exposto, observa-se que as grandes empresas do mercado também estão desenvolvendo técnicas com o objetivo de facilitar a descoberta de conhecimento a partir das iterações dos usuários nas redes sociais, porém nesta pesquisa optou-se por usar uma abordagem baseada em aprendizado de máquina, que pode combinar algoritmos de classificação, conforme proposto por Silva (2016).

Segundo Foschiera (2012, p. 14), “o tratamento computacional da opinião vem representando um desafio na atualidade, quando temos à disposição um grande número de informações relevantes que nos permite entender o ponto de vista dos outros”. Kim e Hovy (2006) alegam que, em decorrência disso, a identificação automática dos sentimentos de documentos eletrônicos tem se tornado o centro da atenção de diferentes campos de pesquisas.

A Análise de Sentimento é uma área que estuda a extração de opiniões usando Recuperação da Informação (RI), Inteligência Artificial (IA) e técnicas de Processamento de Linguagem Natural (PLN). Martha (2005) afirma que a RI é uma ciência que estuda a criação de algoritmos para recuperar dados oriundos de textos livres, que constituem a maior parte de documentos em forma digital, disponível nos dias atuais, principalmente após o advento da internet. Porém, se todos os arquivos na web fossem claramente marcados com palavras-chave e

¹ <https://www.ibm.com/watson/services/natural-language-understanding/>

² <https://azure.microsoft.com/pt-br/services/cognitive-services/text-analytics/>

outros metadados que descrevessem perfeitamente o seu conteúdo, e se os usuários fossem treinados para fazer pesquisas, a recuperação de informações relevantes não exigiria algoritmos sofisticados, elas poderiam ser encontradas através de simples consultas (LINOFF; BERRY, 2002, p.43).

Para Liu (2007, p. 183) RI, teoricamente, estuda a aquisição, organização, armazenamento, recuperação e distribuição de informações. Magalhães (2008) alega que, com os documentos recuperados, a tarefa seguinte é transformar esses dados de forma que algoritmos de mineração e aprendizagem de máquina possam ser aplicados de forma efetiva para a obtenção ou extração de elementos pertencentes a determinados documentos.

Segundo Cordeiro (2003), enquanto a RI pretende extrair documentos importantes, entre um universo de possibilidades, a Extração da Informação (EI) pretende identificar elementos relevantes no interior de determinados documentos que contêm a informação de interesse. Dessa forma, as informações relevantes extraídas são armazenadas em alguma estrutura previamente definida, como, por exemplo, numa tabela de uma base de dados. Pal, Talwar e Mitra (2002) afirmam que é importante salientar a diferença entre as fases de recuperação e extração de informação. De acordo com esse autor, as técnicas de EI procuram retirar conhecimento de documentos recuperados conforme a forma em que o documento está estruturado e representado, enquanto as técnicas de RI visualizam o documento apenas como um conjunto de termos. Segundo Lopes *et al.* (2008), esses métodos “são baseados em estatística, trabalham com informação previamente anotada, com o mesmo intuito: extrair a orientação semântica do texto: positivo, negativo ou neutro”.

De acordo com Konchady (2006), os sistemas de extração geralmente convertem textos não estruturados em um formato que possa ser enviado para uma tabela de um banco de dados ou para uma planilha do Excel. Informações úteis, como opiniões sobre um candidato, podem ser extraídas das redes sociais, sem a necessidade de ler vários comentários. Métodos simples de reconhecimento de padrão ou algoritmo de aprendizado de máquina são suficientes para a extração.

Para Denecke (2008), a classificação de sentimentos pode ser

desenvolvida, sobretudo, por meio de duas técnicas, ou seja, por Aprendizagem de Máquina (estatística) e por seleção de recursos lexicais e Processamento de Linguagem Natural. Nos métodos baseados no modelo estatístico, os termos são extraídos a partir de uma análise de frequência das palavras ou frases em cada documento e em todo o conjunto de informação. Nas técnicas linguísticas, os termos de indexação são extraídos utilizando técnicas de PLN, como, por exemplo, análise lexical, morfológica, sintática e semântica.

Segundo Pak e Paroubek (2010), as redes sociais são fontes ricas de informação que podem ser utilizadas de forma eficiente para a comercialização ou para o desenvolvimento de pesquisas. Esse caráter eficiente é que fundamenta a importância da geração de sistemas capazes de analisar automaticamente os sentimentos dos usuários, ou seja, o desenvolvimento de aplicações computacionais eficientes de Mineração de Opinião. Pensando nisso, Hu e Liu (2004) desenvolveram um estudo em que os autores apresentam um sistema de sumarização automática de opiniões relacionadas a produtos. Nesse trabalho, desenvolveu-se um mecanismo de mineração de características dos objetos, bem como a identificação das sentenças cujo sentimento é positivo ou negativo.

Segundo O'Connor *et al.* (2010), a opinião expressa pelos usuários em redes sociais tem sido monitorada também por governos e empresas na busca pela compreensão do que uma população pensa sobre um determinado produto ou serviço.

Liu (2015), em seu livro *Sentiment Analysis Mining Opinions, Sentiments, and Emotions*, apresenta uma introdução abrangente para o tema a partir de um ponto de vista, principalmente, sobre o PLN, com o objetivo de ajudar os leitores a entender a estrutura subjacente do problema e a construção de linguagem que são comumente usados para expressar opiniões. Além disso, abrange todas as áreas centrais de análise de sentimento, incluindo muitos temas emergentes, como a análise de debates e comentários, mineração de intenção, bem como a detecção de falsos sentimentos, e, ainda, apresenta métodos computacionais para analisar e sintetizar esses comentários.

Para Liu (2015), a análise de sentimento não é uma tarefa simples, mas

um problema multifacetado que contém muitos subproblemas. Assim, o grande desafio é, então, desenvolver métodos para “ensinar” programas de computador a identificar as emoções relativas a cada entidade em textos opinativos.

Muitas outras abordagens desse campo ainda podem ser encontradas na literatura, devido às suas potenciais aplicações. Igualmente importantes são os novos desafios intelectuais que apresenta essa área para a comunidade de pesquisa. Desta forma, se torna imprescindível novos estudos para extrair opiniões de documentos não estruturados e criar versões resumidas que possam melhorar, por exemplo, a percepção do usuário sobre um determinado assunto, sem a necessidade da leitura de centenas de comentários postados na web.

Sabe-se que, uma parte importante do comportamento das pessoas em relação à coleta de informações tem sido, na maioria das vezes, para descobrir o que os cidadãos pensam sobre um determinado assunto, serviço ou produto. Para Pang e Lee (2008), é importante saber o que todos pensam quando se deseja tomar uma decisão. De acordo com Bonette (2011), o uso de dados disponíveis na web, alinhados com ferramentas de extração e descoberta de conhecimento, pode auxiliar as organizações em ambientes competitivos.

Segundo Liu (2015), a técnica de Análise de sentimentos não é desafiadora apenas por causa do processamento da linguagem natural, mas também pela sua utilidade prática, pois empresas querem descobrir a concepção de seu público sobre seus produtos e serviços, clientes potenciais também querem saber as opiniões dos consumidores antes de comprarem um produto ou usarem um serviço e candidatos desejam perceber as reações dos eleitores a respeito de suas propostas de governo. Porém, apesar do acesso às informações relevantes de uma organização ser imprescindível para que os dirigentes possam decidir a melhor maneira de direcionar seus projetos e fazer com que ela atinja seus objetivos mercadológicos, obtê-las normalmente não é uma tarefa fácil (MAGALHAES, 2009).

Assim, o surgimento súbito de atividade na área de Análise de Sentimento, que lida com o tratamento computacional e subjetividade no texto, tem contribuído com o crescente interesse por novos métodos que procuram abordar os novos desafios para análises de pareceres, opiniões, avaliações,

recomendações, estimativas, emoções, subjetividade etc, disponibilizados em formato textual, tais como resenhas, notícias, discussões ou comentários postados na web. Nesse contexto, as redes sociais têm se tornado um interessante objeto de estudos, principalmente, após o crescente uso de smartphones e outros dispositivos móveis conectados à internet.

3 METODOLOGIA

Para a análise dos sentimentos dos usuários do Twitter, usou-se o software Orange Canvas³ e para o experimento, primeiramente, recuperaram-se as informações necessárias para o estudo. Desse modo, o primeiro passo foi adquirir uma coleção de comentários no Twitter sobre os candidatos à presidência que concorreram o segundo turno das eleições no ano de 2018. Para isso, coletaram-se, para cada teste, 1000 *tweets* referentes a cada candidato.

Como os *posts* recuperados, o passo seguinte foi o pré-processamento. Para Silva (2010), isso inclui todas as rotinas, processos e métodos necessários para a preparação dos dados. Tem a função de reduzir o vocabulário e torná-lo menos disperso, característica essencial para o processamento computacional. Apesar de ser um trabalho exaustivo, tem influência direta na qualidade do resultado da análise. Assim, nesta etapa os dados foram transformados, ou seja, converteram-se os sentimentos coletados no Twitter em minúsculo, removeram-se urls e acentos, além de usar um analisador HTML (*Parse HTML*) para detectar *tags* HTML e considerar apenas o texto dessas marcações.

Em seguida, dividiu-se a coleção de sentimentos em unidades menores, denominada *tokens*. Sarkar (2016) os definem como componentes textuais independentes e mínimos que possuem alguma sintaxe e semântica definitivas. Um parágrafo do texto ou um documento tem vários componentes que podem ser separados em cláusulas, frases e palavras. Sendo assim, para dividir um texto em sentenças e cada frase em palavra, é utilizada uma técnica conhecida como tokenização. Esse procedimento foi necessário neste trabalho para a realização das operações de pré-processamento.

³ Disponível para download em: <https://orange.biolab.si/>

Ademais, as palavras foram reduzidas ao seu radical, método conhecido como *stemming*. Segundo Morais e Ambrósio (2007), *stemming* é uma técnica de normalização linguística, na qual as variantes de um termo são reduzidas a uma forma comum denominada *stem* (radical). Isso resulta na eliminação de prefixos, sufixos e características de gênero, número e grau das palavras, reduzindo o número de atributos em até 50%.

Seguidamente, removeram-se as *stopwords*, que são palavras comuns, como artigos, conjunções e preposições, que aparecem nos comentários, repetem muitas vezes, mas não acrescentam e nem retiram informações relevantes. Desse modo, após análise dos dados, elaborou-se uma lista de *stopwords* e a acrescentou ao conjunto já disponibilizado pelo Orange. Ademais, usaram-se expressões regulares para remover os caracteres especiais. E, para trabalhar apenas com os termos mais relevantes do conjunto, utilizou-se a *Document Frequency* (DF). A DF procura eliminar palavras muito frequentes por possuir pouco poder de discriminação e também as palavras que aparecem poucas vezes, por não contribuir com a análise dos sentimentos.

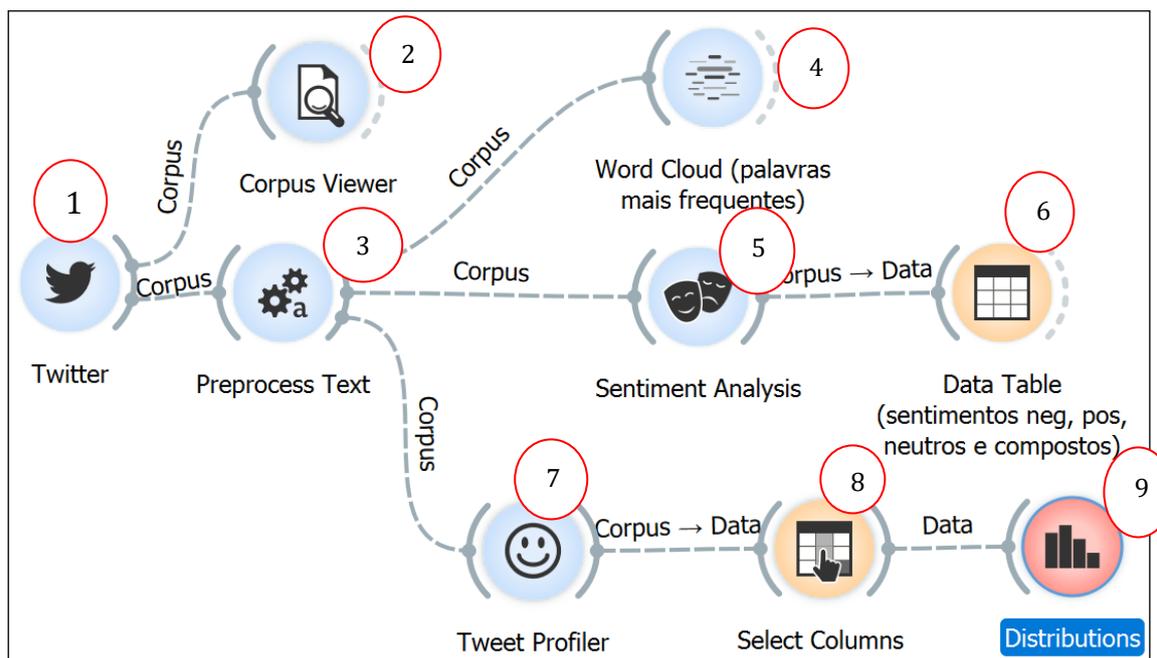
Com os dados preparados, a próxima fase é a análise das opiniões. O Orange utiliza os módulos de sentimento *Liu Hu* (HU; LIU, 2004) e *Vader* (HUTTO; GILBERT, 2014) da NLTK (*Natural Language Toolkit*) para o diagnóstico. O módulo *Liu Hu* é baseado em léxico e suporta somente textos em inglês e esloveno. Desta forma, nesta pesquisa, utilizou-se o *Vader* (*Valence Aware Dictionary for sEntiment Reasoning*), que é baseado em léxico e regra. Esse componente computa os sentimentos com pontuação positiva, negativa e neutra. Assim, os sentimentos são classificados por pontuação, onde -1 é o sentimento mais negativo e 1, o sentimento mais positivo.

Além disso, realizou-se a análise das opiniões dos usuários usando o módulo *Tweet Profiler*. As emoções referentes à *Anger* (raiva), *Disgust* (descontentamento), *Fear* (medo), *Joy* (alegria), *Sadness* (tristeza) e *Surprise* (surpresa) foram exibidas em um gráfico de distribuição de frequência. Ademais, criou-se uma nuvem de palavras para a visualização dos termos mais frequentes da coleção de *tweets* de cada candidato e os resultados referentes a sentimentos positivos, negativos e neutros foram exibidos em uma tabela. Geraram-se,

também, gráficos de distribuição de frequências para melhor análise das emoções dos usuários (raiva, descontentamento, medo, alegria, tristeza, surpresa).

Por fim, a figura 1 ilustra todos os passos dessa metodologia: Em síntese, coletaram-se os *tweets*; visualizaram-se o *corpus*; fez-se o pré-processamento; exibiram-se as palavras mais frequentes em nuvem de palavras; identificaram-se os sentimentos negativos, positivos ou neutros presentes nas postagens; apresentou-se o resultado da análise em uma tabela; usou-se o módulo *Tweet Profiler* para identificar as emoções expressas nos *tweets*. Ademais, selecionaram-se algumas colunas, pois a coleção de *tweets* continha informações referentes ao autor, conteúdo da postagem, idioma, data etc. Porém, para a análise das emoções, utilizou-se somente a coluna de conteúdo, ou seja, os comentários dos usuários. Por fim, as emoções foram exibidas em um gráfico de distribuição de frequência.

Figura 1 - Desenho do experimento



Fonte: Elaborada pelos autores

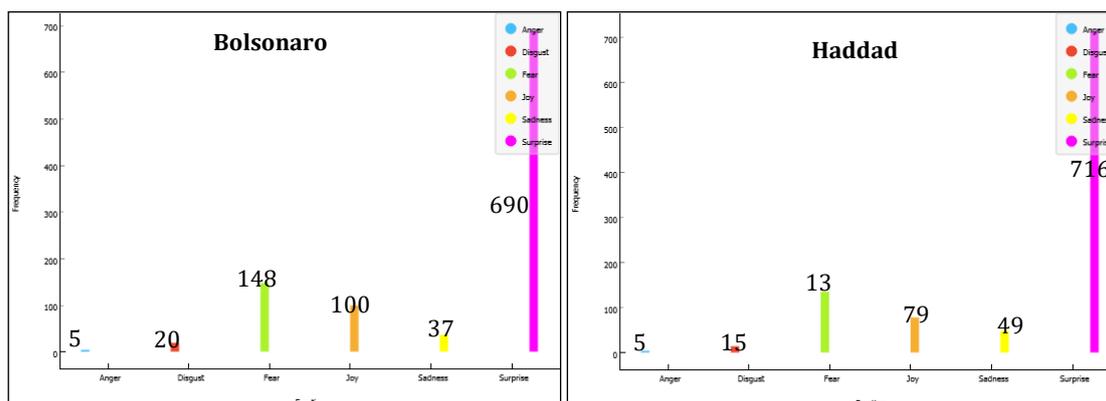
4 ANÁLISE DOS RESULTADOS

A primeira análise realizou-se logo após o debate do dia 09 de agosto de

Nota-se na figura 3 que a palavra “Lula” apareceu com pouco peso nos comentários dos dois candidatos, provavelmente isso se deve ao fato do candidato do Partido dos Trabalhadores ter se desprendido do discurso “Lula é Haddad e Haddad é Lula” e redirecionado a campanha com menos foco no ex-presidente. É notória também a forte presença das palavras “Joaquim” e “Barbosa” na nuvem do candidato Haddad. Isso se deve ao fato do jurista ter declarado voto ao candidato um dia antes da eleição, o que gerou uma série de compartilhamentos e comentários nas redes sociais.

Além da nuvem de palavras, geraram-se gráficos de distribuição de frequência para ilustrar as emoções dos internautas em relação aos políticos. A figura 4 apresenta os sentimentos expressos pelos usuários em um corpus de 1000 *tweets*, coletado, para cada candidato, no dia 11 de outubro de 2018.

Figura 4 - Gráfico de distribuição de frequência de emoções



Fonte: Elaborado pelos autores

Percebe-se na figura 4 que as emoções referentes aos dois candidatos são semelhantes e que muitos sentimentos foram classificados pelo Orange como de surpresa (*Surprise*), o que não demonstra se os internautas apresentaram reações positivas ou negativas em relação aos comentários, visto que a reação de muitas pessoas foi de surpresa, gostando ou não de alguma postagem. Outro sentimento que se destacou no gráfico foi o sentimento de medo (*Fear*) e com um pouco menos de intensidade, o sentimento de alegria (*Joy*).

Assim, para melhor analisar as emoções dos usuários do *Twitter* no período eleitoral, coletaram-se tweets em datas diferentes. A tabela 1 mostra a reação (*anger, disgust, fear, joy, sadness* e *surprise*) dos usuários referentes ao candidato Bolsonaro, em cinco datas.

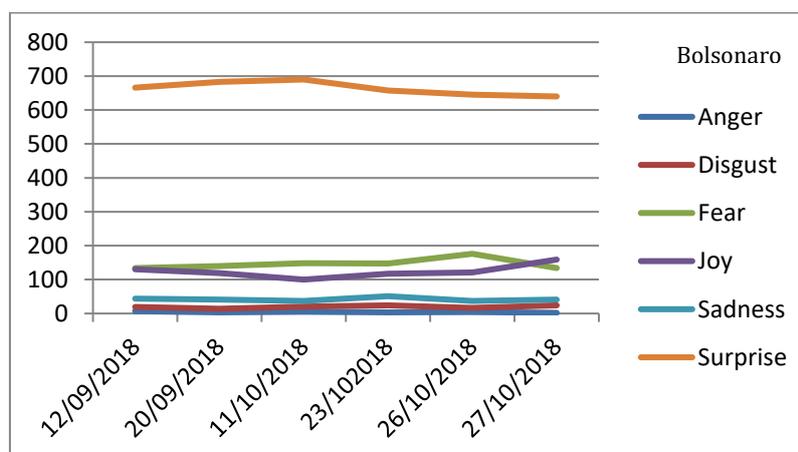
Tabela 1 - Emoções dos usuários em relação ao candidato Bolsonaro

BOLSONARO	12/09/2018	20/09/2018	11/10/2018	23/10/2018	26/10/2018	27/10/2018
Anger	7	3	5	3	4	2
Disgust	19	14	20	24	17	24
Fear	134	140	148	147	176	134
Joy	130	119	100	117	121	159
Sadness	44	41	37	51	37	41
Surprise	666	683	690	658	645	640

Fonte: Elaborada pelos autores

Para facilitar o entendimento, as informações da tabela 1 foram exibidas em um gráfico de linhas (gráfico 1).

Gráfico 1 - Emoções dos usuários em relação ao candidato Bolsonaro em diferentes datas



Fonte: Elaborado pelos autores

Verifica-se no gráfico 1 que, a partir do dia 26 de outubro de 2018, houve uma queda do sentimento de medo referente ao candidato Bolsonaro e o sentimento de alegria começou a aumentar. Quanto ao candidato Haddad, a tabela 3 ilustra, de forma quantitativa, as emoções dos usuários recuperadas nas mesmas datas em que se coletaram os *tweets* sobre o candidato Bolsonaro.

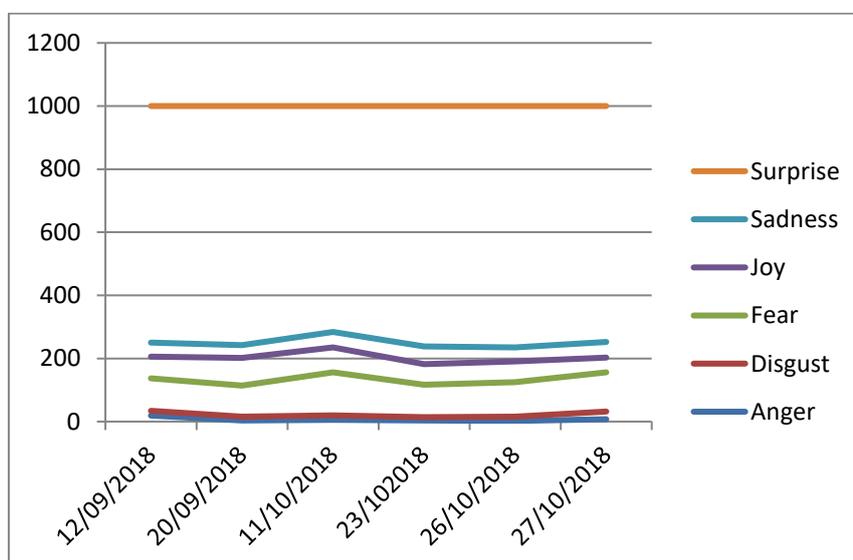
Tabela 2 - Emoções dos usuários em relação ao candidato Haddad

Haddad	12/09/2018	20/09/2018	11/10/2018	23/10/2018	26/10/2018	27/10/2018
Anger	19	3	5	3	2	7
Disgust	15	13	15	11	14	25
Fear	103	98	136	103	109	124
Joy	69	88	79	65	66	47
Sadness	44	40	49	56	44	49
Surprise	750	758	716	762	765	748

Fonte: Elaborada pelos autores

Para facilitar a compreensão, os dados da tabela 2 foram exibidos em um gráfico de linhas (gráfico 2).

Gráfico 2 - Emoções dos usuários em relação ao candidato Haddad em diferentes datas



Fonte: Elaborado pelos autores

Nota-se que, a partir do dia 23 de outubro de 2018, o sentimento de alegria ficou praticamente constante e as emoções referentes à tristeza, ao medo e ao desgosto começaram a aumentar. Porém, como as variações relacionadas aos sentimentos foram pequenas, não foi possível identificar uma conexão entre essas emoções e o resultado das eleições. Por conseguinte, a análise realizada usando o *Tweet Profiler* não foi capaz de identificar qual candidato tinha a preferência dos internautas e, assim, prever os resultados das eleições. Acredita-se que isso se deve ao fato de ser uma eleição muito polarizada, em que os dois candidatos tinham muitos eleitores que os rejeitam e ao mesmo tempo ambos tinham seguidores que os idolatravam.

Dando continuidade à pesquisa, a tabela 3 ilustra a classificação dos sentimentos expressos nos *tweets* como negativos, positivos e neutros.

Tabela 3 - Classificação dos sentimentos (*tweets* coletados no dia 12/09/18)

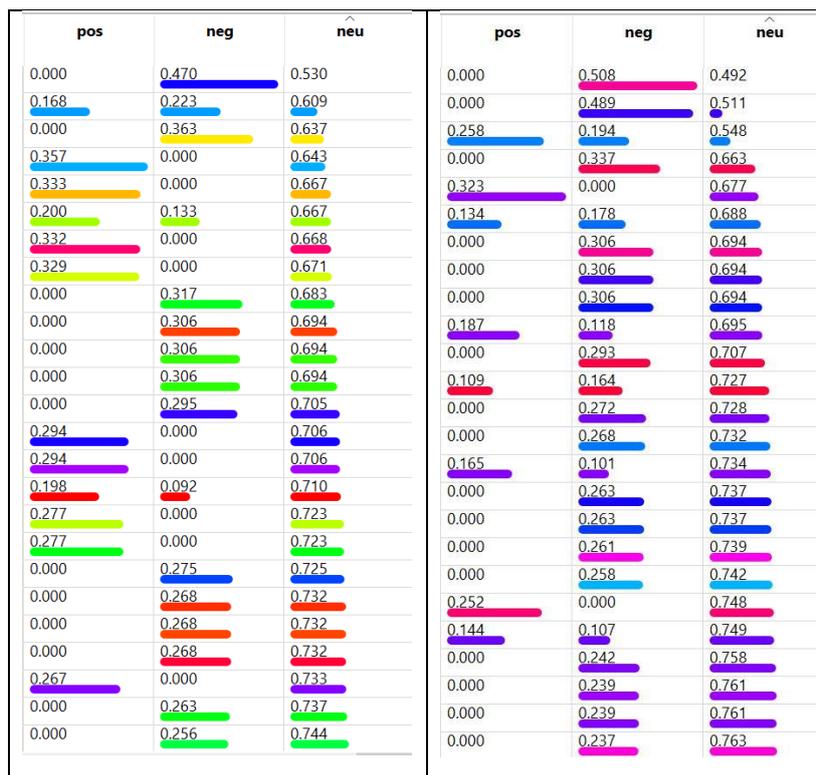
Bolsonaro			Haddad		
pos	neg	neu	pos	neg	neu
0.000	0.454	0.546	0.333	0.204	0.463
0.000	0.454	0.546	0.254	0.169	0.577
0.000	0.370	0.630	0.254	0.169	0.577
0.200	0.157	0.643	0.247	0.136	0.617
0.000	0.355	0.645	0.000	0.355	0.645
0.000	0.355	0.645	0.252	0.093	0.654
0.000	0.338	0.662	0.340	0.000	0.660
0.000	0.320	0.680	0.000	0.328	0.672
0.000	0.306	0.694	0.324	0.000	0.676
0.000	0.306	0.694	0.000	0.311	0.689
0.000	0.293	0.707	0.308	0.000	0.692
0.000	0.292	0.708	0.308	0.000	0.692
0.189	0.104	0.708	0.198	0.109	0.693
0.000	0.286	0.714	0.000	0.306	0.694
0.085	0.188	0.726	0.000	0.293	0.707
0.000	0.268	0.732	0.185	0.104	0.711
0.000	0.268	0.732	0.000	0.289	0.711
0.000	0.268	0.732	0.000	0.286	0.714
0.000	0.263	0.737	0.000	0.284	0.716
0.000	0.258	0.742	0.079	0.201	0.720
0.252	0.000	0.748	0.000	0.280	0.720
0.000	0.251	0.749	0.274	0.000	0.726
0.159	0.087	0.754	0.000	0.268	0.732
0.245	0.000	0.755	0.000	0.268	0.732
0.000	0.244	0.756	0.000	0.268	0.732

Fonte: Elaborada pelos autores

Ao examinar as informações da tabela 3, nota-se que muitos dos usuários do twitter demonstraram neutralidade em relação aos dois candidatos. Porém, no período dessa análise, o candidato do Partido dos Trabalhadores apareceu com mais *tweets* expressando sentimentos positivos do que o candidato Bolsonaro. Acredita-se que a ocorrência foi devido ao período de análise, pois ainda faltava mais de um mês para a eleição. Dessa forma, para uma análise mais próxima da data da votação, coletaram-se *tweets* no dia 26 de outubro de 2018, dois dias antes do segundo turno. O resultado da análise dos sentimentos é apresentado na tabela 4.

Tabela 4 - Classificação dos sentimentos (*tweets* coletados no dia 26/10/18)

Bolsonaro	Haddad
-----------	--------



Fonte: Elaborada pelos autores

Observa-se na tabela 4 que os *tweets* foram classificados como negativo, positivo e neutro. Com base nessas informações, uma previsão do resultado das eleições poderia ser feita comparando o número de *tweets* que apresentam sentimentos positivos em relação a cada candidato. Porém, a classificação foi bastante semelhante para os dois políticos, ambos com muitos *tweets* negativos, o que corresponde ao senso comum esperado, pois se justifica devido à descrença dos usuários do Twitter em relação à política brasileira.

5 CONCLUSÃO

A análise de sentimentos é bastante desafiadora, pois os comentários dos usuários das redes sociais são apenas textos em formato livre que, tradicionalmente, os computadores têm dificuldade em interpretar, por isso a necessidade da etapa de pré-processamento para preparar os textos em formatos que os algoritmos de aprendizagem de máquina possam decifrar. Desta forma, nesse estudo, realizou-se, primeiramente, uma preparação dos *tweets* e, posteriormente, os comentários dos usuários foram classificados em

positivos, negativos ou neutros. Além disso, apresentaram-se nuvens de palavras com os termos que foram mais citados no Twitter, no período eleitoral, que tinha relação com os candidatos Bolsonaro e Haddad.

Apesar de ter usado neste experimento somente os módulos disponibilizados pelo Orange, *scripts* desenvolvidos na linguagem de programação Python também podem ser incorporados para estender novas funcionalidades, uma tarefa interessante a ser implementada em trabalhos futuros. Além disso, a metodologia usada neste trabalho também pode ser utilizada por organizações para mapear a opinião dos usuários do Twitter ou de outras redes sociais, fazendo o uso dos resultados para os mais diversos fins.

Contudo, o fato da análise dos dados provenientes de rede social Twitter, sobre as eleições para presidente do Brasil no ano de 2018, não ter sido útil para predição eleitoral, foi útil em mostrar a variação nos sentimentos dos usuários ao longo dos períodos analisados, para ambos os candidatos, dando oportunidade para que o político altere sua estratégia de campanha a favor da sua aceitação. E isso não significa que a análise de sentimentos não seja relevante para outras finalidades ou para a análise de pleito em outro momento. Além disso, o método é genérico e pode ser facilmente adotados e estendidos a outros domínios. Ademais, outros estudiosos disseram ter obtido sucesso ao predizer resultado das eleições nos Estados Unidos (SALUNKHE; SURNAR; SONAWARE, 2017) e em alguns outros países, como Taiwan (XIE *et al.*, 2018). Porém, como apenas esses estudos ainda não são suficientes para sugerir um algoritmo preditivo de sucesso, novas pesquisas ainda são necessárias. Desse modo, espera-se que novos trabalhos se dediquem a pesquisar os conteúdos eleitorais gerados em redes sociais, a fim de se ter umas visões gerais de como os eleitores reagem nesse espaço.

Por fim, é importante chamar a atenção para as inovações que as redes sociais trazem para a ação política na sociedade da informação, fato que promove novos temas de pesquisa, pois ainda há poucos estudos que examina a complexa relação entre campanhas políticas e redes sociais. É necessário realizar mais pesquisas para compreender de forma mais abrangente as opiniões dos cidadãos em relação às propostas dos políticos bem como outros

assuntos governamentais. Além disso, seria interessante fazer em trabalhos futuros um estudo que possa comparar os resultados da análise de sentimentos utilizando as ferramentas disponíveis no mercado com a metodologia de aprendizado de máquina usada nesta pesquisa.

REFERÊNCIAS

- BONETTE, R. P. **Análise de ferramentas de opinion mining aplicadas a redes sociais com foco em inovação de produtos**. 2011. 87 p. Monografia (Graduação em Sistemas de Informação) - Universidade Federal de Lavras, Lavras, 2011.
- CORDEIRO, J. P. C. **Extracção de Elementos Relevantes em Texto/Páginas da World Wide Web**. 2003. Dissertação (Mestrado em Inteligência Artificial e Computação) Faculdade de Ciências da Universidade do Porto, 2003.
- DENECKE, K. Using sentiwordnet for multilingual sentiment analysis. *In*: INTERNATIONAL CONFERENCE ON DATA ENGINEERING WORKSHOP, 24., 2008, **Anais** [...]. pp. 507–512.
- ESULI, A. **Opinion Mining**. Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche, Pisa, Italy, 14 jun. 2006.
- FOSCHIERA, S. M. P. **A semântica da emoção**: um estudo contrastivo a partir da FrameNet e da roda das emoções. Tese (Doutorado em Linguística Aplicada). São Leopoldo: Universidade do Vale do Rio dos Sinos – UNISINOS. Disponível em: <http://www.repositorio.jesuita.org.br/handle/UNISINOS/4220>. Acesso em: 14 nov. 2018.
- HU, M; LIU, B. Mining opinion features in customer reviews. *In*: AAI Conference on Artificial Intelligence, 4., 2004. **Anais** [...]. [s.l.]. pp. 755–760. 2004. Disponível em: <https://www.aaai.org/Papers/AAAI/2004/AAAI04-119.pdf>. Acesso em: 30 mar. 2020.
- HUTTO, C.J.; GILBERT, E. Vader: A parsimonious rule-based model for sentiment analysis of social media text. *In*: INTERNATIONAL AAI CONFERENCE ON WEBLOGS AND SOCIAL MEDIA, 8., **Anais** [...]. 2014. Disponível em: <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/download/8109/8122>. Acesso em: 14 nov. 2018
- KIM, S.; HOVY, E. Extracting opinions, opinion holders, and topics expressed in online news media text. *In*: WORKSHOP ON SENTIMENT AND SUBJECTIVITY IN TEXT. **Proceedings** [...]. P. 1–8, Sydney, July 2006. Disponível em: <https://dl.acm.org/citation.cfm?id=1654642>. Acesso em: 14 nov. 2018.

KONCHADY, M. **Text Mining Application Programming**. Charles River Media, Boston, 2006.

LINOFF, G.S.; BERRY, M. J. A. **Mining the Web**. Transforming Customer Data into Customer Value. New York: Wiley, 2002.

LIU, B. **Sentiment Analysis and Subjectivity**. In: INDURKHYA, N.; DAMERAU, F. J. (Ed.). [s.l.]: Handbook of Natural Language Processing. 2 ed. 2010. Disponível em: <https://www.cs.uic.edu/~liub/FBS/NLP-handbook-sentiment-analysis.pdf>. Acesso em: 30 mar. 2020.

LIU, B. **Sentiment Analysis: Mining Opinions, Sentiments, and Emotions**. Cambridge University Press, 2015.

LIU, B. Sentiment Analysis: A Multi-Faceted Problem. In: **EEE Intelligent Systems**, 2010b. Disponível em: <http://www.cs.uic.edu/~liub/FBS/IEEE-Intell-Sentiment-Analysis.pdf>. Acesso em set. 2015.

LIU, B. **Web Data Mining**. Exploring Hiperlinks, Contents, and Usage Data. Springer, Chigago, 2007.

LOPES, T. J. P.; HIRATANI, G. K. L.; BARTH, F. J. ; RODRIGUES, O. ; PINTO, J. M. Mineração de opiniões aplicada à análise de investimentos. In: SIMPÓSIO BRASILEIRO DE SISTEMAS MULTIMÍDIA E WEB - WEBMEDIA, 14., **Anais** [...], 2008. p. 117-120.

MAGALHÃES, T. M. **Uma metodologia de mineração de opiniões na Web**. 2009. 102f. Tese (Doutorado em Engenharia Civil) – Universidade Federal do Rio de Janeiro. Rio de Janeiro, 2009.

MARTHA, A. S. Recuperação de informação em campos de texto livre de prontuários eletrônicos do paciente baseada em semelhança semântica e ortográfica. 2005. Dissertação (Mestre em Ciências) Universidade Federal de São Paulo, São Paulo, 2005.

MICROSOFT. **Como fazer**: Detectar o sentimento usando a API de Análise de Texto. 2020. Disponível em: <https://docs.microsoft.com/pt-br/azure/cognitive-services/text-analytics/how-tos/text-analytics-how-to-sentiment-analysis?tabs=version-3>. Acesso em: 30 mar. 2020.

MORAIS, E. A. M.; AMBRÓSIO, A. P. L. **Mineração de Textos**. Universidade Federal de Goiás, 2007. (Relatório Técnico).

OLIVEIRA, Mariana. **Curso de Monitoramento de Mídias Sociais**. Trespontos: Lab Social Media, 2014.

O'CONNOR, B.; BALASUBRAMANYAN, R.; ROUTLEDGE, B. R.; SMITH, N. A.; From tweets to polls: linking text sentiment to public opinion time series. In: **INTERNATIONAL CONFERENCE ON WEBLOGS AND SOCIAL MEDIA**, 4. 2010. Anais [...], Washington, DC, EUA, 2010. Disponível em:

https://www.researchgate.net/publication/221297841_From_Tweets_to_Polls_Linking_Text_Sentiment_to_Public_Opinion_Time_Series. Acesso em 28 fev. 2020.

PAK, A.; PAROUBEK, P. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. *In: INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION*, 2010. **Anais [...]**. Université de Paris-Sud, Laboratoire LIMSI-CNRS, 2010.

PAL, S. K.; TALWAR, V.; MITRA, P. Web Mining in Soft Computing Framework: Relevance, State of the Art and Future Directions. **IEEE Transactions on Neural Networks**, v. 13, n. 5, p. 1163-1177, sep. 2002

PANG, B.; LEE, L. **Opinion Mining and sentiment analysis**. Foundations and Trends in Information Retrieval, v. 2, n. 1-2, 2008, p. 1-135.

SALUNKHE, P; SURNAR, A; SONAWANE, S. A Review: Prediction of Election Using Twitter Sentiment Analysis. **International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)**, v. 6, n. 5, maio 2017. Disponível em: <http://ijarcet.org/wp-content/uploads/IJARCET-VOL-6-ISSUE-5-723-725.pdf>. Acesso em nov. 2018.

SARKAR, D. **Text Analytics with Python: A Practical Real-World Approach to Gaining Actionable Insights from your Data**. [s.l.]: Apress. 2016.

SILVA, E. R. C. C. **Técnicas de Data e Text Mining para anotação de um arquivo digital**. 2010. 96 f. Dissertação (Mestrado em Engenharia Eletrônica e Telecomunicações) – Departamento de Eletrônica, telecomunicações e Informática, Universidade de Aveiro, 2010.

SILVA, N. F. F. d. Análise de sentimentos em textos curtos provenientes de redes sociais. 2016. 112 f. Tese (Doutorado em Ciências de Computação e Matemática Computacional) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2016.

JUNIOR SIRI. **Profissional de métricas, monitoramento social analytics no Brasil**: 2014. Disponível em: <https://pt.slideshare.net/juniorsiri/profissional-metricas-monitoramentosocialanalytics2014>. Acesso em: 28 fev. 2020.

SOUSA, G. L. S. **Tweetmining**: análise de opinião contida em textos extraídos do twitter. Monografia (Bacharel em Sistemas de Informação). Lavras: Universidade Federal de Lavras, 2012. Disponível em: <http://www.bsi.ufla.br/wp-content/uploads/2013/09/TWEETMINING-AN%C3%81LISE-DE-OPINI%C3%83O-CONTIDA-EM-TEXTOS-EXTRA%C3%8DDOS-DO-TWITTER-.pdf>. Acesso em: 27 mar. 2020.

XIE, Z.; LIU, G.; WU, J.; TAN, Y. Big data would not lie: prediction of the 2016 Taiwan election via online heterogeneous information. **EPJ Data Science**, v. 7, n. 32, 2018. Disponível em:

<https://epjdatascience.springeropen.com/track/pdf/10.1140/epjds/s13688-018-0163-7>. Acesso em nov. 2018.

ZANDAVALLE, A. C. O mercado de inteligência de mídias sociais. *In*: SILVA, T.; STABILE, M. (Orgs.). **Monitoramento e Pesquisas em Mídias Sociais: metodologias, aplicações e inovações**. São Paulo: Uva Limão, 2016. Disponível em: <https://www.ibpad.com.br/wp-content/uploads/2016/12/Monitoramento-e-pesquisa-em-midias-sociais.pdf>. Acesso em: 28 fev. 2020.

RECOVERY AND CLASSIFICATION OF TWITTER USER FEELINGS IN ELECTORAL PERIOD

ABSTRACT

Introduction: Social networks have become an important space for users to express their feelings. These comments are valuable for governors to know the citizens 'point of view about their political proposals and candidates perceive voters' reaction to the election campaign. **Objectives:** To analyze the feelings expressed by users on Twitter, referring to the candidates that competed for the presidency of Brazil in the year 2018, and to predict the result of the elections based on these posts. **Methodology:** The posts about the candidates that disputed the second round of the elections were the object of study. The software Orange Canvas, a free and open-source machine learning tool, was used to collect the sample and to extract the relevant information. The technique of opinion analysis was applied to automatic classification of feelings into positive, negative and neutral. For better analysis and interpretation of the results, the most important words of comments in word clouds and emotions were shown in frequency distribution charts. **Results:** There were a lot of negative feelings in the posts and the surprise emotion was the one that stood out the most for both competitors. **Conclusions:** The study showed that Twitter is an interesting place for users to express their feelings during the election period. However, the work was not able to predict the outcome of the elections based on the emotions. It is believed that this is due to the high rates of rejection of voters regarding the candidates and the polarization that has characterized Brazilian politics in recent times.

Descriptors: Social networks analysis, Automatic classification, Information retrieval, Information extraction, Text mining.

RECUPERACIÓN Y CLASIFICACIÓN DE SENTIMIENTOS DE USUARIOS DEL TWITTER EN PERÍODO ELECTORAL

RESUMEN

Introducción: Las redes sociales se han convertido en un espacio importante para los usuarios expresar sus sentimientos. Estos comentarios son valiosos para los gobernantes saber el punto de vista de los ciudadanos sobre sus propuestas políticas y candidatos a percibir la reacción de los votantes acerca de la campaña electoral. **Objetivos:** Analizar los sentimientos expresados por los usuarios en Twitter, referentes

a los candidatos que concurrieron a la presidencia de Brasil en el año 2018, y predecir el resultado de las elecciones con base en esos post. **Metodología:** Los posts sobre los candidatos que disputaron la segunda vuelta de las elecciones fueron el objeto de estudio. Se utilizó el software Orange Canvas, una herramienta de aprendizaje de máquina libre y de código abierto, para la recolección de la muestra y para la extracción de la información relevante. La técnica de análisis de opinión fue aplicada para la clasificación automática de los sentimientos en positivos, negativos y neutros. Para un mejor análisis e interpretación de los resultados, se mostraron las palabras más importantes de los comentarios en nubes de palabras y las emociones, en gráficos de distribución de frecuencia. **Resultados:** Se detectaron muchos sentimientos negativos en las entradas y la emoción de sorpresa fue la que más se destacó para ambos competidores. **Conclusiones:** El estudio mostró que Twitter es un sitio interesante para los usuarios expresar sus sentimientos en el período electoral. Sin embargo, el trabajo no fue capaz de predecir el resultado de las elecciones sobre la base de las emociones. Se cree que esto se debe a las altas tasas de rechazo de los votantes en cuanto a los candidatos y la polarización que ha caracterizado la política brasileña en los últimos tiempos.

Descriptores: Análisis de redes sociales, Clasificación automática, Recuperación de la información, Extracción de la información, Minería de texto.

Recebido em: 25/11/2018

Aceito em: 20/11/2019