

Twitter-Eleições: Analisando e minerando o Twitter durante o segundo turno das eleições presidenciais no Brasil em 2014

Thiago L. Andrade
Centro Integrado de Vocação
Tecnológica, UFRN
Natal, Brasil
thiago@limaandrade.com

Renato M. Soares
Centro Integrado de Vocação
Tecnológica, UFRN
Natal, Brasil
rmsnatal@gmail.com

Francisco J. S. Macário
Centro Integrado de Vocação
Tecnológica, UFRN
Natal, Brasil
franciscomacario@outlook.com

ABSTRACT

Twitter-Elections is a data mining web application developed specifically to analyze posts on Twitter during the second round of elections and post-election. Thereby carry out an assessment of the most used hashtags, which candidate is the most popular among other assessments. The results of these analyzes were processed and presented by statistical graphics in a web project.

Keywords

Twitter, Github, mineração de dados, eleições, hashtags, Highchart, metadados.

1. INTRODUÇÃO

Em outubro de 2014 ocorreu no Brasil as eleições políticas e como de costume, durante o período de eleição, e mais fortemente nas semanas que antecedem o dia da eleição, fontes como jornais, bancadas políticas e interessados encomendam pesquisas de intenções de voto à institutos e fundações especializada em pesquisas de consulta popular.

Também nesta época, é normal e democrática a vontade de alguns eleitores de expressar suas intenções de voto, debater propostas e posicionamento político, ideologias e ideias com amigos, familiares ou o público em geral. Com base nisso, é notório um fenômeno interessante que se popularizou desde as eleições anteriores. Cada vez mais as pessoas utilizam redes sociais para expressar posicionamentos e debater política, assim sendo, um grande volume de dados é gerado e lançado em redes sociais em época de eleição. Geralmente isto é feito em redes bem difundidas atualmente e de acesso fácil a grande parte da população mundial, como o Facebook e no microblog Twitter e em específico também são muito utilizados aqui no Brasil como ferramenta de difusão política.

Essas ferramentas são bem aproveitadas por diversas pessoas

para diversas finalidades, seja pela população para defender suas ideias e posicionamentos políticos, ou então por candidatos e seus assessores para realizar a promoção de suas candidaturas e campanhas. Com isso, várias técnicas de marketing são implantadas, como exemplo: uso de hashtags, vídeos, fotos e demais conteúdos de mídia.

Segundo [2], no Brasil, as eleições foram quase três vezes mais movimentadas, com 674,4 milhões de interações no Facebook em três meses e meio de campanha, uma média de 5,96 milhões por dia.

As eleições 2014 foram exploradas nas redes sociais, principalmente no Twitter e Facebook, sendo essas redes utilizadas para promover tudo, desde boca de urna a uma opinião crítica dos eleitores. O intuito de fazer um trabalho relacionado as eleições 2014, foi para responder a pergunta: Quem foi o candidato a presidência mais popular das redes sociais, em particular no Twitter? Para podermos responder essa pergunta foi utilizado técnicas de mineração de dados sobre o microblog. O Twitter foi escolhido para ser o alvo da mineração devido que conseguimos obter informações mais simples e mais intuitivas sobre as intenções de voto dos usuários, devido principalmente ao fato de que cada tweet tem um limite de caracteres, logo o usuário precisa expressa em poucas palavras o que pensa, assim muitas vezes tendo que dizer de forma direta em qual candidato está apoiando. Postagens nessas redes geram um grande volume de informação que vão bem além do conteúdo publicado, pois outras informações estão atreladas ao conteúdo principal, são esses os metadados. Localização da postagem, data e hora, pessoas envolvidas, hashtags, links, mídias como fotos e vídeos, quantidade de vezes que a postagem foi curtida/compartilhada, retweets, dentre outros são exemplos de metadados que podem estarem relacionados a uma única postagem. Por meio deles grandes estudos podem ser encorajados, como por exemplo localidades onde certa hashtag foi usada. Este volume de dados a ser explorado incentiva a criação do Twitter-Eleições.

Na primeira seção falamos sobre a mineração de dados e porque o uso da mineração de dados foi importante para o nosso trabalho. Na segunda seção falamos sobre os trabalhos relacionados ao nosso trabalho. Já na terceira seção, o foco é nas ferramentas utilizadas para realizar a mineração de dados, além de ferramentas para gerenciamento de tarefas. A quarta seção ficou destinada as métricas de comparação,

onde falamos sobre as medidas utilizadas para comparar os dados. A quinta seção aborda sobre os resultados do nosso trabalho e a sexta seção sobre a qual conclusão chegamos após a mineração e análise dos dados.

2. MINERAÇÃO DE DADOS

Nesta seção, será apresentado sobre a técnica de análise de dados utilizada para obter as informações sobre as eleições 2014. Com o intuito de obter informações relevantes a partir de rede sociais, foi necessário realizar mineração de dados sobre as informações obtidas, utilizando por exemplo técnicas apresentadas em [5].

A [9] define a mineração de dados como um processo de explorar grandes quantidades de dados à procura de padrões consistentes, para detectar relacionamentos sistemáticos entre variáveis, detectando assim novos subconjuntos de dados. A mineração de dados é formada por um conjunto de ferramentas e técnicas que através do uso de algoritmos de aprendizagem ou classificação baseados em redes neurais e estatística, são capazes de explorar um conjunto de dados, extraindo ou ajudando a evidenciar padrões nestes dados e auxiliando na descoberta de conhecimento. Esse conhecimento pode ser apresentado por essas ferramentas de diversas maneiras: agrupamentos, hipóteses, regras, árvores de decisão, grafos ou dendrogramas.

O uso da mineração de dados, foi bastante importante para podermos coletar as informações relevantes e assim responder ao nosso questionamento, sobre: Qual o candidato da eleição para presidência no segundo turno era o mais popular e importante na rede social Twitter.

3. TRABALHOS RELACIONADOS

Uma pesquisa semelhante foi realizada por [1] na Alemanha, onde se desejavam fazer uma investigação que comparem por meio de análise de sentimento o que as pessoas falavam on-line no Twitter com informações que eram colhidas off-line, assim perceberam que uma análise do sentimento político nos tweets demonstraram na Alemanha em 2010 uma estreita correspondência com os partidos políticos e posições políticas dos políticos, indicando assim que o conteúdo das mensagens no Twitter se repetem de forma plausível também off-line.

4. FERRAMENTAS UTILIZADAS

Nesta seção, será apresentado as ferramentas utilizadas para compor o trabalho desse artigo. Com o intuito de obter informações relevantes a partir de rede sociais, foi necessário realizar mineração de dados sobre as informações obtidas.

Uma série de ferramentas foram utilizadas para realizar uma análise dos dados do Twitter no período eleitoral. A integração dessas tecnologias foram importantes em todas as fases do projeto e por isso serão descritas as tecnologias utilizadas e suas importâncias, foram elas:

- GitHub
- Twitter API
- API do Twitter para Python

- Twitter4J
- HighCharts
- Ruby on Rails

GitHub é um serviço de web para compartilhar projetos que usam o controle de versão *Git*. Escrito em *Ruby on Rails* pelos desenvolvedores da Logical Awesome, esse sistema possui funcionalidades de uma rede social como feeds, followers, wiki e um gráfico que mostra como os desenvolvedores trabalham as versões de seus repositórios. Além disso, o GitHub possui um sistema de *issues*, onde pode ser utilizado como um gerenciador de tarefas e/ou processo, os *issues* podem ser considerados as tarefas e os milestones podem ser a *Sprint*.

A rede social Twitter, disponibiliza para quem quiser uma API [7], onde através dela é possível buscar informações de dados (os tweets) em sua plataforma. É possível fazer uma busca direta por tweets, podendo passar na consulta parâmetros como, período da busca, o conteúdo da busca, podendo ser frase, hashtags, usuários, etc. Também é possível utilizar o mecanismo de stream, onde fica-se escutando o em um canal de stream sobre informações, exemplo, quando uma hashtag específica é utilizada ou o que um usuário específica fica twitando.

Inicialmente para facilitar o uso da API do Twitter, foi utilizado uma biblioteca que simplifica os recursos da API do Twitter, a biblioteca é fornecida em python [6]. Utilizamos ela no início para poder coletar as informações que queríamos. Mas devido principalmente a linguagem em si, onde nenhum dos integrantes do trabalho tinha o domínio, paramos de usá-la, mas os dados coletados foram guardados para ser utilizados na análise dos dados.

Após a dificuldade da utilização da API de python, foi necessário buscar uma outra alternativa para continuar com a mineração dos dados, foi então que encontramos a biblioteca *Twitter4J* [8], onde é uma biblioteca que faz o uso da API do Twitter, mas escrita em Java, facilitando muito o trabalho, devido ao domínio de todos os integrantes do trabalho. Para facilitar a análise dos dados, utilizamos de banco de dados relacional, para guardar as informações que queríamos e futuramente fazer a análise dos dados.

Com a análise dos dados concluída, sentiu-se a necessidade de mostrar esses dados de uma maneira mais simples e mais objetiva, foi nesse momento onde resolvemos criar um sistema web, para criar gráficos utilizando uma biblioteca em Javascript chamada *HighCharts* [4] e assim podemos publicar de maneira fácil e rápida os nossos resultados. Para fazer o sistema web, foi utilizado o framework *Ruby on Rails*, devido ao seu desenvolvimento ágil. Não tivemos muitos problemas ao utilizar o framework *Ruby on Rails*, pois um dos integrantes já tinha trabalho com o mesmo em trabalhos anteriores.

5. MÉTRICAS DE COMPARAÇÃO

Para obter informações com base nos dados minerados do Twitter, foram utilizadas as hashtags mais populares de cada

candidato, para assim, se ter um novo conhecimento quantitativo de uma determinada informação.

Abaixo, poderemos ver alguns exemplos de hashtags que foram mais utilizadas para consulta de tweets na mineração:

- #Dilma
- #Aecio
- #Dilma13
- #Aecio45
- #Voto13
- #Voto45

Foi realizada também uma análise quantitativa referente aos tweets da candidata vencedora da eleição, Dilma Rousseff. Para isso, foram utilizadas as seguintes hashtags mais populares após os resultados da eleição do segundo turno. Vejamos exemplos de algumas delas:

- #Dilmais
- #Dilmais4Anos
- #DilmaDeNovo
- #DilmaMudaMais
- #MudaMais
- #Dilmãe

6. RESULTADOS

Nesta sessão serão avaliados os resultados obtidos por nosso projeto por meio de um tratamento dos tweets colhidos usando a metodologia descrita anteriormente, e em seguida faremos uma comparação com pesquisas eleitorais de intenção de voto encomendadas por diferentes órgãos e divulgadas em mídias do Brasil como TV e jornais.

Para esta análise iremos comparar os dados no Twitter por proporção, ou seja, os dados das pesquisas off-line serão comparados em porcentagem relativa a quantidade de tweets colhidos naquele período.

No resultado abaixo temos a quantidade de ocorrências de cada hashtag a favor de Dilma, ou seja, representa as hashtags mais utilizadas nos tweets referenciados à presidenta Dilma Rouseff. Com isso, podemos observar que a hashtag #DilmaNovamente foi a mais utilizada. Além disso, no segundo gráfico, podemos observar a hashtag presente nos tweets que fizeram mais sucesso, ou seja, a hashtag de maior sucesso, que foi a #Dilmais4Anos. **Ver Figure-1.**

Ainda se tratando da popularidade dos tweets, no gráfico seguinte, podemos observar que a hashtag mais popular #Dilmais4Anos está presente no tweet mais popular, com a seguinte descrição: "RT @dilmabr: Muito obrigada! #Dilmais4anos <http://t.co/LcPPKMO>. Podemos ver com isso, que o tweet mais popular durante as eleições foi criado pelo twitter oficial

da própria presidenta vencedora, "@dilmabr". **Ver Figure-2.**

Além disso, tratando a popularidade do tweet de acordo com a quantidade de vezes favoritadas, temos ainda que a hashtag #Dilmais4Anos é a mais popular. O que confirma o tweet mais popular da própria presidenta. Podemos ver no gráfico abaixo o resultado. **Ver Figure-3.**

Foi realizada também duas consultas dos tweets momento antes da eleição, nos dias 25 e 26 de outubro. O primeiro gráfico representa a popularidade dos candidatos de acordo com a ocorrência do nome de cada um. De acordo com essa pesquisa, foi observado que a presidenta Dilma estaria mais popular no Twitter, com 56,8%, já o presidente Aécio estaria com 43,2%. Já no segundo gráfico, a métrica utilizada foi a popularidade dos tweets que referencia cada candidato de acordo com a soma do número de favoritos e o número de retweets. Com essa segunda pesquisa, a presidenta Dilma ficou com 60,1% e o presidente Aécio com 39,9%. **Ver Figure-4 e Figure-5.**

No mesmo dia, ou seja, 25 de outubro, as consagradas fontes de pesquisas brasileiras *IBOPE* e *Datafolha* também realizaram pesquisas de intenções de votos publicadas em [3]. Na pesquisa do Datafolha deste dia a candidata Dilma Rousseff expressava 52% das intenções de votos validos, contra 48% do candidato Aécio Neves. Também no dia 25 de outubro, o IBOPE apresentou uma pesquisa onde Dilma Rousseff tinha 57% dos votos validos contra 47% de Aécio Neves. **Ver Figure-6.**

Na próxima pesquisa, temos a análise das ocorrências de hashtag de voto, de acordo com as hashtags #Dilma13 e #Aecio45 após os resultados do segundo turno da eleição. Analisando o gráfico, podemos observar que a hashtag que representa o voto de Aécio já não tinha mais representatividade comparando com Dilma. No dia após a eleição, podemos ver que Aécio ainda tinha uma boa quantidade de declarações de voto (129), mas após o dia 28 essa quantidade cai absurdamente para 14, o que representou uma queda de 89,14%. As pessoas que votaram em Aécio já não queriam mais declarar seu voto. Já Dilma, após ter saído vencedora, teve uma quantidade de declarações de voto quase 4 vezes mais do que Aécio no primeiro dia após a eleição, caindo menos de 10% no dia seguinte (28), ou seja, dois dias após as eleições, as pessoas ainda declaravam seu voto. **Ver Figure-7.**

7. CONCLUSÃO

Com a implementação da aplicação e a obtenção dos resultados, concluímos que realizar pesquisas em redes sociais é mais econômico e rápido de se obter informações sobre determinado assunto, para pesquisas quantitativas. No caso da nossa aplicação, podemos observar resultados interessantes e com base em períodos determinados, mesmo tendo feito a mineração dos dados fora daquele período, sendo assim, o pesquisador tem o poder não só da informação daquele momento, mas sim, pesquisar por determinado período em que a situação pôde se modificar.

No Boca de Twitter, foi feita a mineração de dados antes e após os resultados do segundo turno. Com essa análise, foi

observado que houve uma variação significativa após a divulgação do candidato eleito, como foi relatada na sessão de resultados. Além disso, a implementação do Boca de Twitter propiciou um aprendizado em diferentes tecnologias, incluindo duas APIs do Twitter, como explicitadas na sessão sobre as ferramentas utilizadas.

8. REFERENCES

- [1] T. O. S. Andranik Tumasjan. Predicting elections with twitter: What 140 characters reveal about political sentiment. *Association for the Advancement of Artificial Intelligence*, 2010.
- [2] R. Barifouse. Eleicoes 2014: Novos habitos criam pleito mais conectado do mundo, 2014.
- [3] U. eleicoes 2014.
<http://eleicoes.uol.com.br/2014/pesquisas-eleitorais/brasil/2-turno/>. Datafolha, TV Globo e Folha de S.Paulo, Last access: 2014-11-20.
- [4] Highcharts. Highcharts and highstock documentation.
<http://www.highcharts.com/docs>. Last access: 2014-10-11.
- [5] M. A. Russell. *Mining the Social Web*. O'Reilly, 2014.
- [6] G. sixohsix. Api twitter python doc.
- [7] Twitter Developers. Rest apis / twitter developers.
- [8] Twitter4J. Twitter4j - a java library for twitter api.
<http://twitter4j.org/en/index.html>. Last access: 2014-10-11.
- [9] Wikipedia. Mineracao de dados. Last access: 2014-11-20.

APPENDIX

A. FIGURAS



Figure 1: Hashtags e retweets a favor de Dilma.

Tweets mais populares a favor de Dilma



Figure 2: Tweets populares a favor de Dilma.

Hashtags Favoritas a favor de Dilma

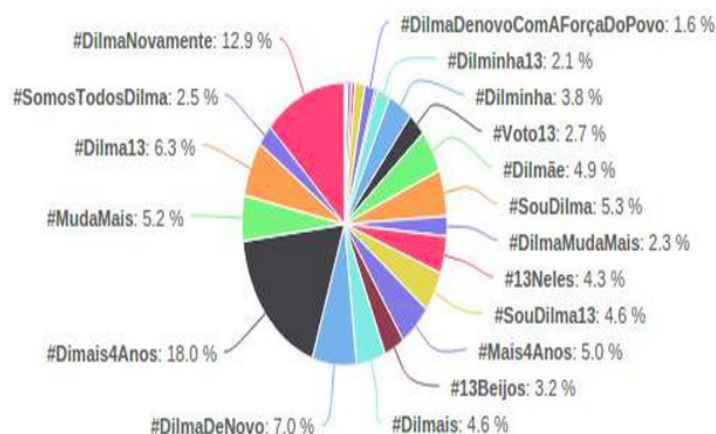


Figure 3: Hashtags favoritas a favor de Dilma.

Ocorrências do nome dos candidatos antes do segundo turno (entre 25 e 26 de Outubro)

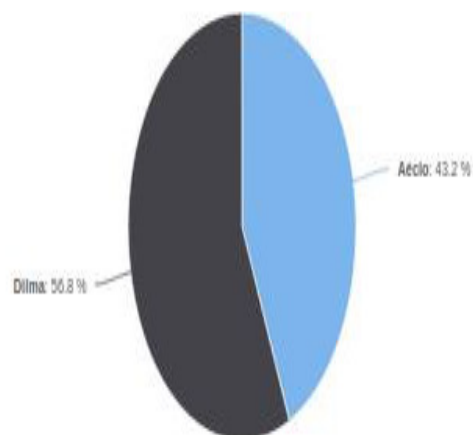


Figure 4: Ocorrência dos nomes dos candidatos no dia 25 de outubro

Candidatos com mais tweets retweetados + favoritos (entre 25 e 26 de Outubro)

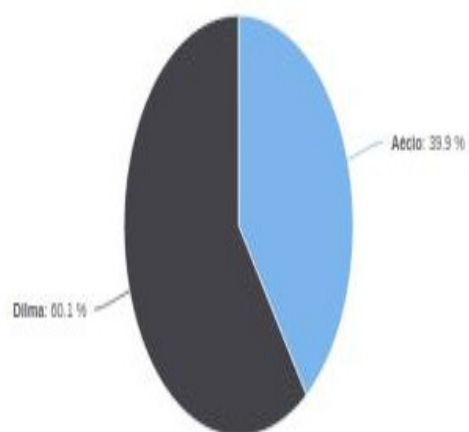


Figure 5: Ocorrência de favoritos e retweets no dia 25 de outubro por candidato



Figure 6: Pesquisa IBOPE e DATAFOLHA no dia 25 de outubro.

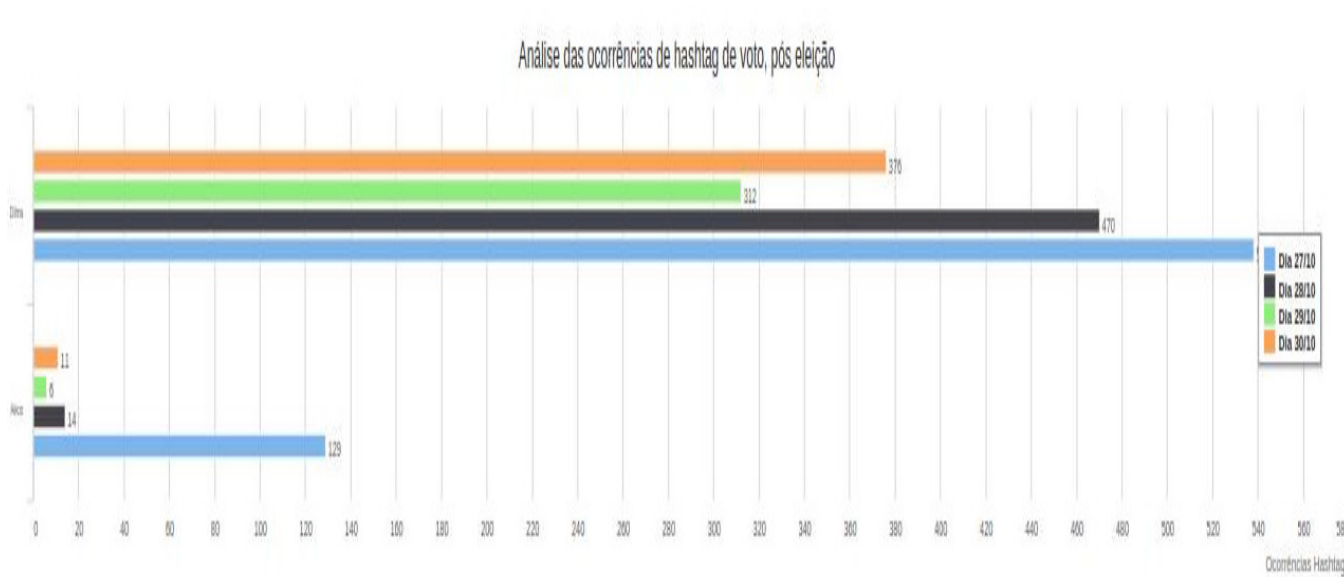


Figure 7: Hashtags de votos pos-eleicao