



ANÁLISE DE TWEETS POLÍTICOS POR MEIO DO APRENDIZADO DE MÁQUINA

Eduardo Pioli do AMARAL¹; Renan Magalhães CICÍLIO²; Diego Saqui³

RESUMO

O estudo realiza, através do *Latent Dirichlet Allocation*, uma análise de texto de diversas postagens de políticos famosos brasileiros, com o propósito de verificar as estratégias de discurso que eles utilizam, bem como os assuntos que geralmente aborda, e se há alguma compatibilidade com o que é observado normalmente. É interessante abordar essa questão tendo em vista a grande influência que políticos são capazes de atingir com uso das redes sociais, podendo trazer uma visão mais crítica e fria sobre um assunto abordado muitas vezes de forma emocional. O trabalho por fim trata os dados e aplica ao modelo LDA para gerar os tópicos que resumem a essência das postagens.

Palavras-chave:

LDA; Discurso Político; Machine Learning; Análise de Texto.

1. INTRODUÇÃO

O discurso político possui muitas formas e varia conforme o tempo, além de ser construído com estratégias e técnicas específicas para se atingir um fim (DYLGJERI, 2023, p. 19-20). Dentre estes fins destacam-se a coerção, resistência, protesto e dissimulação; muito em prol de angariar a simpatia do povo e guiá-los a uma opinião específica. Atualmente ferramentas políticas comuns são as redes sociais, já que desempenham um papel fundamental na disseminação de informações, mobilização de eleitores e influência nas opiniões públicas (STIEGLITZ e DANG-XUAN, 2012).

Estudos recentes destacam a importância das redes sociais na moldagem da opinião pública e no comportamento político dos indivíduos. Pesquisas conduzidas mostram como ela tem a capacidade de criar e fortalecer comunidades políticas virtuais (CASTELLS, 2012, BUNZ, 2016). Uma forma de melhor identificar padrões em texto e possíveis técnicas empregadas em falas e discursos é através da análise de texto, performada por meio do aprendizado de máquina (ML). *Latent Dirichlet Allocation* (LDA) é um modelo probabilístico de ML não supervisionado e processamento de linguagem natural para a tarefa de modelagem de tópicos. É baseado na suposição de que cada documento em um corpus é feito por uma mistura de vários tópicos latentes, e cada tópico é uma distribuição de probabilidade sobre um conjunto de palavras. (BLEI, 2003).

Com isso, o foco do trabalho é aplicar o LDA a um *corpus textual* que reúne textos atribuídos a políticos, com o intuito de melhor identificar as estratégias utilizadas e se há perfis

¹Bolsista Eduardo Pioli do Amaral, IFSULDEMINAS – Campus Muzambinho. E-mail: 12201000758@muz.ifsuldeminas.edu.br.

²Bolsista Renan Magalhães Cicilio, IFSULDEMINAS – Campus Muzambinho. E-mail: 12201000343@muz.ifsuldeminas.edu.br.

³Orientador Diego Saqui, IFSULDEMINAS – Campus Muzambinho. E-mail: diego.saqui@muz.ifsuldeminas.edu.br.

gerais que podem ser traçados por meio do discurso político, tendo em vista não só o potencial coercitivo dos mesmos, como a difusão do acesso a eles. Tendo em vista o potencial coercitivo destes discursos, e o quão difundido podem ser, muito pelas redes sociais, este trabalho propõe analisar tweets de 3 candidatos à presidência em 2022, Luiz Inácio Lula da Silva, Jair Messias Bolsonaro e Simone Nassar Tebet, com o objetivo de encontrar padrões e temas presentes nas principais palavras que utilizaram (S. GEETHA e VISHNU KUMAR KALIAPPAN, 2018).

3. MATERIAL E MÉTODOS

Para realização do trabalho foi usado um *dataset*⁴ que reúne 23.165 postagens (*tweets*), feitas no Twitter, divididas entre os políticos Jair Messias Bolsonaro, Luiz Inácio da Silva e Simone Nassar Tebet, compreendendo um período de 25 de Junho de 2018 a 19 de Maio de 2023.

Antes de criar o modelo LDA, foi feito o pré-processamento dos dados através da ferramenta *Natural Language Toolkit* (NLTK)⁵. Primeiramente, normalizou-se todos os *tweets* colocando-os em caixa-baixa. Enfim, declarou-se quais são as palavras vazias, palavras de baixa ou nula significância, comuns em textos, para isso removeu-se as chamadas *stopwords*. Por fim, transformou-se cada palavra em *tokens*, somente se formados por letras e não fossem palavras vazias. O *dataset* pré-processado foi então usado na modelagem do LDA pela biblioteca *Gensim*⁶.

Ao aplicar o modelo LDA no dataset, espera-se identificar e categorizar os principais tópicos discutidos pelos políticos Bolsonaro, Lula e Tebet em suas postagens no Twitter, permitindo uma compreensão mais profunda dos assuntos abordados por cada um deles.

4. RESULTADOS E DISCUSSÃO

Tabela 1 - Tabela de tópicos dos *tweets* do candidato Bolsonaro.

Palavra	Frequência	Palavra	Frequência
ano	0.011	governo	0.006
país	0.010	brasileiro	0.005
estado	0.007	obra	0.005
emprego	0.006	povo	0.005
ação	0.006	investimento	0.005

⁴ [2022_Brazilian_Presidential_Candidates_Tweets](#)

⁵ [NLTK](#)

⁶ [Gensim: Topic modelling for humans](#)

Fonte: Elaborada pelos autores.

Tabela 2 - Tabela de tópicos dos *tweets* do candidato Lula.

Palavra	Frequência	Palavra	Frequência
povo	0.028	gente	0.012
país	0.027	pessoa	0.011
governo	0.016	mundo	0.009
presidente	0.014	vida	0.008
ano	0.012	fome	0.006

Fonte: Elaborada pelos autores.

Tabela 3 - Tabela de tópicos dos *tweets* da Tebet.

Palavra	Frequência	Palavra	Frequência
mulher	0.011	presidente	0.009
senado	0.011	projeto	0.008
país	0.010	vida	0.007
governo	0.009	pessoa	0.006
ano	0.009	apoio	0.006

Fonte: Elaborada pelos autores.

Como era de se esperar, todos os políticos fazem menção frequente a pautas governamentais. Nota-se também certas divergências de comportamento entre eles. O político Jair Bolsonaro utiliza palavras que exaltam uma ação e tração do governo, e faz alusões ao nacionalismo, bem como salienta a parte econômica de seu governo, como pode ser visto na Tabela 1. Já o político Lula emprega mais as pautas sociais, e faz referência à própria população tendo um discurso que apela mais ao pessoal, como nota-se na Tabela 2. Por fim, a política Simone Tebet apresenta foco em projetos políticos, em ações, tração e apoio; evocando também aproximação pessoal e vocabulário mais técnico político; além de referenciar as mulheres no meio sócio-político, uma forma de

associar à sua causa, vide a Tabela 3.

5. CONCLUSÃO

Com a análise de texto das postagens políticas verificou-se a presença de um discurso direcionado à base aliada de cada político. Pelos tópicos foi possível inferir que, com ajuda dos textos de apoio, os tipos de reações políticas buscam evocar em seus eleitores e simpatizantes.

REFERÊNCIAS

- 1- BLEI, David M.; JORDAN, Michael I.; NG, Andrew Y. Latent Dirichlet Allocation. **Journal of Machine Learning Research**, [s.l.], vol. 3, p. 993-1022, Jan. 2003. Disponível em: <https://jmlr.csail.mit.edu/papers/v3/blei03a.html>. Acesso em: 26 Jun. 2023.
- 2- BUNZ, Mercedes. **Book Review: The Internet of Things: tracing a new field of enquiry**. *Media, Culture & Society*, v. 38, n. 8, p. 1278–1282, 22 Set 2016. Acesso em: 28 Jun. 2023.
- 3- CASTELLS, Manuel. **Redes de indignação e esperança: movimentos sociais na era da internet**. Tradução Carlos Alberto Medeiros. Rio de Janeiro: Zahar, 2013. 271 p. *Horizontes Antropológicos*, v. 21, n. 44, p. 407–410, Dez 2015. Disponível em: <http://www.scielo.br/pdf/ha/v21n44/0104-7183-ha-21-44-0407.pdf>. Acesso em: 27 Jun. 2023.
- 4- DYLGJERI, Ardita. ANALYSIS OF SPEECH ACTS IN POLITICAL SPEECHES. **European Journal of Social Sciences Studies**, [s.l.], vol. 2, n. 2, p. 19-26, Fev. 2017. Disponível em: <https://oapub.org/soc/index.php/EJSSS/article/view/66>. Acesso em: 26 Jun. 2023.
- 5- KEYVAN KHOSROVIAN e PFAHL, Dietmar e VAHID GAROUSI. **GENSIM 2.0: A Customizable Process Simulation Model for Software Process Evaluation**. p. 294–306, 5 Maio 2008. Acesso em: 26 Jun. 2023.
- 6- S. GEETHA e VISHNU KUMAR KALIAPPAN. **Tweet Analysis Based on Distinct Opinion of Social Media Users**. p. 251–261, 12 Dez 2018. Acesso em: 27 Jun. 2023.
- 7- STIEGLITZ, Stefan e DANG-XUAN, Linh. **Social media and political communication: a social media analytics framework**. *Social Network Analysis and Mining*, v. 3, n. 4, p. 1277–1291, 25 Ago 2012. Disponível em: <https://link.springer.com/article/10.1007/s13278-012-0079-3>. Acesso em: 26 Jun. 2023.