

# COMO O DEBATE ECONÔMICO NAS REDES SOCIAIS NORTEOU O VOTO NAS ELEIÇÕES BRASILEIRAS DE 2022

**Palavras-Chave:** twitter; clusterização; voto econômico; eleições 2022; debate econômico

**Autores(as):**

**Thiago Lopes Garcia, UNICAMP – IE**

**Prof. Dr Alexandre Gori Maia (orientador), UNICAMP - IE**

---

## INTRODUÇÃO:

Em 2022, as eleições presidenciais foram marcadas por uma grande polarização, especialmente, no 2º turno, onde ambos os candidatos, Jair Messias Bolsonaro (“Bolsonaro”) e Luiz Inácio Lula da Silva (“Lula”), apresentavam opiniões políticas e econômicas antagônicas. Além disso, desde 2018, as redes sociais vêm apresentando um papel fundamental e determinante no resultado das eleições.

Perante isso, este se tornou um momento ímpar para se compreender como os diferentes discursos econômicos podem atuar em favor de um candidato dentro da corrida eleitoral. Ainda mais, com o crescente papel das redes sociais nesse resultado, aliado à maior facilidade de acesso às informações dos usuários por meio de API's (Application Programming Interface) permitiram as ferramentas necessárias para se realizar este estudo.

Portanto, a pesquisa em questão busca analisar como a percepção das variáveis econômicas influenciam no processo decisório dos agentes de uma economia. Em especial, como que este foi responsável por nortear os resultados das eleições presidenciais de 2022 no Brasil.

Para isso, se utilizou do “Twitter” (Atualmente chamado “X”), afinal, esta é a principal rede social que permite que seus usuários recebam e enviem atualizações sobre a percepção de suas vidas, assim, sendo ideal para conseguir compreender a interpretação dos agentes de uma sociedade sobre os seus recentes eventos. Além disso, o Brasil se mostra com uma altíssima taxa de uso da rede social, com mais de 19 milhões de usuários em 2021 (Statista, 2022).

O objetivo geral deste projeto é, através da linguagem de programação Python, tratar os dados coletados pela API do Twitter (tweets) e através do uso de algoritmos não supervisionados (como por exemplo K-Mean), analisar o perfil dos usuários desta rede social e seu possível comportamento perante os discursos econômicos pontuados.

## REFERENCIAL TEÓRICO:

De um modo geral, a pesquisa em questão leva consigo a noção do “voto econômico”, conceito este que foi muito bem pontuado e problematizado a seguir:

“O chamado voto econômico constitui uma escolha individual discreta na qual os indivíduos atribuem valores às alternativas políticas (candidatos e/ou partidos) com base em suas avaliações do desempenho econômico do país, estado, ou qualquer outro nível de governo (Duch & Stevenson 2008, p. 42). Além disso, também é possível que os indivíduos tenham como referência diferentes períodos de tempo ao emitirem suas avaliações (Lewis-Beck 1985). Eleitores podem avaliar retrospectiva ou prospectivamente o desempenho da economia, ou podem simplesmente levar em consideração a presente situação sem qualquer referência a

passado ou futuro. Além do mais, ainda que a definição dada acima focalize apenas o desempenho econômico de coletividades (tal avaliação também é chamada de "sociotrópica"), parte da literatura também inclui na rubrica "voto econômico" o ato de basear a escolha na avaliação da situação econômica pessoal." (Pereira, 2014, p. 151).

A partir deste excerto, é possível notar que uma das grandes dificuldades desse campo de estudo é que a noção do voto econômico não é única, indivíduos levam em conta diversos fatores para escolher qual será seu voto. Sendo que destes fatores, cada indivíduo terá o seu próprio e que, provavelmente, difere do próximo indivíduo.

A hipótese central é que o discurso nas eleições estava muito mais centrado em problemas cotidianos, como preço dos alimentos e combustíveis, que no discurso de políticas macroeconômicas. Isto é, o voto é pautado muito mais em questões palpáveis e resultados econômicos do que necessariamente na teoria econômica que o candidato propõe.

### **METODOLOGIA:**

A coleta de dados havia sido realizada previamente pelo professor orientador, Alexandre Gori Maia, para uso em pesquisa própria pelo uso da API disponibilizada pelo até então "Twitter". A estratégia de extração dos dados foi feita de maneira similar à referência a seguir:

"The selected keywords included the words (translated from Portuguese to English): "COVID", "pandemic", "mask", "isolation", "quarantine", "distancing", "vaccine", "coronovac", "health", "Bolsonaro", "communism", and "dictatorship" (these two latter words were selected because many tweets linked vaccination to communist dictatorships, also exemplifying the ideological aspects of the COVID-19 pandemic). The API matched each keyword within the body of a tweet ignoring special characters (Twitter Developer Platform, 2022b). For example, the texts including the hashtag "#covid19" or the word "Covid-19" were both matched with the keyword "covid." Characters with accents or diacritics are treated the same as normal characters" (GORI MAIA, A. et al. 2023).

Entretanto, neste caso, as palavras-chaves referenciadas foram os nomes dos principais candidatos participantes da corrida eleitoral de 2022; sendo neste caso "Bolsonaro", "Lula", "Ciro", "Dória" e "Simone Tebet". Sendo assim, a API mapeou todos os tweets realizados dentro do período das eleições que referenciam estas palavras e as armazenou.

De modo geral, a base é composta por uma série de informações sobre o usuário, contudo, as informações mais relevantes para este trabalho, são que há um identificador para cada usuário, o "id", e seu respectivo tweet. Assim, a base compreende os tweets referentes à semana anterior ao primeiro turno das eleições (25/09/2022 até 01/10/2022), com um total de 1.043.027 observações.

Perante isso, buscou-se métodos que pudessem compreender, inicialmente, quais eram as principais variáveis econômicas que estavam sendo atreladas aos candidatos. A estratégia utilizada aqui foi de fazer uma pré-seleção de termos econômicos cotidianos e fazer uma busca de quantas vezes essas palavras eram mencionadas dentro da base de dados, caso a contagem apresentasse aderência, isto é, fosse mencionada um número relevante de vezes, ela seria adicionada à lista de palavras-chaves do macro-grupo "Economia", caso contrário, elas seriam descartadas.

A partir disso, com uma seleção das palavras correlatas à economia que mais apareciam na base de dados, juntamente com outras adicionadas posteriormente, que foram julgadas como termos fundamentais às ciências econômicas, foi criada a lista com todas as palavras relacionadas à economia. A mesma coisa foi feita para termos políticos e relacionados à saúde, visto que foram assuntos muito debatidos na época, em especial o último, por causa da pandemia de COVID-19 em 2022.

Vale pontuar que, de um modo geral, dentro das redes sociais, há uma despreocupação com a grafia exata das palavras, assim, para que a pesquisa destes termos das listas fossem processadas, necessitou-se que todos os tweets da base tivessem sua grafia corrigida.

## CLUSTERIZAÇÃO:

Clusterização é uma técnica de aprendizado não-supervisionado com o intuito de agrupar os dados similares entre si. Escolheu-se o método de “K-means” como método de agrupamento dos dados uma vez que ele seria o mais adequado para tratar grandes conjuntos de dados, além de permitir facilmente a incorporação de novas observações ao modelo, caso necessário.

Ao contrário de outros algoritmos, onde o número de agrupamentos é determinado posteriormente, no K-means é necessário definir previamente os K clusters antes da execução. Esta técnica começa selecionando aleatoriamente K pontos iniciais como os centros dos grupos (centróides) e distribuindo os dados em grupos baseados em qual centróide está mais próxima daquela observação. Em seguida, o algoritmo calcula a distância entre cada dado e seu centróide e, posteriormente, calcula a média dessas distâncias. Com base nessa média, uma nova coordenada cartesiana é gerada, que se torna a nova posição do centróide, possivelmente resultando na alteração dos itens que compõem aquele grupo.

Em seguida, através da linguagem de programação Python, submeteu-se a base de dados à metodologia de clusterização “K-Mean”. No caso, para que os dados fossem distribuídos e com isso, pudesse ser calculadas as centróides do modelo, utilizou-se metodologia semelhante à seguinte:

“To Identify the tweets that are related to family and the time of the sentence, we used the LIWC-22 software (R. L. Boyd et al., 2022). This software works by using internal dictionaries in different languages. These dictionaries were constructed using a combination of human expertise, algorithms and statistical models (R. L. Boyd et al., 2022). The internal English dictionary, for example, consists of over 12,000 words, word stems, phrases, and emojis; and each dictionary entry can belong to more than one category (R. L. Boyd et al., 2022). LIWC-22 uses word counting to build standardized scores expressed as percentages, as explained in the LIWC-22 webpage documentation: ‘LIWC reads a given text and compares each word in the text to the list of dictionary words and calculates the percentage of total words in the text that match each of the dictionary categories. For example, if LIWC analyzed a single speech containing 1,000 words using the built-in LIWC-22 dictionary, it might find that 50 of those words are related to positive emotions and 10 words related to affiliation. LIWC would convert these numbers to percentages: 5.0% positive emotion and 1.0% affiliation.’ (LIWC-22, n.d.)” (GIL-CLAVEL, S.; MULDER, C. H. 2024)

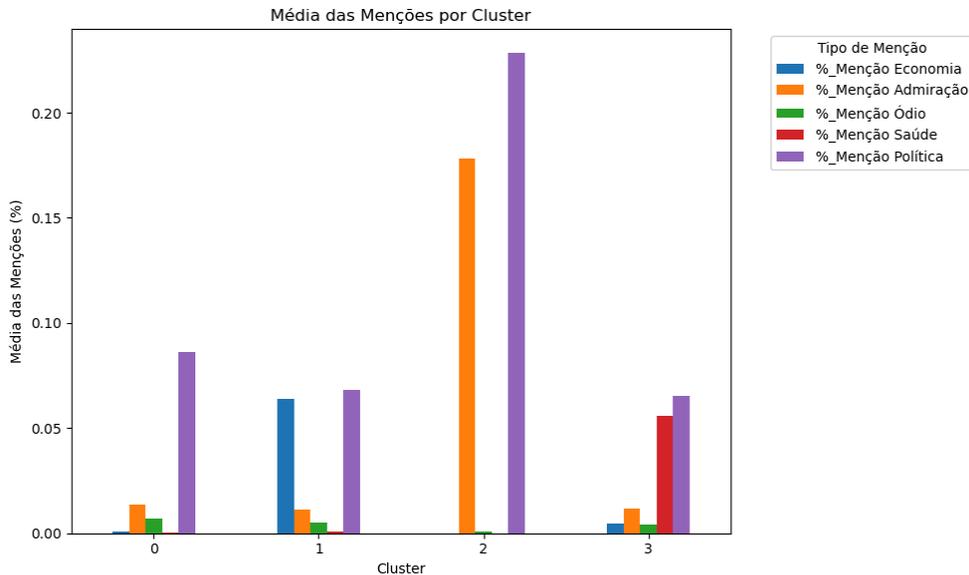
No caso, a estratégia utilizada baseou-se em contar a quantidade total de palavras em cada um dos tweets e, em seguida, calcular a porcentagem de palavras perante as variáveis chaves. Isto é, se em um tweet houvesse 30 palavras e apenas 2 palavras que ele mencionasse pertencesse à lista de palavras de Economia já pré-estabelecidas, então, naquele tweet haveria 6,6% de Menção à termos econômicos. Portanto, para agrupar os dados, além da porcentagem de menção à economia, também foram calculadas menções à saúde, política, ódio e admiração. Sendo que estas duas últimas foram estabelecidas como uma forma de se fazer uma análise de sentimentos dos tweets.

Com isso, passado a clusterização, as observações foram divididas da seguinte forma:

- Cluster 0: 794.349 (76,2%)
- Cluster 1: 36.963 (3,5%)
- Cluster 2: 199.896 (19,2%)
- Cluster 3: 11.819 (1,1%)

## ANÁLISE DO CONTEÚDO DOS CLUSTERS

Feito isso, como uma forma de poder compreender os clusters e, a partir disso, o conteúdo dos tweets, utilizou-se a média da menção a cada uma das variáveis selecionadas.



**Figura 1:** Gráfico indicando a média da porcentagem de menção das variáveis selecionadas. Fonte: Elaboração própria (2024).

### Cluster 0

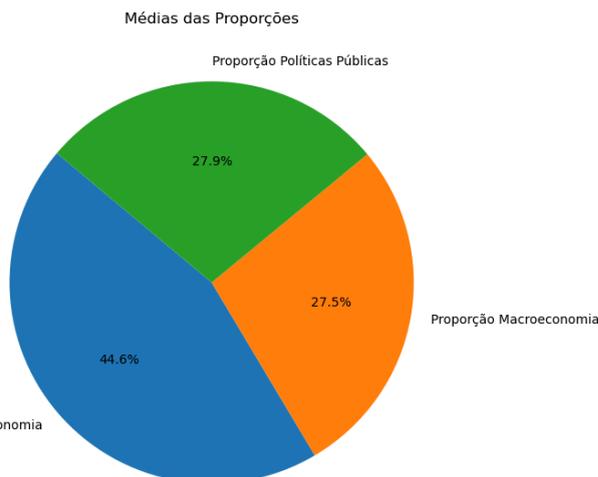
Esse cluster é dominado, principalmente pelo o que não passou no crivo dos outros clusters. Isto é, aqui são tweets que por muitas vezes estão mencionando o nome dos candidatos mas o usuário não necessariamente está colocando em pauta uma questão política ou econômica. Por mais que seja sim possível inferir qual o posicionamento do tweet em relação às eleições, ainda assim este é um fato secundário e pouco desenvolvido.

### Cluster 1

A partir das variáveis utilizadas, este cluster é caracterizado, predominantemente, pela discussão da economia. Sendo que aqui já é possível notar mais claramente qual o posicionamento das pessoas dentro dessa discussão econômica, além de que, mesmo com o espaço limitado dos tweets, aqui as temáticas são um pouco mais desenvolvidas.

Como o enfoque principal do trabalho é compreender o debate econômico, buscou-se aprofundar a análise do Cluster 1, visto que, este é o Cluster onde mais se debate economia.

Assim, para se compreender melhor a temática do que estava sendo discutido, utilizou-se a já utilizada lista “Economia” e esta foi subdividida em 3 novas listas, sendo estas referentes à aspectos micro, macro e de políticas públicas. Sendo que a composição de cada uma dessas novas listas foi criada a partir de curadoria.



**Figura 2:** Gráfico indicando a média da porcentagem de menção das variáveis selecionadas. Fonte: Elaboração própria (2024).

A partir do gráfico, é possível averiguar que a principal temática debatida nas redes sociais estava muito mais centrada no cotidiano dos usuários (temas microeconômicos) e em sua experiência econômica individual do que

necessariamente em questões universais e públicas.

### **Cluster 2**

O mais latente de todos os clusters, sendo marcado principalmente por tweets enaltecendo principalmente os candidatos durante a eleição. Interessante notar que as outras variáveis, como a de Menção à Saúde, Economia e até mesmo Palavras de Ódio são praticamente nulas. Isto é, este é um cluster que é inteiramente caracterizado por pessoas que idolatram seus respectivos políticos.

Cabe dizer também que uma parte muito considerável do cluster apenas reproduzia frases, ou algo nas mesmas linhas, como: 'Bolsonaro Presidente, 'Bolsonaro Primeiro Turno' ou Lula Presidente, Lula Primeiro Turno'. Isto é, por muitas vezes, ainda que fosse bem explícito o posicionamento da pessoa, não havia detalhamentos além sobre a opinião da pessoa.

### **Cluster 3**

Este, é caracterizado principalmente pelo debate relacionado a temas de saúde. Visto que este foi um tema ainda muito latente, dado que as eleições se situavam em um momento pós-pandemia do COVID-19,

Em sua grande maioria, os tweets eram contrários a Bolsonaro indicando a forma como conduziu a pandemia no Brasil e como isso acabou potencializando o número de mortes. Assim, encontravam a alternativa na figura do Lula que, em seus discursos, repudiava a atitude do presidente anterior enquanto enaltecia, por exemplo, a necessidade de reforço do Sistema Única de Saúde (SUS).

### **CONCLUSÃO**

A partir do que foi disposto acima, a conclusão primária é que para os usuários, pouco importa o discurso econômico dos candidatos, isto é, se eles terão um viés mais liberal ou keynesiano, por exemplo. Mas sim, se o salário aumentar e o preço dos produtos diminuir, independente de como isso será feito, ou seja, o debate econômico no Twitter e, provavelmente, o voto dos usuários concentra-se especialmente dentro dos aspectos cotidianos da economia e pouco em qual vertente será adotada.

### **Referência Bibliográfica**

Gori Maia, A.; Martinez, J. D. M.; Marteleto, L. J.; Rodrigues, C. G.; Sereno, L. G. *Can the Content of Social Networks Explain Epidemic Outbreaks? Population Research and Policy Review*, v. 42, p. 9, 2023. Disponível em: <https://doi.org/10.1007/s11113-023-09753-7>. Acesso em: 09/11/2023.

do Amaral, O. E.; Ribeiro, P. F. *Por que Dilma de novo? Uma análise exploratória do Estudo Eleitoral Brasileiro de 2014. Revista de Sociologia e Política*, v. 23, n. 56, p. 107–123, 2015. Disponível em: <https://doi.org/10.1590/1678-987315235605>. Acesso em: 20/02/2024.

Pereira, F. B. *Voto econômico retrospectivo e sofisticação política na eleição presidencial de 2002. Revista de Sociologia e Política*, v. 22, n. 50, p. 149–174, 2014. Disponível em: <https://doi.org/10.1590/1678-987314225010>. Acesso em: 21/02/2024.

Statista. Social Media & User-Generated Content, 2022. Disponível em: [Brazil: Twitter users 2019-2028 | Statista](https://www.statista.com/pt-br/estatisticas/brasil/twitter-usuarios-2019-2028/). Acesso em: 11/03/2024.

GIL-CLAVEL, S.; MULDER, C. H. *Does Twitter data mirror the European North-South family ties divide? A comparative analysis of tweets about family*. Disponível em: <osf.io/c379h>. . Acesso em: 21/06/2024