

Political Forecasting White Paper

Heitor Ponchio, Joao Medeiros

June 2024

info@researchbluesky.com



Contents

Resumo Executivo	2
1 Introdução	2
2 Metodologia	2
3 Resultados	2
4 Conclusões e Recomendações	3
Referências	3

Resumo Executivo

O avanço das plataformas de mídia social, especialmente o Twitter, transformou significativamente o discurso político e os processos de tomada de decisão no século 21. Este estudo investiga como a análise de dados do Twitter pode prever o comportamento de votação dos congressistas brasileiros em propostas legislativas importantes. Focamos no período de 2019 a 2021, particularmente durante os debates sobre a reforma da previdência, para classificar e prever as posições dos políticos—favorável, neutra ou contra—usando análise de sentimentos e modelos probabilísticos.

1 Introdução

À medida que avançamos no século 21, a influência das plataformas de mídia social no discurso político e nos processos de tomada de decisão cresceu exponencialmente. Em particular, o Twitter tornou-se um espaço crucial para a comunicação política, onde políticos e o público expressam seus pensamentos, opiniões e reações a assuntos legislativos em andamento. Essa mudança digital é especialmente evidente no contexto das atividades parlamentares, onde a imediatividade e o alcance amplo das mídias sociais podem moldar os cenários políticos e potencialmente prever os resultados legislativos.

Neste estudo, investigamos esse fenômeno analisando dados do Twitter para prever o comportamento de votação dos congressistas brasileiros em propostas legislativas significativas. Utilizando análise de sentimentos e um modelo probabilístico, buscamos classificar as posições dos políticos—favorável, neutra ou contra um projeto de lei—e prever a probabilidade de sua aprovação. O período de 2019 a 2021 na política brasileira oferece um rico conjunto de dados, particularmente em relação à reforma da previdência, que foi um tema altamente debatido nesse período. Nossa metodologia envolve a coleta e análise de tweets de congressistas brasileiros, aplicando modelos estatísticos para interpretar os dados e fazer previsões informadas sobre os resultados legislativos.

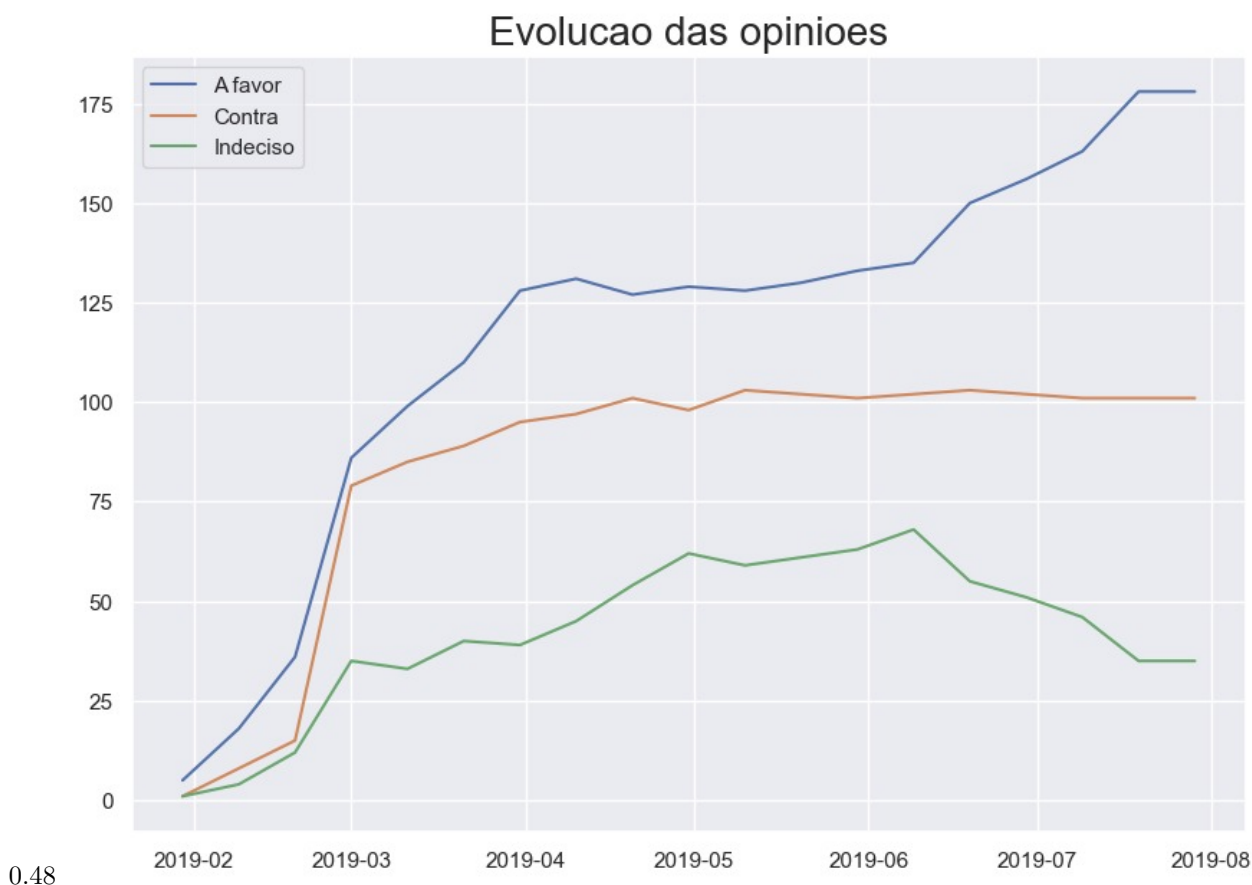
2 Metodologia

Nossa metodologia combina a coleta de dados do Twitter, a análise de sentimentos (NLP e Machine Learning) e a modelagem probabilística. Utilizamos técnicas de processamento de linguagem natural para extrair e categorizar sentimentos expressos nos tweets dos congressistas. Em seguida, aplicamos um modelo probabilístico para prever o comportamento de votação, considerando fatores como histórico de votos, afiliação partidária e engajamento nas redes sociais.

3 Resultados

Os resultados indicam uma correlação significativa entre o sentimento expresso nos tweets e o comportamento de votação dos congressistas. Visualizamos esses achados com gráficos que mostram a distribuição das posições (favorável, neutra, contra) e a probabilidade de aprovação dos projetos de lei. A análise detalhada dos debates sobre a reforma da previdência destaca como as mídias sociais podem antecipar tendências legislativas.

Conseguimos analisar a dinâmica das opiniões:



4 Conclusões e Recomendações

Nossas conclusões apontam que a análise de dados do Twitter pode ser uma ferramenta poderosa para prever resultados legislativos. As implicações práticas incluem a possibilidade de prever e influenciar decisões políticas com maior precisão. Recomendamos a integração de análises de mídias sociais nas estratégias de advocacy e na formulação de políticas públicas, aproveitando o potencial das redes sociais para compreender e antecipar movimentos políticos.

Referências

Andranik Tumasjan, Timm O. Sprenger, Philipp G. Sandner, Isabell M. Welp, **"Election Forecasts With Twitter: How 140 Characters Reflect the Political Landscape"** Social Science Computer Review, vol 29 no 4. 2010 — DOI: [10.1177/0894439310386557](https://doi.org/10.1177/0894439310386557)

Stefan Stieglitz, Christian Meske, Björn Ross, Milad Mirbabaie, **"Going Back in Time to Predict the Future - The Complex Role of the Data Collection Period in Social Media Analytics"**, Information Systems Frontiers - Springer - 2018 — <https://doi.org/10.1007/s10796-018-9867-2>

Stefan Stieglitz, Linh Dang-Xuan, **"Political Communication and Influence through Microblogging – An Empirical Analysis of Sentiment in Twitter Messages and Retweet Behavior"** 2012 45th Hawaii International Conference on System Sciences 2012 — DOI [10.1109/HICSS.2012.476](https://doi.org/10.1109/HICSS.2012.476)

Andrea Ceron, Luigi Curini, Stefano M Iacus and Giuseppe Porro, **"Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France"** New Media Society 16: 340, 2014 — DOI: [10.1177/1461444813480466](https://doi.org/10.1177/1461444813480466)

Frederick A. Hayek, **"The Use of Knowledge in society"** The American Economic Review 1945

Gregory D. Larosiliere, Lemuria D. Carter, Christian Meske **"How Does the World Connect? Exploring the Global Diffusion of Social Network Sites"** Journal of The Association For Information Science and Technology, 2017 — DOI: 10.1002/asi.23804

Gilad Mishne, Natalie Glance, **"Predicting Movie Sales from Blogger Sentiment "** American Association for Artificial Intelligence, 2006

Cass R. Sunstein, **Neither Hayek nor Habermas** Public Choice, Public Choice 134: 87–95, 2008 — DOI 10.1007/s11127-007-9202-9

Isaak, J., Hanna, M. J. **User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection.** Computer, 51(8), 56–59, 2008. — doi:10.1109/mc.2018.3191268

Kosinski, M., Matz, S. C., Gosling, S. D., Popov, V., Stillwell, D. **Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines,** American Psychologist, 70(6), 543–556. — DOI:10.1037/a0039210

<https://github.com/twintproject/twint>