ANNALS OF DATA ANALYSIS

OPINIÃO PÚBLICA E A FRAUDE DO INSS: UMA ANÁLISE DE SENTIMENTOS DE USUÁRIOS DO X/TWITTER SOBRE LULA E BOLSONARO

Carla Vitória Alves Cavalcante [1] * - https://orcid.org/0009-0003-7145-6160
Amanda Rodrigues Ferreira [2] - https://orcid.org/0009-0005-7873-0841
Rebeca Barbosa da Luz [3] - https://orcid.org/0009-0002-7691-6253
Gabriel Erik Moura de Albuquerque[4]- https://orcid.org/0009-0001-0114-9259
Euri Gurgel de Amorim Neto[5]- https://orcid.org/0009-0007-4046-1097
Maria do Carmo Soares de Lima[6]- https://orcid.org/0000-0002-5480-3103

[1]* Graduanda em Ciência Política pela Universidade Federal de Pernambuco
E-mail: carla.vitoria@ufpe.br

[2] Graduanda em Ciência Política pela Universidade Federal de Pernambuco
E-mail: amanda.rferreira@ufpe.br

[3] Graduanda em Ciência Política pela Universidade Federal de Pernambuco
E-mail: rebeca.bluz@ufpe.br

[4] Graduando em Ciência Política pela Universidade Federal de Pernambuco
E-mail: gabriel.ealbuquerque@ufpe.br

[5]Bacharel em Direito pela Universidade Federal do Rio Grande do Norte e Graduando em Ciência Política pela Universidade Federal de Pernambuco
E-mail: euri.neto@ufpe.br

[6] Doutora em Estatística pela Universidade Federal de Pernambuco e Professora Adjunta do Departamento de Ciência Política da mesma instituição
E-mail: maria@de.ufpe.br

RESEARCH ARTICLE

Annals of Data Analysis
https://periodicos.ufpe.br/revistas/ADA/

Submitted on: 12.12.2025.
Accepted on: 01.08.2026.
Published on: 01.08.2026.

Resumo

O estudo analisa a atribuição de responsabilidade pela fraude bilionária no INSS a partir de 4.748 postagens no X/Twitter. Foram comparadas menções ao presidente Luiz Inácio Lula da Silva e ao ex-presidente Jair Messias Bolsonaro, por meio de análise de sentimento em nível de entidade com o modelo BERTimbau. Os resultados revelam escrutínio assimétrico, com maior volume de menções negativas a Lula. O trabalho demonstra como métodos de redes neurais podem ampliar a compreensão da opinião pública em contextos políticos nas ciências sociais.

Palavras-chave

Análise de Sentimentos, Bolsonaro, Fraude, INSS, Lula, Opinião Pública.

1. Introdução

A operação "Sem Desconto", deflagrada no dia 23 de abril de 2025, e operacionalizada pela Polícia Federal (PF) e pela Controladoria-Geral da União (CGU), revelou uma fraude bilionária no Instituto Nacional do Seguro Social (INSS). A investigação identificou descontos mensais indevidos na folha de pagamento de milhões de aposentados e pensionistas. Essas cobranças, afirma a PF, aconteciam muitas vezes sem a autorização do beneficiário e por meio, dentre outras formas, de falsificação de assinaturas (Brasil, 2024). A "fraude do INSS", como ficou conhecida popularmente, apresenta um prejuízo estimado em mais de R$ 6 bilhões entre os anos de 2019 e 2024.

O atual caso de fraude do INSS, tomado como enfoque desta análise, não é um evento isolado no complexo sistema brasileiro de previdência social, atingido consideráveis vezes no curso da história da nação (Domingues Júnior, 2023). Esse evento, entretanto, é especialmente alarmante pelo volume de impactados e pela enorme difusão de opiniões *online* sobre o episódio.

A literatura sobre os casos de corrupção no sistema previdenciário brasileiro costuma focar no aspecto institucional e sistemático de tais eventos, bem como nos impactos jurídicos desses (Domingues Junior, 2023; Savazzoni, 2016). Apesar disso, o teor político, todavia, torna-se uma fração fundamental na narrativa da fraude deflagrada em 2025, e será o aspecto principal da análise desta pesquisa, assim como os possíveis efeitos sociais que derivam de escândalos de tal magnitude.

Em meio às apurações do caso, nota-se que os dois principais atores da política brasileira nos últimos anos tomaram os holofotes da discussão pública sobre a responsabilização pela fraude: o atual Presidente da República, Luiz Inácio Lula da Silva, e o ex-presidente Jair Messias Bolsonaro. Embora pouco medida pela literatura local, a polarização entre as duas "facções políticas" lideradas por esses agentes não é tão recente como o conhecimento público sobre o caso tratado neste artigo, e evidencia um embate de ideologias na sociedade brasileira (Alfaro, 2025).

É a partir desse evento disruptivo, porém, que a disputa entre esses dois polos é alavancada de maneira extrema na rede social X/Twitter, com usuários buscando realizar *accountability* e atribuir culpa aos indivíduos que têm, ou tiveram, o papel de regulamentar as instituições que falharam no controle da corrupção dessa instituição já no seu princípio.

O termo *accountability*, que não possui tradução literal para a língua portuguesa, é utilizado para descrever tal ato de responsabilização por se tratar de uma ferramenta de prestação de contas que pertence à sociedade civil sob um regime democrático, e ultrapassa um simples movimento jurídico e político de fiscalização e aplicação de sanções para agentes da administração pública, influenciando o direcionamento das perspectivas do eleitorado sobre eventos políticos (Campos, 1990; Nascimento, 2023). É o exercício do *accountability* que permite que a emissão da opinião pública em redes sociais se torne um mecanismo essencial para o mapeamento da culpabilização social desses indivíduos.

Nesse sentido, as mídias sociais demonstram ter uma participação considerável no impacto causado pela fraude, uma vez que amplificam esse contexto político polarizado (Machado;

Miskolci, 2019). É a atribuição popular dessa culpa difundida na internet que a presente pesquisa busca mensurar, com foco nos presidentes que atuavam no período marcado pela fraude, a fim de entender as narrativas derivadas do evento.

No cenário traçado pela amostra considerada por esse estudo, é importante considerar a animosidade entre os grupos de esquerda e direita mais engajados em atividades políticas, como afirmam Ortellado, Ribeiro e Zeine (2022), e notar a existência de preferências definidas internamente por esses espectros políticos na análise dos resultados adquiridos, que declaram opiniões políticas públicas em um recorte temporal específico e não fatos concretos, uma vez que a investigação ainda está em curso. Adicionalmente, embora não haja pesquisas desse tipo especificamente para o Brasil, deve-se considerar o viés político da plataforma, documentada na literatura como estimuladora de postagens com conteúdo conservador (Huszár; Ktena; O'Brien, 2021).

A fim de investigar a quem os usuários do X/Twitter atribuíram a culpa pela fraude do INSS, esta pesquisa empregou a Análise de Sentimento baseada em Entidade, em um *corpus* com 4.734 (quatro mil, setecentos e trinta e quatro) postagens, realizadas entre o dia 23 de abril e 12 de maio (dia que antecedeu o começo do mapeamento da escala da fraude, com envio de mensagens, pelo INSS, para os beneficiários).

O artigo está estruturado em outras cinco seções principais, para além da Introdução. A segunda Seção, intitulada "Cenário dos acontecimentos", tem como objetivo delinear o contexto político-institucional que possibilitou a ocorrência de práticas de corrupção no INSS. A Seção 3, "Impactos de escândalos políticos e a formação da opinião pública", analisará o referencial teórico da subárea de comportamento político na Ciência Política, com foco na compreensão dos efeitos provocados por escândalos e no entendimento da formação da opinião pública. A Seção 4, corresponde à "Metodologia", que se dedica a descrever o processo de coleta e elaboração do modelo utilizado na análise proposta neste trabalho. A Seção 5, "Resultados", apresentará os dados empíricos obtidos a partir da amostra selecionada, discutindo as implicações desses achados no contexto brasileiro. Por fim, a Seção 6 reunirá os principais achados da pesquisa, oferecendo uma síntese das análises desenvolvidas ao longo do trabalho e suas contribuições para o campo.

2. Cenário dos acontecimentos

O evento discutido no presente artigo decorre das revelações trazidas a público pela operação da Polícia Federal, cujo objetivo principal foi combater um esquema criminoso de alcance nacional responsável por realizar descontos indevidos nos benefícios de aposentados e pensionistas do INSS, conforme exposto na seção anterior.

Na operação autorizada pela Justiça Federal do Distrito Federal, foram cumpridos 211 mandados de busca e apreensão e 6 mandados de prisão temporária, havendo determinação de afastamento de servidores do INSS, inclusive do presidente da autarquia, que acabou por ser demitido na mesma data, e de um policial federal (Brasil, 2024).

Dias após a operação, ocorreu também a demissão do Ministro da Previdência, Carlos Lupi, Presidente do Partido Democrático Trabalhista (PDT), que fora substituído pelo Deputado Federal Wolney Queiroz, também integrante daquele partido (Brasil, 2024).

A investigação teve início em 2023, quando a CGU iniciou a apuração de um aumento expressivo no número de entidades que realizavam descontos associativos e nos valores subtraídos dos benefícios previdenciários. A partir dessa apuração inicial, foram realizadas auditorias em 29 entidades que mantinham Acordos de Cooperação Técnica (ACTs) com o INSS para a realização dos descontos. Além disso, no período entre 17/04/2024 e 04/07/2024, foram entrevistadas presencialmente pessoas que tinham descontos em seus benefícios.

As entrevistas realizadas com 1.273 beneficiários nas 27 unidades da federação foram um elemento central para a comprovação da fraude. Segundo informações disponibilizadas pela CGU, os entrevistados foram selecionados de forma aleatória, e divididos em dois grupos: o primeiro correspondia a 90 beneficiários em relação aos quais o órgão possuía documentação fornecida pelas próprias entidades associativas/sindicatos ao INSS; o segundo, totalizando 1.183 pessoas, contemplava beneficiários cujos descontos iniciaram em 2024, visando uma maior precisão nos dados coletados, haja vista o pouco tempo transcorrido.

Os resultados foram contundentes: 97,6% dos entrevistados (1.242 pessoas) afirmaram não ter autorizado os descontos, e 95,9% (1.221 pessoas) disseram sequer fazer parte de qualquer associação. Os achados acima demonstraram que, com alta probabilidade, os descontos ocorreram de forma indevida e contra a vontade dos beneficiários.

A análise da folha de pagamento dos benefícios do INSS, conhecida como "Maciça", também revelou o crescimento exponencial dos descontos e permitiu identificar as entidades mais problemáticas. Dezenove entidades concentravam 94,7% dos repasses feitos até maio de 2024.

A fraude explorava uma brecha legal que permite o desconto de mensalidades associativas diretamente na folha de pagamento dos beneficiários do INSS. A Lei nº 8.213, de 24 de julho de 1991, que dispõe sobre os Planos de Benefícios da Previdência Social, prevê em seu artigo 115, inciso V, a possibilidade de descontos nos benefícios para pagamento de mensalidades de associações e entidades de aposentados e pensionistas.

Para que esses descontos sejam legais, no entanto, é necessária a celebração de um Acordo de Cooperação Técnica (ACT) entre a entidade associativa e o INSS. Além disso, o benefício do aposentado ou pensionista precisa estar desbloqueado para esse tipo de desconto, e a entidade deve obter do beneficiário um termo de filiação e uma autorização expressa para a realização do desconto, devidamente assinados.

O que a investigação da CGU revelou foi um desvirtuamento completo desse mecanismo. A auditoria identificou um crescimento atípico e alarmante nos valores descontados nos últimos anos. Em 2021, os descontos somaram R$ 536,3 milhões, saltando para R$ 1,3 bilhão em 2023. A projeção para 2024, caso nada fosse feito, era de que esses descontos alcançariam a cifra de R$ 2,6 bilhões.

O esquema fraudulento se desdobrava em várias etapas, cada uma com seus riscos associados:

1. Celebração do ACT: O primeiro passo era a celebração do ACT com o INSS. A investigação apontou que muitos desses acordos foram firmados com entidades sem a devida capacidade operacional para atender aos seus supostos filiados.

2. Captação Indevida de Beneficiários: A etapa seguinte era a captação dos dados dos beneficiários. As entidades utilizavam diversas artimanhas para obter as informações e

as assinaturas dos aposentados e pensionistas. Em muitos casos, os beneficiários eram abordados com a promessa de vantagens, como acesso a planos de saúde, seguros de vida ou até mesmo a revisão de seus benefícios, e acabavam assinando documentos sem saber que estavam se filiando a uma associação e autorizando um desconto mensal em sua aposentadoria.

3. Envio de Listas Fraudulentas à Dataprev: De posse dos dados dos beneficiários, as entidades enviavam listas para a Dataprev, a empresa de tecnologia da Previdência Social, para a inclusão dos descontos na folha de pagamento. Essas listas, no entanto, continham nomes de pessoas que nunca haviam consentido com a filiação ou o desconto.

4. Desconto em Folha sem Autorização: Com base nas listas fraudulentas, os descontos eram efetuados diretamente nos benefícios, muitas vezes sem que o beneficiário sequer percebesse, dado o valor relativamente baixo de cada desconto individual. A fraude se tornava massiva ao ser aplicada a milhares de benefícios.

5. Repasse às Entidades: Por fim, o INSS repassava os valores descontados às entidades que se beneficiaram do esquema. A investigação da CGU apontou deficiências no processo de acompanhamento desses repasses por parte do INSS.

A vulnerabilidade dos beneficiários do INSS, em sua maioria idosos e pessoas com baixa escolaridade, era um fator crucial para o sucesso da fraude. Muitos tinham dificuldade em compreender os extratos de pagamento de seus benefícios, obtidos apenas por meio do aplicativo "Meu INSS", em identificar os descontos indevidos e em saber como proceder para solicitar o cancelamento.

Com isso, o Governo Federal, visando conter a crise e minimizar os danos políticos, passou a adotar medidas para a identificação dos descontos realizados indevidamente e efetuar a devolução completa dos valores aos beneficiários do INSS, sem depender das inúmeras entidades que realizavam os descontos, as quais serão cobradas por meios próprios.

3 Impacto de escândalos políticos e a formação de opinião pública

Compreender os efeitos que a perda de confiabilidade institucional proveniente de escândalos como o do INSS podem ter sobre a ordem política é um ponto fundamental, dado que a literatura tradicionalmente interpreta tais eventos como um dos fatores corrosivos da legitimidade institucional (Bowler, 2004). Sobretudo no cenário brasileiro, em que os índices de confiança no governo, por exemplo, são historicamente baixos, conforme atesta a série histórica do Latinobarómetro (2024).

Nesse sentido, autores como Maier (2011) argumentam ser possível avaliar as consequências dos escândalos quanto ao seu impacto na confiança institucional a partir de duas vias: a da lógica funcional, em que o impacto de tais eventos na credibilidade é positivo, uma vez que ações institucionais de contenção a esses acontecimentos podem atuar como reforço simbólico da legitimidade política; e da lógica disfuncional – com mais evidências empíricas – em que tais episódios impactam negativamente na confiança institucional na medida em que comprometem a integridade das instituições.

Assim, sob essa perspectiva disfuncional, tais eventos contribuem para fragilizar a relação entre cidadãos e Estado, causando uma queda do apoio difuso, – isto é, do apoio ao sistema político como um todo – um aumento do cinismo na política, e o desenvolvimento de mais uma arena discursiva atravessada pela polarização. Esse cinismo – ou seja, a descrença do eleitorado acerca da integridade dos políticos (Dancey, 2011) – pode ser notado nas reações de opositores, que interpretam o caso como um instrumento de ganho pessoal do atual presidente, evidenciando a crença de que políticos agem motivados por interesses pessoais.

Tal interpretação cínica do tecido social se manifesta, por exemplo, ao associarem uma possível omissão deliberada de Lula ou ao inferirem uma participação do presidente a partir do seu relacionamento com seu irmão, Frei Chico, um dos responsáveis por gerir o Sindicato Nacional dos Aposentados, Pensionistas e Idosos, sendo esse órgão um dos investigados na Operação Sem Desconto.

A queda do apoio difuso como produto disfuncional dos escândalos, por sua vez, torna-se evidente quando essa análise se estende às próprias estruturas do Estado, como a Previdência Social e os órgãos de controle, cuja atuação passa a ser questionada em ambas as gestões, seja em razão da suposta origem da fraude em uma gestão anterior, ou seja pela demora em sua identificação e resposta na gestão subsequente.

Além de impactar na confiança do governo, os escândalos políticos são mais um espaço de manifestação da opinião pública, cuja formação é igualmente relevante e complexa socialmente. Estudos demonstram que os cidadãos não reagem de modo uniforme perante denúncias de corrupção e má conduta política, variando de acordo com o alinhamento de regras morais, perspectiva de ganhos pessoais e predisposições ideológicas. (Botero et al., 2019; Fernández-Vázquez et al., 2015; Lee, 2015).

Esta dimensão é fundamental nesse contexto, uma vez que a formação de opinião pública não se constitui apenas por fatos, mas por um processo de raciocínio motivado (Lee, 2015), o qual, enquadrado em um cenário de polarização, será profundamente afetado pelos vieses partidários, influenciando diretamente no processo de interpretação (Bartels, 2002). Essas diferentes perspectivas são direcionadas e alinhadas a conclusões preferidas, diferentemente de uma motivação voltada à acurácia, que busca conclusões factuais. É dessa forma que eventos políticos tornam-se maleáveis e ganham múltiplas interpretações. No caso aqui estudado, ainda que não houvesse questionamento acerca dos malefícios do esquema, o raciocínio motivado do corpo social torna-se evidente no dissenso sobre a instauração da CPI, transformando o debate em um jogo de narrativas.

Tal comportamento é elucidado pela Teoria do Raciocínio Motivado, segundo a qual os indivíduos reinterpretam os escândalos conforme suas preferências (Nir, 2011), de forma que os apoiadores do acusado podem tender a ignorar ou a minimizar a gravidade da situação, enquanto os opositores tendem a reagir com maior reprovação (Anduiza et al., 2013). Nesse sentido, a polarização atua como um fenômeno que acentua o efeito do filtro cognitivo que molda o comportamento político, e consequentemente, a responsabilização. Entretanto, é de suma importância ressaltar que essa percepção assimétrica é acentuada nesse cenário de fraude, pois, o contexto brasileiro encontra-se afetivamente polarizado e atrelado a candidatos, com uma intensa hostilidade entre campos políticos rivais (Fuks; Marques, 2022).

Considerando isso, a polarização afetiva, enquanto um intensificador do enviesamento, traduz-se na utilização de manobras lógicas para ambos os lados, personalismo exacerbado e na

estigmatização do adversário, visível no uso de apelidos, no limiar de um antagonismo existencial, demonstrando como o afeto precede qualquer análise factual da realidade (Fuks; Marques, 2022). Em uma disputa ideológica entre entes polarizados, assim, defender a integridade de um dos agentes políticos é, consequentemente, questionar a do outro.

Somado a esse cenário de polarização, tal processo insere-se em uma configuração digital na qual as redes sociais tendem a potencializar esse fenômeno de radicalização afetiva. Isso ocorre porque os escândalos, na configuração atual, constituem dinâmicas comunicativas que ultrapassam a mera violação de normas, sendo caracterizados por debates intensos, amplificados pelas transformações trazidas pelas redes sociais e pelo contexto político brasileiro, que possui diversas singularidades, em razão de sua configuração partidária fragmentada. Esse cenário tem incitado questionamentos de base afetiva entre os segmentos de direita e esquerda nos últimos anos, oriundos do contraste entre governos ideologicamente divergentes (Ortellado, 2022).

Assim, nota-se que o julgamento político é fortemente influenciado por vieses ideológicos, evidenciando como a formação de opinião pública não é apenas um processo de resposta a informações, mas um mecanismo de filtragem ideológica e cognitiva contínua (Stanovich, 2021). Esse panorama revela-se essencial para a compreensão mais ampla do caso do INSS, tanto como um escândalo com consequências institucionais, quanto como um campo para desdobramento da polarização afetiva.

4. Metodologia

Esta seção tratará do processo de coleta, tratamento e processamento dos dados, considerando todas as minúcias metodológicas adotadas.

4.1. Coleta de dados

Esta Seção explora o processo de raspagem dos dados utilizados nesta análise. Inicialmente, foram coletadas 4.734 postagens na plataforma X, feitas entre os dias 23 de abril, data de deflagração da operação pela CGU e a PF, e 12 de maio de 2025. O Gráfico 4 aponta a distribuição de postagens coletadas a cada dia para cada termo de busca utilizado, para as entidades Bolsonaro e Lula. O período de 20 dias permite verificar a oscilação da percepção popular ao longo do avanço das investigações, bem como após a divulgação do vídeo do deputado Nikolas Ferreira (publicado no dia 7 de maio).

A primeira etapa da coleta utilizou o termo de busca “INSS Lula”, enquanto a segunda utilizou “INSS Bolsonaro”, termos que foram escolhidos por abrangerem postagens que falassem tanto do evento analisado (Fraude/Escândalo do INSS), como por incluir as entidades analisadas (Lula e Bolsonaro). Por outro lado, esses termos não restringem o *corpus* a postagens que usem termos específicos, como fraude ou escândalo, ou se refiram a uma mas não a outra entidade. Ao total, foram analisadas 3.214 postagens com o primeiro termo de busca citado e 1.520 postagens para o segundo.

A extração foi realizada via *web scraping*, em que as postagens foram obtidas via Nitter, uma aplicação *web* de código aberto que funciona como *front-end* do X/Tweet, que permite o acesso a conteúdos da plataforma sem login e autenticações. As postagens foram coletadas tendo como critério a data de publicação, sendo assim, uma média de 119 postagens foram extraídas para

cada dia (para cada termo de busca). A coleta foi feita a depender da disponibilidade das postagens e estabilidade da conexão, de maneira automatizada (0xFFFF08, 2025).

Vale salientar que a quantidade predominante de postagens acerca da entidade Lula presente na amostra analisada neste estudo é, em sua maioria, fruto da disponibilidade das publicações no X, uma vez que a ferramenta utilizada coleta as publicações encontradas a cada página carregada. Sendo assim, a coleta para cada entidade depende da quantidade de postagens disponíveis em cada página e não de uma explícita escolha dos autores. Tal cenário pode refletir um viés da própria plataforma que, após a aquisição por Elon Musk, passou a atrair um volume maior de usuários de direita (Mariani, 2024), o que possivelmente contribui para a maior quantidade de menções negativas ao presidente Lula encontradas na rede social, como será discutido na Seção 5.

Nesse sentido, o *corpus* final usado para análise do presente trabalho é a união das postagens obtidas na primeira e segunda etapas de coleta especificadas acima, resultando em um *corpus* com 4.734 observações. Dentre essas, 19,8% (941 postagens) foram usadas para o processo de *fine-tuning* do modelo para classificação. Essas observações foram rotuladas manualmente entre as classes positiva, neutra ou negativa, conforme explicado na Seção 4.2. Ademais, os outros 80,2% das postagens tiveram seus *labels* preditos pelo modelo em questão.

Dessa forma, dos 19,8% utilizados para a construção do modelo para inferência, 80% foram utilizados para treinamento do modelo e os outros 20% foram utilizados para a avaliação. Assim, de forma geral foram utilizadas para o treinamento do modelo 752 postagens, enquanto que para avaliação foram usadas 189 observações.

4.2. Análise descritiva dos dados

Os dados destinados ao treinamento do modelo foram manualmente rotulados. O *label* de cada observação foi adicionado levando em conta os seguintes critérios: o valor do *label* era determinado pelo sentimento relacionado à entidade, mesmo que ambas as entidades fossem citadas, o *label* assumiria o valor referente a uma só das entidades, identificadas em uma coluna à parte do banco de dados; por fim, observações escritas pelo *Grok*, inteligência artificial do X, foram rotuladas como neutras.

Para além dessas informações, os critérios utilizados para classificar as postagens entre positivo, neutro ou negativo foram:

- Positivo: postagens que afirmam que a entidade influenciou positivamente no combate à fraude, ou que negam a culpa da entidade;

- Neutro: postagens descritivas, que narram acontecimentos, contam fatos ou que foram redigidas pelo *Grok*;

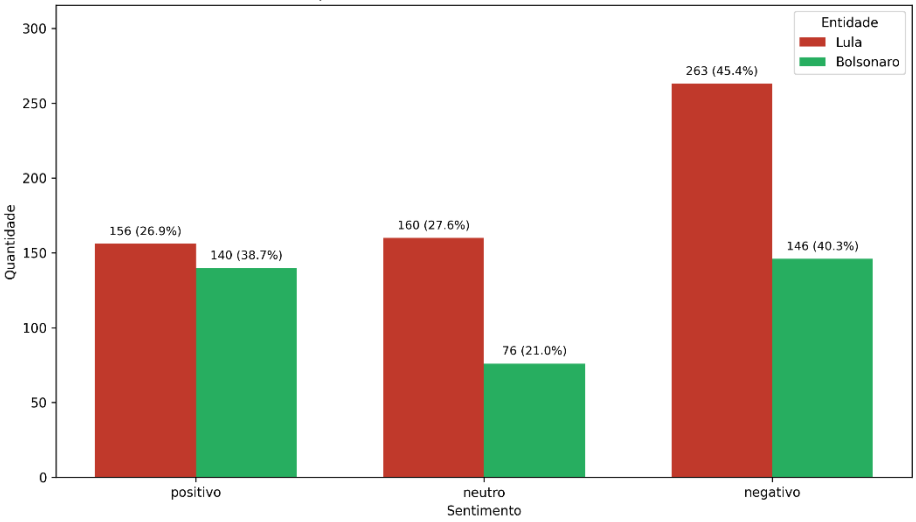
- Negativo: postagens que responsabilizam diretamente à entidade pela fraude.

Dessa forma, 19,8% das postagens totais foram rotuladas manualmente e usadas no processo de treinamento supervisionado. Esse modelo foi escolhido pela sua flexibilidade, já que é capaz de se adaptar ao conjunto de dados de treinamento utilizado, diferente de estratégias *lexicon-based* com dicionários mais generalistas, que poderiam não englobar estruturas de linguagens específicas do cenário político analisado neste estudo.

A distribuição da quantidade de *labels* para cada entidade analisada no *Corpus* de treinamento está exposta no Gráfico 1. Pode-se observar um certo desbalanceamento na distribuição das classes nos dados de treinamento, havendo, por exemplo, poucas observações negativas para a entidade "Bolsonaro" (146) em comparação com as observações negativas para a entidade "Lula" (263).

Gráfico 1

Distribuição de Sentimento por Entidade - Corpus de Treinamento



Nota. Elaboração própria (2025).

Buscando solucionar esse desbalanceamento, foi implantado um instrumento que penaliza erros de classificação que ocorrem em classes minoritárias, o *class weight*. Dessa forma foram atribuídos os respectivos pesos para cada classe analisada: para classe positiva, 1.060; para a neutra, 1.329 e para a negativa, 0.767. Sendo o peso w_k para cada classe k, os respectivos pesos representam a razão entre o número total de postagens no *corpus* de treinamento (N) e o produto do número de classes analisadas (K) e do número de postagens de cada classe analisada (n_k) (Scikit-learn, 2025).

Com isso, devido aos pesos incorporados à função de perda, erros nas classes positiva e neutra acarretam penalidades mais severas no processo de otimização da função de perda, uma vez que essas classes estão menos presentes no *corpus* de treinamento, em comparação com a classe negativa. Outra medida implantada foi utilizar a média harmônica entre *precision e recall* (F1-*Score*), como critérios para escolher o melhor modelo, e seus respectivos hiperparâmetros no processo de otimização, e não a acurácia, como normalmente ocorre, já que essa medida pode mascarar inconsistências de um modelo desbalanceado (Jurafsky; Martin, 2025). Essas modificações proporcionaram um melhor aprendizado para o modelo, mitigando classificações errôneas.

4.3. Elaboração do modelo

A Análise de Sentimento é uma tarefa fundamental de estruturas que lidam com o Processamento de Linguagem Natural, existindo diversos modelos que podem ser usados para isso, como *Naive Bayes, Maximum Entropy e Support Vector Machines* (Pang et al., 2002).

Para além disso, a análise ainda pode ser realizada em quatro níveis: *document level, sentence level*, *entity level e aspect-based* (Kumar; Sebastian, 2012).

Neste trabalho, utiliza-se uma análise em *word level,* parte da chamada *Targeted Sentiment Analysis* ou *Entity-Level Sentiment Analysis*. Esse tipo de análise consegue compreender o sentimento em relação a diferentes entidades presentes em uma mensagem. Dessa forma, esta pesquisa busca compreender o posicionamento dos brasileiros quanto a duas entidades, "Lula" e "Bolsonaro". Para isso, aquela análise é utilizada, pois não capta o sentido global do texto, mas o sentimento relacionado a cada entidade.

Ademais, pela análise em questão se tratar de publicações na rede social X, os dados utilizados nesta pesquisa possuem uma estrutura semântica própria, com gírias, ironia e trocadilhos particulares ao ambiente virtual. Para lidar com os desafios de analisar esse cenário, optou-se pela utilização de um modelo de aprendizado supervisionado, que utiliza a arquitetura *Transformers*, em vez de um modelo tradicional de *machine learning lexicon-based*, como alguns dos citados anteriormente. Aquela, é uma arquitetura de rede neural que possui um mecanismo denominado *self-attention*, que permite incorporar representações contextuais de cada *token*, o que possibilita que o processo de aprendizado integre informações de maneira adjacente (Jurafsky; Martin, 2025). Objetivamente, o *Transformers* viabiliza a análise de sentimento dado o contexto de cada sentença.

Para além disso, o modelo de aprendizado utilizado neste trabalho é o BERTimbau, uma versão do modelo BERT (Devin et al, 2018), tendo este sua arquitetura formada pelo *encoder* do *Transformers*. A principal contribuição existente em usar um modelo BERT é o seu mecanismo de processamento textual bidirecional, que considera tanto o contexto anterior, quanto o posterior a cada palavra. Por outro lado, diferente do modelo BERT tradicional, o modelo BERTimbau foi pré-treinado com textos e notícias já em português, o que permite uma aplicação direta ao *corpus* deste trabalho.

Indo mais a fundo, a aplicação de um modelo BERT possui duas partes fundamentais, o pré-treinamento e o *fine-tuning*, que consiste no ajuste dos parâmetros usados no modelo com base nos dados rotulados para tarefas específicas, no caso desse trabalho, para análise em *word level* (Devlin et al., 2019). Esse ajuste é feito pela adição da camada de saída (*head*) que será responsável pelo ajuste dos pesos e predição dos *logits*. Foi esse o processo realizado neste trabalho, um *fine-tuning* que preparou o modelo — já pré-treinado — para inferir as classes de aproximadamente 80% (oitenta por cento) do *corpus* coletado.

Por fim, o modelo final alcançou as seguintes métricas de desempenho: um F1-*Score* igual a 0.59, valor que, embora indique espaço para melhorias, demonstra a capacidade do modelo em lidar com classes desbalanceadas; quanto à acurácia, ficou igual a 0.60, o que é consistente com um modelo que aprende significativamente bem ao longo do processo. Para uma análise mais detalhada é possível observar a matriz de confusão do modelo em questão no Gráfico 2.

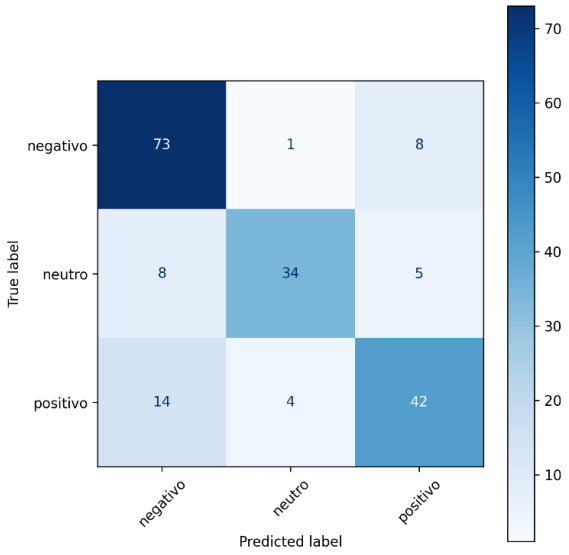
Para obter esses resultados, foi empregado um mecanismo de otimização de hiperparâmetros com *Optuna*, a fim de maximizar o aprendizado do modelo de maneira eficiente, para isso os intervalos passados para busca dos hiperparâmetros ótimos visaram ser compatíveis com o tamanho do *corpus*, com a tarefa de inferir classes desbalanceadas e com as restrições de processamento, além de limitar o desperdício de tentativas. Assim, o intervalo de busca da taxa de aprendizagem foi estabelecido entre 10-5e 410-5. Por outro lado, o *batch size* foi avaliado

com os seguintes valores: 8, 16 e 32. Por fim, o *weight decay* ficou estabelecido entre 0.01 e 0.1. Com isso, os melhores parâmetros encontrados foram os seguintes:

- Taxa de aprendizagem : 2.32e-05
- *batch size*: 8
- *weight decay:* 0.08

Gráfico 2

Matriz de confusão



Nota. Elaboração Própria (2025).

Assim, o treinamento foi realizado com o tokenizer do BERTimbau base, com truncamento de 512 tokens, *padding* até o comprimento máximo. Foi ainda utilizado o método *Early Stopping*, para que o treinamento findasse se a perda começasse a crescer. Para fins de replicabilidade da pesquisa, utiliza-se somente a *seed* 42 e o número total de épocas foi de 5.

Dessa forma, a metodologia adotada permitiu desenvolver um modelo de classificação robusto, que se adequa às necessidades desta pesquisa, principalmente ao contexto discursivo da plataforma X e à tarefa de classificação de sentimentos direcionada às entidades políticas, com base em dados atuais. A seguir, apresentam-se os resultados obtidos com a aplicação do modelo ao *corpus* coletado.

5. Resultados

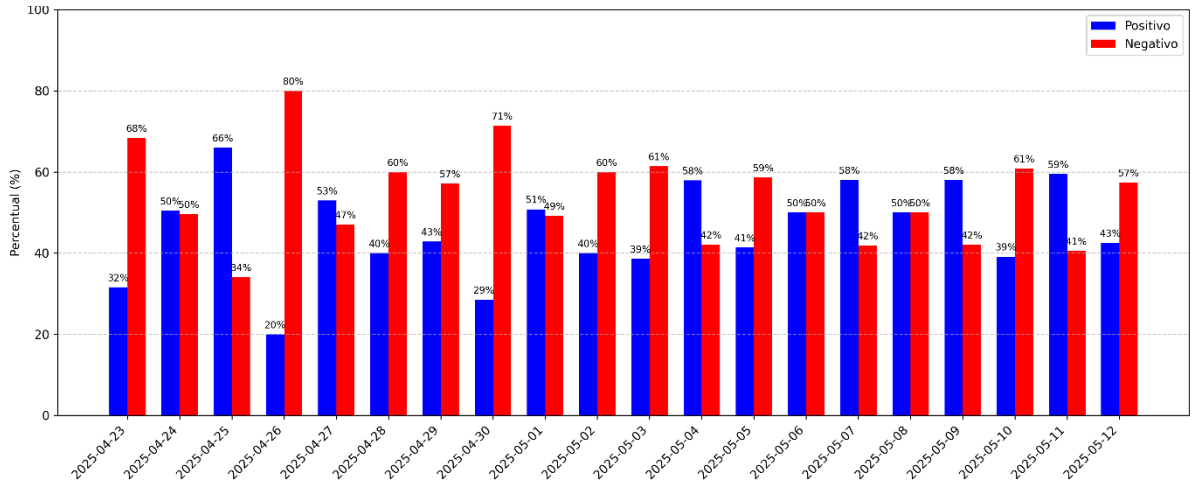
A análise que se segue correlaciona os principais eventos do escândalo com a intensidade da conversa *online*, medida pelo volume diário de postagens e os sentimentos expressos em relação a cada uma das entidades estudadas. Dessa forma, é fundamental ressaltar que as conclusões são inerentes a esta amostra específica, e que a distribuição não linear de postagens ao longo dos dias (Gráfico 4) é um fator que pode impactar a interpretação dos picos de

sentimento, assim como outras limitações metodológicas contempladas na conclusão deste artigo.

Inicialmente, a análise dos sentimentos sobre o ex-presidente Bolsonaro revela um padrão de respostas com perfil contestatório, mas que se manifesta por dinâmicas distintas ao longo do período. Em primeiro momento, o pico de negatividade observado no dia 26 de abril (80%) deve ser interpretado com cautela, configurando-se majoritariamente como um efeito residual da operação deflagrada três dias antes e estatisticamente inflacionado pelo baixo volume de postagens (29) naquela data.

Gráfico 3

Frequência de sentimento diário – Entidade Bolsonaro



Nota. Elaboração Própria (2025).

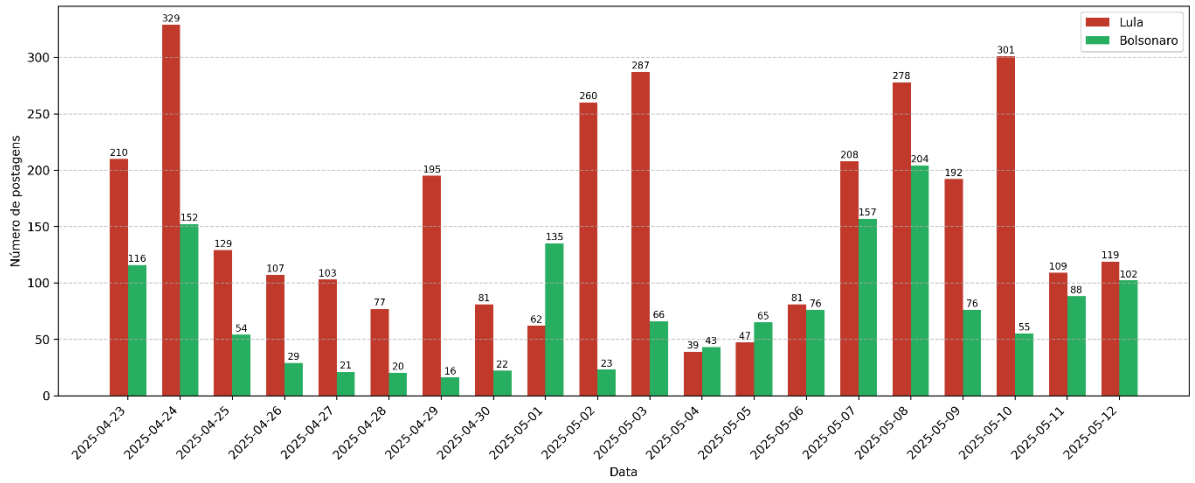
Em contrapartida, o pico do dia 30 de abril (71%) reflete uma reação direta à divulgação do relatório da Polícia Federal, ocorrida no dia anterior (29). Diferente do dia 26, neste momento a contestação é fundamentada em fatos novos, como evidenciado por postagens que utilizam os dados revelados para atacar a gestão passada: “vai falar do principal citado no relatório da pf? o ex-ministro da previdência e do inss do governo de jair messias bolsonaro, há indícios gravíssimos de que suas empresas e sócios receberem dinheiro desviado.” Esse exemplo ilustra a natureza do debate, indicando que a negatividade, neste segundo momento, é impulsionada pela apropriação das evidências oficiais para contestar a integridade da figura do ex-presidente.

Ademais, no dia da publicação do vídeo do deputado Nikolas Ferreira (7 de maio), no qual o político endossa a narrativa utilizada para responsabilizar a entidade Lula, observa-se que as menções positivas para a entidade Bolsonaro (58%) superaram as negativas (42%), contrariando a tendência de equilíbrio vista no dia anterior. Esse dado sugere que a mobilização digital aliada demonstrou força suficiente para disputar a narrativa naquele momento, sobrepondo-se até mesmo à repercussão da agenda oficial do governo Lula sobre o ressarcimento das vítimas. Contudo, a análise dos dias subsequentes revela uma instabilidade: o retorno ao empate técnico no dia 8 e a queda de popularidade no dia 10 indicam que o efeito do vídeo foi pontual, incapaz de blindar a imagem do ex-presidente de forma duradoura.

Para além disso, o Gráfico 4 aponta a forte oscilação na quantidade de postagens coletadas ao longo dos vinte dias analisados nesta pesquisa. Por isso, as justificativas usadas para explicar cada um dos picos do Gráfico 3 precisam levar em conta a quantidade de postagens disponíveis. Por exemplo, o pico de menções negativas atribuídas a Bolsonaro pode existir em decorrência da divulgação do relatório da PF, ou simplesmente porque a amostra coletada está abaixo da média, uma vez que os dias 29 e 30 de abril somaram apenas 16 e 22 postagens, respectivamente. Esse baixo volume impede a generalização dos resultados e deve ser levado em conta na interpretação das conclusões apresentadas ao longo do texto.

Gráfico 4

Distribuição de postagens por dia – Lula x Bolsonaro



Nota. Elaboração Própria (2025).

Em contraste, a análise dos dados referentes ao presidente Lula aponta como o governo foi massivamente atingido pelo escândalo do INSS. Desde a deflagração da operação, as alusões ao Presidente Lula são preponderantemente negativas, com exceção para o dia 7 de maio, em que a quantidade de postagens com sentimentos positivos supera a com sentimento negativo em cerca de 32 pontos percentuais (Gráfico 5). Esse pico, que rompe a tendência crítica, coincide com a ampla cobertura midiática das reuniões ministeriais ocorridas nos dias 6 e 7 de maio, nas quais foi definido o plano de ressarcimento aos pensionistas/aposentados lesados. Esse movimento sugere que a sinalização de uma solução institucional rápida gerou uma percepção popular favorável e imediata, antecedendo inclusive o anúncio oficial, por meio da coletiva de imprensa no dia 8, e de uma peça publicitária do governo, no dia 11.

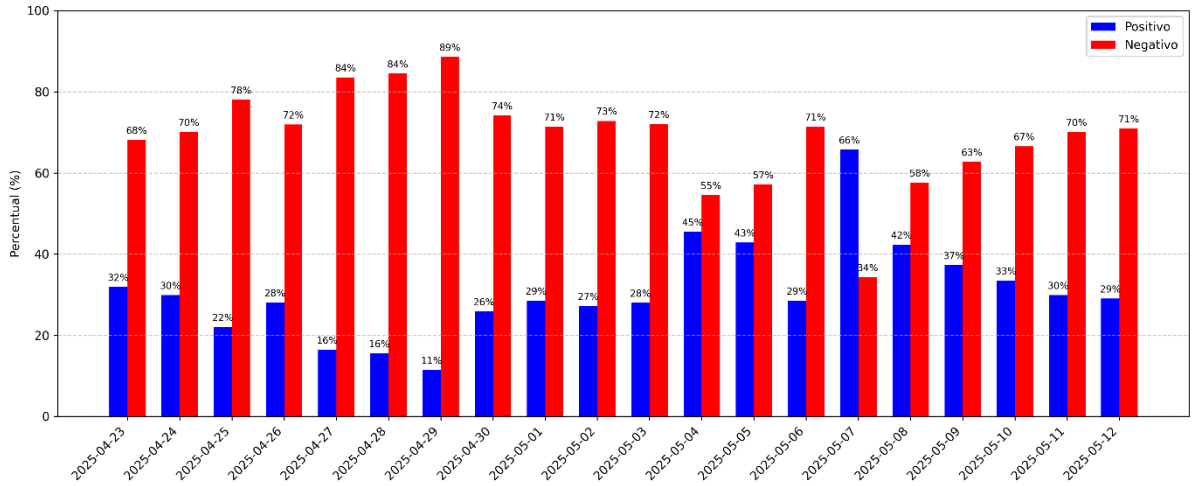
Observa-se, portanto, a predominância de uma narrativa que atribui a Lula a responsabilidade pela fraude do INSS no X (antigo Twitter). Esse discurso se manifesta tanto na associação do presidente a figuras como Frei Chico e o ministro Carlos Lupi, quanto na ênfase ao crescimento do número de beneficiários lesados em sua gestão, dado corroborado pelo relatório da CGU.

Além dos pontos mencionados, a narrativa que responsabiliza o presidente é muitas vezes construída sobre o argumento de que ele teria revogado a Lei Antifraude (Lei 13.846/2019), o que teria viabilizado os crimes; embora, na realidade, apenas o acesso do INSS aos dados da Receita Federal tenha sido vetado. A narrativa também se apoia em supostas omissões,

sugeridas pela resistência da base governista à criação de uma CPI, e na insistente vinculação do mandatário a um histórico de corrupção.

Gráfico 5

Frequência do sentimento por dia - Lula



Nota. Elaboração Própria (2025).

Ademais, ao realizar uma análise textual simples via TF-IDF (*Term Frequency–Inverse Document Frequency*), visando identificar as 15 palavras mais importantes no *corpus* analisado para cada entidade — como pode ser visualizado no Gráfico 6 —, percebe-se que tais palavras desenham os contornos das estratégias de responsabilização e defesa que dão tom à conversa no X. Embora o termo “governo” ocupe o topo da relevância para ambas as entidades, a observação dos vocábulos subsequentes permite identificar dinâmicas de associação distintas no caso de Lula. Nota-se uma investida discursiva que busca ultrapassar o campo institucional para atingir o âmbito pessoal, como sugere a forte importância do termo "irmão". Essa conexão, que remete à figura de Frei Chico, demonstra um gatilho para associar o escândalo atual a um histórico familiar e de corrupção, transformando uma falha de fiscalização em uma narrativa de proximidade pessoal. Em contrapartida, o surgimento do termo "salvou" sinaliza a reação da base governista, que tenta inverter a polaridade do debate ao posicionar o presidente como o agente responsável por expor o esquema e iniciar o ressarcimento das vítimas.

Essa dinâmica de personificação contrasta com o cenário observado nas menções ao ex-presidente Bolsonaro, onde a disputa se desloca para uma arena predominantemente temporal e legislativa. A proeminência de termos como "2019", "começou" e "lei" revela o esforço em situar a gênese da fraude no período de sua gestão, utilizando a cronologia como principal ferramenta de transferência de culpa. Ao mesmo tempo, o destaque para "PT", "esquerda" e "CPI" indica que as citações a Bolsonaro funcionam muitas vezes como um mecanismo de contra-ataque ideológico; aqui, a narrativa de omissão do atual governo ganha força por meio da resistência à abertura de comissões parlamentares, como citado anteriormente, alimentando a percepção de um viés partidário na condução do caso.

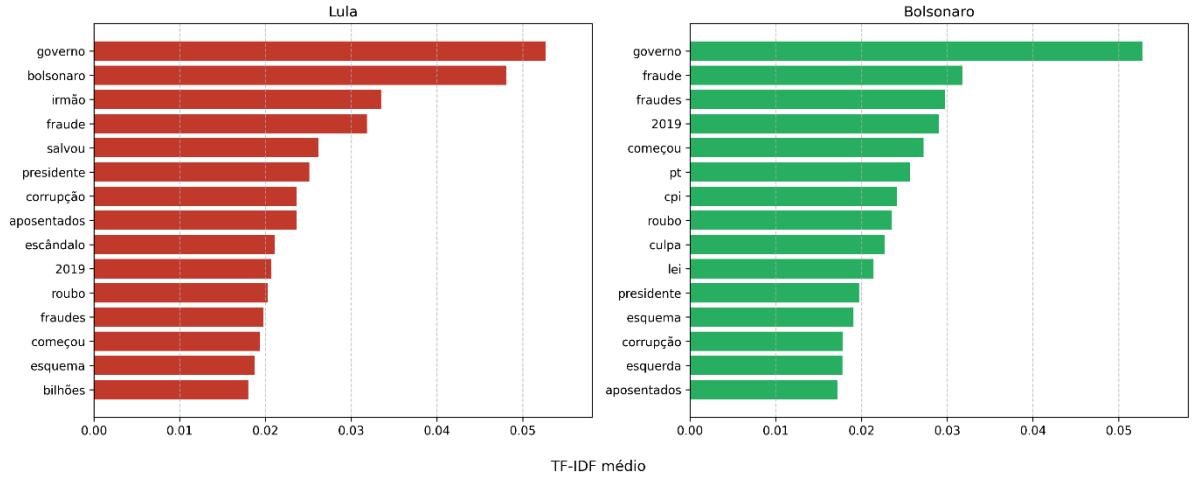
Por conseguinte, vale pontuar que, mesmo sem a aplicação de um tratamento algorítmico específico para detecção de ironia, a análise qualitativa revelou a capacidade surpreendente do

modelo em lidar com esse fenômeno, captando a inversão semântica típica do discurso político. Teoricamente, a ironia exige que o classificador detecte a dissonância entre o sentido literal e a intenção pragmática, o que induz métodos *lexicon-based* ao erro, uma vez que as palavras são aplicadas com o intuito oposto ao seu sentido literal.

Contudo, duas razões podem elucidar o satisfatório desempenho com ironia alcançado pelo modelo. Em primeiro lugar, o grande número de postagens com teor irônico no *corpus* de treinamento contribuiu para o aprendizado do modelo de padrões semânticos típicos das postagens irônicas. Como também, o mecanismo de *self-attention* da arquitetura *Transformers*, que permite a incorporação de informações contextuais por meio de uma interpretação relacional e não isolada dos *tokens*.

Gráfico 6

Palavras mais relevantes por entidade (TF-IDF)



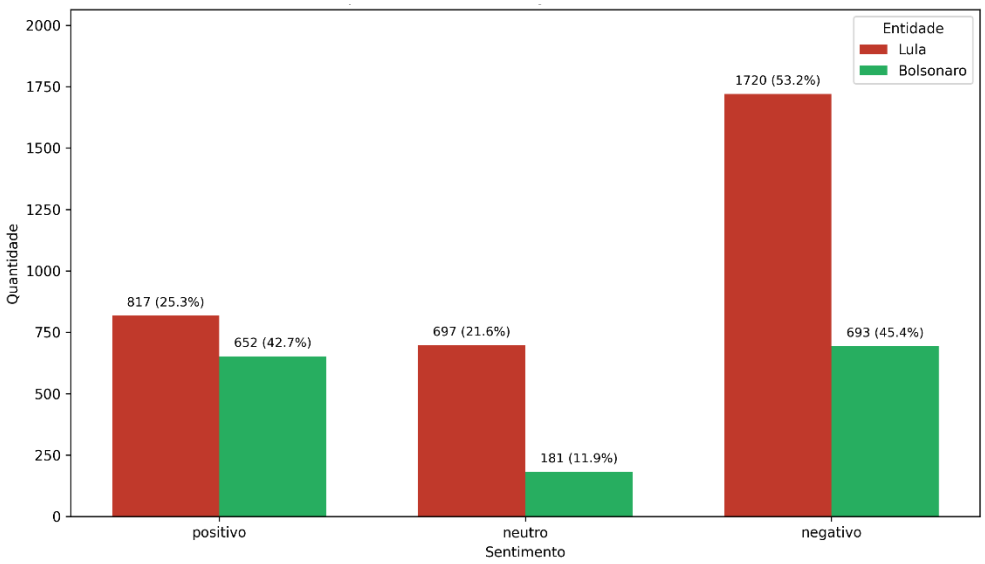
Nota. Elaboração Própria (2025).

Como exemplo, a publicação “lula quer ressarcir roubos do inss com dinheiro nosso... me digam, se esse cara não é um gênio! kkkkk” foi classificado corretamente como negativo, mesmo contendo a palavra “gênio”. O modelo entendeu que o contexto transformava o elogio em crítica. Da mesma forma, houve acerto na classificação da frase “É não é q a farra do INSS começou no governo do impoluto e super honesto Bolsonaro, lá em 2019. Pois é.”, onde o modelo detectou que os adjetivos de honestidade foram usados ironicamente para responsabilizar o ex-presidente. Esses resultados indicam que o modelo conseguiu ir além das palavras isoladas, captando nuances do contexto pragmático da mensagem.

Na sequência, o Gráfico 7 demonstra como o presidente Lula foi fortemente responsabilizado pelo escândalo do INSS, sendo cerca de 10% mais culpado – por usuários do X – que o ex-presidente Bolsonaro pela fraude. Por outro lado, Jair Bolsonaro obteve algo em torno de 18% a mais de menções positivas, quando comparado a seu opositor político. Esses resultados possivelmente estão relacionados tanto ao maior engajamento de usuários apoiadores de Bolsonaro nas redes sociais, acentuado pelo manuseio político das redes sociais por políticos da base de direita, o que ocasiona uma participação mais engajada em debates nas redes sociais, quanto à decrescente avaliação do governo Lula (Quaest, 2025).

Gráfico 7

Comparativo da distribuição final de sentimentos



Nota. Elaboração Própria (2025).

Conclui-se nessa seção que, embora o escândalo tenha servido como um catalisador para a polarização afetiva no país, o escrutínio público não foi distribuído de forma simétrica. Como governante em exercício, Lula tornou-se o foco do debate, atraindo um volume de menções negativas quase três vezes superior ao do seu antecessor. Assim, apesar de a base governista ter usado a cronologia da fraude para criticar a gestão anterior, a análise dos dados sugere que a maior carga de responsabilização pela gestão da crise e pela vulnerabilidade das instituições recaiu sobre a figura do atual presidente.

6. Conclusão

A divisão da sociedade brasileira em dois núcleos políticos centrais, analisada no artigo, é notada particularmente a partir da eleição de Jair Messias Bolsonaro em 2018, quando ocorre um aumento na escala de posicionamento ideológico, associado a um salto na associação entre ideologia e voto (Fuks; Marques, 2020). Isso, dentre outros fatores explicitados no corpo da pesquisa, é o bastante para considerar a existência de uma polarização das opiniões políticas de massa no país, assim como a presença de uma polarização identitária entre esquerda e direita, como explicitado por Ortellado, Ribeiro e Zeine (2022), principalmente após a mudança de polo político vigente com a eleição de Luís Inácio Lula da Silva em 2022.

Diante da operação para apurar a fraude do INSS, essa divergência entre polos se torna bastante notável. Os seguidores dos dois agentes políticos analisados aqui — Lula e Bolsonaro — demonstram certo teor de polarização afetiva em relação a essas lideranças políticas, algo também percebido em um momento de fragilidade partidária em 2018 (Fuks; Marques, 2022). Esse afeto com a persona política demonstra impactar de forma considerável a percepção do eleitorado sobre os fatos disponíveis sobre a fraude e, consequentemente, sobre quem é responsabilizado por ela.

Nota-se que a defesa do eleitorado de direita ou dos eleitores abertamente apoiadores do Bolsonaro para com ele é manifestado - em grande parte das postagens negativas feitas por esse grupo sobre Lula na amostra aqui considerada - por meio de uma clara desafeição com o Partido dos Trabalhadores (PT), como corroborado no Gráfico 6 pela importância do termo "pt" no *corpus* analisado da endidade Bolsonaro, como também pela analise qualitativa realizada nesse mesmo conjunto de postagens. Esse antipetismo utilizado pela direita estava igualmente presente em 2018, denotando o crescimento desse polo no país (Fuks; Marques, 2022).

A defesa do eleitorado do PT, ou dos eleitores abertamente apoiadores do Lula, por outro lado, direciona-se ao ex-presidente Jair Bolsonaro especificamente, e não a um partido político, e busca alavancar as ações de mitigação realizadas pelo governo atual.

A entidade Lula ter sido mais responsabilizada pela fraude do INSS pode ser explicado por uma gama de fatores, dentre esses: o conhecimento público do caso em seu mandato; o impacto de outras decisões políticas que levaram à queda em popularidade antes da divulgação da fraude; o crescente bolsonarismo nas redes sociais (Amaral; Silveira, 2023); o viés político da plataforma analisada; e, por último, a presença de uma polarização afetiva na sociedade brasileira, discutido no corpo deste trabalho.

A responsabilização de Bolsonaro – mesmo que menor – também denota um cenário alarmante de desconfiança em instituições de controle, com usuários questionando como tal fraude não pode ter sido percebida antes.

Vale ainda ressaltar as implicações de um *F1-Score* modesto – de 0.59 – nos resultados vislumbrados no presente trabalho. Os dados analisados tendem apontar que Lula possui um maior número de menções negativas, quando comparado à entidade Bolsonaro. No entanto, essa conclusão precisa ser ponderada nas métricas de avaliação do modelo utilizado para classificação das postagens analisadas, que foram abaixo dos níveis esperados. Os valores insuficientemente adequados revelam a fragilidade do modelo em aprender alguns padrões textuais utilizados, resultando em erros de omissão e de atribuição indevida (falsos negativos e falsos positivos).

Por exemplo, apesar do modelo classificar corretamente frases como "lula não sabia de nada, mas bolsonaro sabia que os sindicatos ligados ao pt estavam mamando tudo no inss. conta outra "amante"" (postagem classificada como negativa para entidade Lula); ele ainda pode falhar em identificar padrões como o exposto na frase "ai, e vocês estão muito enganados. a coisa ficou bem feia durante o governo de jair bolsonaro. é fácil culpar o lula pq o ministro colega de seu partido, duda! bolsonaro roubou o inss" (classificada como negativa para entidade Lula), o que pode acontecer, devido a erros gramaticais, tamanho do *corpus* de treinamento reduzido, ou por outros fatores. Esse cenário revela que o número de menções negativas para a entidade Lula pode, em certa medida, estar inflado devido aos falsos negativos classificados erroneamente pelo modelo.

6.1 Limitações do modelo

O modelo desenvolvido neste estudo apresenta algumas limitações procedimentais que merecem menção e que serão discutidas nesta subseção.

As alterações nas políticas de acesso e moderação de conteúdo sofridas pelo X/Twitter com a compra do aplicativo por Elon Musk foi uma das principais deficiências encontradas na coleta de postagens, dificultando o tamanho da amostra analisada. A partir disso, foi gerado um desbalanceamento de informações, devido à falta de padronização na quantidade de postagens por dia, o que pode impactar a generalização dos resultados. O uso de termos de pesquisa alternativos não contemplados nesta coleta também é uma variável a ser considerada.

Por outro lado, não houve nenhum tratamento expresso para publicações realizadas pelo Grok, Inteligência Artificial do X/Twitter, ou para publicações realizadas pelo mesmo usuário. Sendo assim, as postagens realizadas pelo Grok foram meramente rotuladas como neutras no *corpus* de treinamento, enquanto que o *corpus* geral pode possuir mais de uma publicação realizada por um mesmo usuário em dias diferentes, ou até no mesmo dia. Tal fator pode impactar a interpretação dos resultados discutidos neste estudo ao limitarem a abrangência de opiniões individuais analisadas. Portanto, esses pontos são especialmente relevantes, pois impactam diretamente na interpretação das postagens e, consequentemente, nas conclusões derivadas delas.

Para pesquisas futuras, é fundamental considerar as limitações do modelo adotado, bem como os vieses inerentes às pesquisas de opinião pública. Recomenda-se, igualmente, a utilização conjunta de diferentes redes sociais como fontes de dados, a fim de ampliar a diversidade do público analisado e obter uma visão mais representativa das percepções sociais. Essa abordagem pode enriquecer a compreensão sobre a opinião pública em torno de temas complexos, como a fraude do INSS.

Por fim, neste estudo, percebe-se a presença de polarização afetiva no Brasil, e entende-se os possíveis impactos da fraude do INSS à legitimidade e à confiança institucional. Os resultados da pesquisa evidenciam o embate de narrativas entre dois crescentes grupos políticos nas redes sociais, e demonstram o impacto da percepção pública seletiva sobre fatos para uma tomada de decisão sobre responsabilidade, acima de tudo, com informações que permeiam vias ideológicas personalizadas.

Referências

1. 0xFFFF08. (2025). pyscraper [Código-fonte]. GitHub. https://github.com/0xffff08/pyscraper
2. Alfaro, F. G. G. de. (2025). Polarização política no Brasil e a influência (direta) das mídias sociais. *Derecho y Cambio Social*, 22(79). https://doi.org/10.54899/dcs.v22i79.156
3. Amaral, A., & Silveira, F. (2023). Bolsonarismo e o fascismo na era digital. Revista Brasileira de Estudos Políticos, (127), 55–94.
4. Anduiza, E., Gallego, A., & Muñoz, J. (2013). Turning a blind eye: Experimental evidence of partisan bias in attitudes toward corruption. *Comparative Political Studies*, 46(12), 1664–1692. https://doi.org/10.1177/0010414013489081
5. Bartels, L. M. (2002). Beyond the running tally: Partisan bias in political perceptions. *Political Behavior*, 24(2), 117–150.

6. Botero, S., et al. (2019). Are all types of wrongdoing created equal in the eyes of voters? *Journal of Elections, Public Opinion and Parties*. https://doi.org/10.1080/17457289.2019.1651322

7. Bowler, S., & Karp, J. A. (2004). Politicians, scandals, and trust in government. *Political Behavior*, 26(3), 271–287. https://doi.org/10.1023/B:POBE.0000043456.87303.3a

8. Brasil. Controladoria-Geral da União. (2024). Relatório de avaliação: Descontos de mensalidades associativas – INSS: Exercícios de 2023 e 2024 (77 p.). CGU.

9. Campos, A. M. (1990). Accountability: Quando devemos traduzi-la para o português? Revista de Administração Pública, 24(2).

10. Dancey, L. (2011). The consequences of political cynicism: How cynicism shapes citizens' reactions to political scandals. *Political Behavior*, 34(3), 411–423. https://doi.org/10.1007/s11109-011-9163-z

11. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 *Conference of the North American Chapter of the Association for Computational Linguistics*. https://aclanthology.org/N19-1423/

12. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv. https://doi.org/10.48550/arXiv.1810.04805

13. Domingues Júnior, A. C. (2023). A incidência criminosa nas fraudes à previdência social e sua persecução penal (Trabalho de Conclusão de Curso, Bacharelado em Direito). Unifucamp. http://repositorio.fucamp.com.br/jspui/handle/FUCAMP/683

14. Espinoza, R. M. (2012). Dicionário de políticas públicas. In C. L. F. de Castro, C. R. B. Gontijo, & A. E. N. Amabile (Orgs.). EdUEMG.

15. Fernández-Vázquez, P., et al. (2015). Rooting out corruption or rooting for corruption? The heterogeneous electoral consequences of scandals. *Political Science Research and Methods*, 4(2), 379–397.

16. Fuks, M., & Marques, P. H. (2020). Contexto e voto: O impacto da reorganização da direita sobre a consistência ideológica do voto nas eleições de 2018. Opinião Pública, 26(3), 401–430. https://doi.org/10.1590/1807-01912020263401

17. Fuks, M., & Marques, P. H. (2022). Polarização e contexto: Medindo e explicando a polarização política no Brasil. Opinião Pública, 28(3), 560–593. https://doi.org/10.1590/1807-01912022283560

18. Huszár, F., Ktena, S., & O'Brien, C. (2021). Algorithmic amplification of politics on Twitter. *Proceedings of the National Academy of Sciences*.

19. Jurafsky, D., & Martin, J. H. (2025). Speech and language processing (3rd ed.). Stanford University. https://web.stanford.edu/~jurafsky/slp3/

20. Kumar, A., & Sebastian, T. M. (2012). Sentiment analysis: A perspective on its past, present and future. *International Journal of Intelligent Systems and Applications*. https://doi.org/10.5815/ijisa.2012.10.01

21. Latinobarómetro. (2024). Online Data Analysis (ODA-JDS). *Corporación Latinobarómetro*. https://www.latinobarometro.org/odajds/

22. Lee, F. L. F. (2014). How citizens react to political scandals surrounding government leaders: A survey study in Hong Kong. *Asian Journal of Political Science*, 23(1), 44–62.

23. Machado, J., & Miskolci, R. (2019). Das jornadas de junho à cruzada moral: O papel das redes sociais na polarização política brasileira. Sociologia & Antropologia. https://doi.org/10.1590/2238-38752019v9310

24. Maier, J. (2011). The impact of political scandals on political support: An experimental test of two theories. International Political Science Review, 32(3), 283–302. https://doi.org/10.1177/0192512110378056

25. Mariani, D. (2024, 13 de novembro). Contas de direita crescem três vezes mais que as de esquerda no X de Musk. Folha de S.Paulo.

26. Nascimento, P. (2023). Accountability: Um debate inicial. In O conceito de accountability na ciência política brasileira (cap. 2). EDUEPB.

27. Nir, L. (2011). Motivated reasoning and public opinion perception. Public Opinion Quarterly, 75(3), 504–532. https://doi.org/10.1093/poq/nfq076

28. Ortellado, P., Ribeiro, M. M., & Zeine, L. (2022). Existe polarização política no Brasil? Opinião Pública, 28(1). https://doi.org/10.1590/1807-0191202228162

29. Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. In Proceedings of the 2002 *Conference on Empirical Methods in Natural Language Processing (pp. 79–86). Association for Computational Linguistics.* https://aclanthology.org/W02-1011/

30. Quaest. (2025). *Pesquisa nacional – Avaliação do governo Lula*. Quaest.

31. Savazzoni, S. de A. (2016). Crimes contra a previdência social. Revista do Tribunal Regional Federal da Terceira Região, 27(129).

32. Scikit-learn. (s.d.). *sklearn.utils.class_weight.compute_class_weight*. https://scikit-learn.org/stable/modules/generated/sklearn.utils.class_weight.compute_class_weight.html

33. Stanovich, K. (2021). The many faces of myside bias. In The bias that divides us: The science and politics of myside thinking. *MIT Press*.