

Comunicação política no Facebook e previsão eleitoral:

Análise de big data da eleição presidencial brasileira de 2018 no Brasil

Leonardo Magalhães Firmino¹ Felipe Murta²

Resumo

O presente artigo busca entender como a análise quantitativa da comunicação política nas mídias sociais, mediante indicadores de engajamento ao discurso de candidatos, é capaz de prever o resultado eleitoral. Trata-se de uma análise de big data do 1º turno do pleito presidencial de 2018 no Brasil. Concretamente, foram coletadas cerca de 10 mil postagens das páginas oficiais de campanha no Facebook entre 1 de junho e 7 de outubro de 2018. No que se refere à intenção de voto, foi realizada para o mesmo período uma série de surveys de opinião com frequência diária, representativa da população brasileira. Com o fim de entender se há uma relação causal entre o engajamento ao candidato no Facebook e a sua intenção de voto, a análise preditiva foi realizada testando duas abordagens empíricas: uma com dados agregados e a outra com 90 modelos de previsão mediante regressão linear múltipla. Concluímos a análise comparando os resultados reais das eleições com todos os modelos preditivos. Os resultados mostraram que tanto a abordagem agregada quanto a regressiva demonstram que o índice de engajamento ao candidato no Facebook é um bom preditor eleitoral. Os resultados obtidos reforçam as teorias que defendem a relevância dos dados advindos do comportamento político nas mídias sociais como bons preditores eleitorais.

Palavras-chave

Comunicação Política; Previsão Eleitoral; Opinião Pública; Mídias Sociais; Big Data.

¹ Doutorando em comunicação política pela PUC-Rio e Coordenador de Projetos para a América Latina e Caribe em Atlas Político. E-mail:leonardo_firmino@msn.com.

² Doutorando em comunicação política pela PUC-Rio e fundador da Arena Digital. E-mail: femurpi@gmail.com.

Political communication on Facebook and election prediction:

Big data analysis of the 2018 Brazilian presidential election

Leonardo Magalhães Firmino¹ Felipe Murta²

Abstract

This article aims to understand how the quantitative analysis of political communication in social media, through indicators of engagement to the political speech, is able to predict the electoral outcome. This study is a big data analysis of the first round of the 2018 presidential election in Brazil. Specifically, around 10,000 posts were collected from official Facebook campaign pages between June 1 and October 7, 2018. Regarding voting intent, a series of daily representative of the Brazilian population opinion polls were conducted for the same period. In order to understand whether there is a causal relationship between candidate engagement on Facebook and voting intent, predictive analysis was performed by testing two empirical approaches: one with aggregated data and the other with 90 multiple linear regression prediction models. We conclude the analysis by comparing the actual election results with all predictive models. The results showed that both the aggregate and regressive approaches demonstrate that the candidate engagement rate on Facebook is a good electoral predictor. The results reinforce the theories that defend the relevance of data coming from political behaviour in social media as good electoral predictor.

Keywords

Political Communication; Election Prediction; Public Opinion; Social Media; Big Data.

¹ Doutorando em comunicação política pela PUC-Rio e Coordenador de Projetos para a América Latina e Caribe em Atlas Político. E-mail:leonardo_firmino@msn.com.

² Doutorando em comunicação política pela PUC-Rio e fundador da Arena Digital. E-mail: femurpi@gmail.com.

As eleições de 2018 representaram o começo de uma nova etapa, tanto no que se refere a campanhas eleitorais nas mídias sociais quanto ao campo da comunicação política no Brasil. No país, o advento como observamos em 2018 da campanha digital, esperado para 2014 e 2016, finalmente se fez sentir claro e forte em uma eleição fortemente polarizada, com novos atores da comunicação política e uma nova ecologia midiática.

Neste sentido, se faz necessária a realização de estudos preditivos que possam explicar mediante, métodos estatísticos, como a nova dinâmica mediática e comunicacional está afetando à comunicação política e às eleições no Brasil. Seguindo esta linha, o presente estudo se propõe como uma contribuição ao campo da Comunicação, Internet e Política, assim como o da Opinião Pública, realizando uma análise de *big data* das campanhas presidenciais no Brasil em 2018. Seguindo de maneira multidisciplinar consolidadas linhas de pesquisa sobre *media effect* e previsão eleitoral, integramos as contribuições da informática e da estatística aplicada ao presente estudo de caso, localizando-nos no que vem sendo chamado de Ciências Sociais Computacionais.

O presente estudo, portanto, se justifica pelo fato que ficou notório hoje, mais do que em anteriores eleições, o potencial das mídias digitais e de aplicativos como o WhatsApp, como eficientes plataformas de campanha eleitoral no que se refere à capacidade de disseminar mensagens por parte do candidato; dialogar com seus eleitores; divulgar de notícias falsas; utilizar de robôs; e impactar públicos segmentados com conteúdo específico. Com a massificação do uso do Facebook pelos diversos públicos que compõem a esfera pública online brasileira, abrem-se diversas possibilidades para entender como a plataforma em questão pode influenciar disputas eleitorais e, em particular modo, se e como o uso dessa ferramenta poderia ajudar a prever os resultados das eleições no Brasil mediante o estudo do comportamento político online.

Afirmar que estudos sobre a manifestação da opinião pública em redes sociais como o Facebook são capazes auxiliar na previsão de resultados eleitorais reforça a ideia de que a plataforma já vem sendo usada como meio de expressão da própria subjetividade, e de forma tão ampla que poderia considerar-se como análoga à uma amostra representativa da população brasileira. Com as devidas estratificações amostrais seria possível realizar estudos tão ou mais precisos que os tradicionalmente usados para este fim – como *surveys* e modelos estruturais, porém com a vantagem de

poder realizar, em tempo real, análises mais profundas e variadas. De fato, o estudo da opinião pública nas redes sociais, ou seja, da comunicação política e da interação de diversos atores com a mesma, só é possível graças à exposição pública e voluntária da própria opinião em diversos formatos e meios. Tais dados ficam armazenados e possuem um rico repertório de informações, onde acredita-se ser possível, inclusive, entender suas conexões com outros elementos online.

Nesta linha, o presente estudo é dividido em quatro partes incluindo esta introdutória. Em seguida, uma breve discussão teórica acerca das pesquisas que sustentam e as que consideram inconclusivo o uso de dados digitais coletados em redes sociais durante períodos eleitorais como índices válidos e influentes no processo de previsão de votos. Em terceiro vem a metodologia da pesquisa, a qual se baseou em analisar métricas quantitativas produzidas durante o período de campanha presidencial no Facebook para identificar se houve, de fato, indícios que possam apontar para um modelo de previsão eleitoral voltado para as interações do público com os candidatos mediante a comunicação política. Sobre isso, é necessário lembrar que esta pesquisa reconhece sua importância parcial por considerar fundamental e de igual relevância uma análise qualitativa de sentimento sobre as interações contabilizadas. A pesquisa se desenvolve na busca por comprovar a hipótese (H1) de que o engajamento no Facebook aos candidatos no 1º turno das eleições presidenciais do Brasil em 2018 prevê a intenção de voto aos mesmos. Para tanto, mediante a plataforma Atlas Político, foram coletadas todas as postagens dos candidatos à presidência da república para calcular o índice diário de engajamento. No que se refere aos dados de intenção de voto, os mesmos foram obtidos mediante *surveys* de opinião com frequência diária, representativos da população brasileira. Na conclusão, parte final, e em respeito à análise dos dados e resultados, foi encontrada uma forte correlação entre a intenção de voto nas urnas no primeiro turno e o engajamento aos candidatos no Facebook mediante a sua comunicação de campanha nas postagens.

Neste cenário, isoladamente observado, o Facebook se demonstrou um razoável indicador de intenção de voto. Mediante regressões lineares múltiplas foram testados 90 modelos combinatórios ao todo, assim como um modelo adicional baseado em dados agregados, confirmando a hipótese de relação causal entre as variáveis.

Comunicação, política e previsão eleitoral online

A campanha eleitoral é o momento mais intenso da disputa política entre representantes políticos, ao passo que constituem um dos elementos mais importantes

para o processo eleitoral democrático (NORRIS, 2000; GOMES, 2004). Porém, mesmo que as campanhas se desenvolvam como *media events* (DAYAN; KATZ, 1992) ou lutas reais entre ideias e programas partidários (GOMES, 2004), na prática, segundo Salgado (2012), constituem esforços organizados para mobilizar e convencer o eleitor com o fim de conquistá-lo, captar seus votos e fidelizá-lo, reduzindo com isso as possibilidades de desistência.

Para Norris (2000), as campanhas eleitorais estão divididas em três fases: a) pré-moderna, b) moderna e c) pós-moderna. A primeira implica formas mais diretas de comunicação entre candidatos e eleitores e uma forte tendência de identificação e adesão aos partidos. A segunda surge na segunda metade do século XX e implica o advento da televisão e a substituição da relação direta por uma relação mediada pelas mídias de massa. A terceira fase começa com a adoção das TICs a partir da década dos 90, em um contexto em que a internet e as redes sociais se somam ou complementam sem substituir, porém, as formas anteriores de expressão, tais como o rádio, os jornais e a televisão. Esta fase, pela sua maior competitividade, determina a passagem da massificação à segmentação do público e marca também a crescente profissionalização das campanhas com profissionais, consultores e ‘marqueteiros’ políticos que desenvolvem um papel relevante na criação, difusão e promoção dos candidatos.

No contexto das campanhas eleitorais pós-modernas, o aumento constante do uso da internet nas estratégias de comunicação, causa a ampliação das ações e escopo das equipes de comunicação e marketing político que se transformam em comunicação política online. As mídias sociais pelas suas estruturas de rede e as suas lógicas interativas e colaborativas, distinguem-se da rotina comunicacional dos meios tradicionais, caracterizada pelo fluxo unidirecional da informação (WEIMANN; BROSIUS, 2016).

Perguntando-se como as mídias sociais afetam às eleições, não é de hoje que a busca por previsões de votos em períodos eleitorais mobiliza o mundo acadêmico. O uso cada vez mais disseminado e frequente das mídias sociais por parte da sociedade levou estudiosos a perceber a necessidade de incorporar os dados nelas disponíveis nas recentes pesquisas de previsão eleitoral, indo muito além dos métodos costumeiramente utilizados nas pesquisas tradicionais de intenção de voto ou em modelos estruturais típicos da Ciência Política (FIGUEIREDO, 2007).

Mesmo com o recente protagonismo assumido pelas mídias sociais, especificamente no contexto eleitoral, são usuais os trabalhos inseridos tanto no campo da Comunicação Política Digital quanto nas Ciências Sociais e da Computação

que defendem a influência de determinados tipos de dados digitais em processos eleitorais. Porém, bilhões de usuários ativos compartilhando incessantemente mensagens e arquivos fizeram com que a relação entre mídia social e campanha eleitoral fosse considerada parte fundamental não apenas do ambiente de mídia atual como do processo eleitoral. Dado o grande volume de informações pessoais que produzem, essas novas mídias vêm sendo cada vez mais exploradas para fins de monitoramento e previsão de eventos no mundo. Não à toa muitos consideram que a evolução tecnológica dos meios de comunicação, assim como a amplificação do acesso ao ambiente online em escala global, fizeram com que estudos na área da comunicação e política passassem a coletar dados que os eleitores produziam nas redes sociais e outras informações extraídas do ambiente online para buscar validações em suas análises e conclusões científicas (DI GRAZIA, *et al.*, 2013; KRISTENSEN, *et al.*, 2017; SAEZ-TRUMPER, MEIRA; ALMEIDA, 2011; TUMASJAN, *et al.*, 2013). Twitter, YouTube, Facebook, Instagram e WhatsApp são exemplos de ambientes digitais que desde 2008 estão sendo, cada uma no seu tempo, massivamente utilizados por campanhas eleitorais na disseminação do seu conteúdo diretamente para o seu público sem a necessidade de intermediários, papel até então desempenhado pela mídia tradicional. De fato, estudos recentes sobre agenda-setting estão evidenciando que, embora a mídia tradicional ainda tenha grande poder de pautar a agenda do público (GUO; MCCOMBS, 2015), novos atores políticos e comunicacionais estão enfraquecendo o papel tradicional do jornalista *como único gatekeeper*, conseguindo estabelecer novos fluxos informacionais (WEIMANN; BROSIUS, 2016).

Ao investigar se as redes sociais possuem indicadores válidos e expressivos sobre o comportamento político do eleitor, autores como Tumasjan, Sprenger e Sandner (2010) dedicaram-se a observar o comportamento de pessoas em plataformas digitais durante campanhas eleitorais e escolheram o Twitter como objeto de suas pesquisas por apresentar características propícias ao debate político, discussões, trocas de ideias e acesso à informações globais em tempo real. Argumentavam que, como previsor de resultado de eleições, a simples contagem de menções a partidos ou candidatos poderia ser capaz de acompanhar pesquisas eleitorais tradicionais e apresentar resultados próximos aos resultados finais da eleição (MIRANDA *et al.*, 2014). Metodologia com tendência semelhante também pôde ser vista com Trumper, Meira e Almeida (2011). Os pesquisadores realizaram reduzidas, porém significativas mudanças na metodologia de Tumasjan *et al.* (2010). Ao invés de considerar todas as menções a todos os perfis analisados, os cálculos das projeções foram feitos considerando apenas uma menção a um termo de cada (TRUMPER; MEIRA; ALMEIDA,

2011). No final da pesquisa, obtiveram resultados satisfatórios de previsão eleitoral com um erro absoluto médio de 4,07 pontos percentuais.

Outros pesquisadores (BENEVENUTO *et al.* 2010; LUMEZANU *et al.* 2012) optaram por debruçar-se sobre a identificação de perfis pessoais de usuários nas redes sociais. Di Grazia *et al.* (2013), por sua vez, mostrou em pesquisa sobre o cenário político dos Estados Unidos que a porcentagem de menções a candidatos republicanos no Twitter durante parte do ano de 2010 se correlacionam com a margem de votos que os mesmos obtiveram nas eleições daquele ano. Os autores também trazem o debate sobre a falta de ênfase que os recentes estudos da área dão a variáveis como a incumbência dos candidatos, o partido ao qual eles pertencem, cobertura midiática e a composição sociodemográfica do eleitorado, mostrando que, mesmo levando em consideração essas variáveis, a correlação se mantinha (DI GRAZIA, *et al.*, 2013). As redes sociais, segundo os autores, podem ser um indicador válido do eleitorado americano. São muitos os estudos sobre as características das contas de eleitores, que assiduamente interagem com conteúdo político de campanha. Assim como são inúmeros e extensos os trabalhos que abordam o comportamento específico dos perfis voltados para a disseminação de informação política em período eleitoral (LUMEZANU *et al.*, 2012), identificando padrões de uso e mensagens (GHOSH, 2012).

Há também os estudos analíticos quantitativos focados em comprovar se as formas conhecidas de interação entre eleitor e campanha eleitoral nas redes sociais, como por exemplo o compartilhamentos, a curtidas e o *retweets*, entre outras modalidades, podem ser aceitas como índices relevantes e capazes de alimentar pesquisas de intenção de voto em um sistema multipartidário de eleições democráticas (KRISTENSEN *et al.*, 2017). Kristensen produziu notório levantamento sobre o comportamento dos eleitores nas redes sociais durante as eleições locais na Dinamarca (2017). Para tal, realizou pesquisas tradicionais de intenções de voto e as correlacionou com as informações obtidas em bancos de dados coletado a partir das publicações em páginas públicas de partidos e políticos dinamarqueses em campanha entre janeiro de 2015 e janeiro de 2017. O objetivo era não apenas comprovar se havia alguma relação direta entre os resultados da pesquisa tradicional com os dados digitalmente coletados, como também comprovar se é possível identificar as preferências políticas do eleitor a partir da forma como estes interagiram com as campanhas eleitorais (KRISTENSEN *et al.*, 2017). Em posse de novos métodos para previsões eleitorais, o autor afirmou ser possível não somente separar público por perfis e preferências políticas como também generalizar as conclusões para níveis nacionais e até além da fronteira. Assim, Kristensen desafiou a tendência contemporânea de buscar métodos

baseados em big data para atingir conclusões científicas ao mostrar que uma curtida a uma publicação ou página no Facebook pode ter grande importância científica (KRISTENSEN *et al.*, 2017).

Além disso, argumentou que a maioria dos estudos de previsão de comportamento individual de eleitores consideram contextos bipartidários, como no caso da política americana. O contexto eleitoral em questão, portanto, é uma variável determinante neste caso, já que o sistema de partidos define qual a melhor metodologia a ser empregada em cada cenário. Um esquema metodológico para sistemas multipartidários também foi apresentado por Tumasjan *et al.*, (2013). Os pesquisadores concluem não só que o número de menções ao partido refletiu o resultado das eleições, como também que o sentimento das mensagens corresponde às preferências políticas dos eleitores (TUMASJAN *et al.*, 2013).

Pesquisas como essas geraram muitos testes metodológicos (JUNGHERR, 2012; GAYO-AVELLO, 2010) sob a premissa de buscar uma confirmação sobre algumas das hipóteses anteriormente mencionadas. Muitos autores acabaram não comprovando a afirmativa de que é possível prever um resultado eleitoral apenas com índices quantitativos como, por exemplo, a contagem de menções. Foram apontadas inúmeras limitações básicas como a falta de regras bem fundamentadas para a coleta dos dados e o recorte temporal ideal para esse tipo de análise de previsão. Entre os testes metodológicos mais relevantes podemos destacar a análise feita sob tweets coletados durante as eleições federais na Alemanha em 2009 (JUNGHERR, 2012) e para o congresso norte americano em 2010 (GAYO-AVELLO, 2010). Ambos não observaram nenhuma relação entre os dados coletados e o resultado das urnas. Gayo-Avello chegou a realizar uma grande pesquisa bibliográfica (2014) sobre trabalhos que objetivassem previsões eleitorais a partir de dados coletados em plataformas sociais digitais, focando sua análise em levantar erros, apontar falhas e limitações metodológicas até então empregadas.

Outra grande deficiência recorrente em pesquisas de monitoramento de rede e previsões eleitorais (MUKHERJEE *et al.*, 2013) é a falta de mecanismos utilizados para distinguir, por exemplo, uma simples menção a um candidato de um voto em si. Essa lacuna evidenciou, entre tantas outras coisas, a importância da compreensão sobre o que é publicado textualmente por esses usuários. Assim foi introduzida a análise de sentimento sobre o conteúdo produzido por usuários de redes sociais durante períodos eleitorais como sendo uma determinante que não pode ser ignorada por quem busca associar dados coletados nas mídias digitais com o resultado obtido em disputas eleitorais (BORA, 2014).

Está claro e evidente, portanto, que relativizações são necessárias nesse campo de estudo. Pesquisadores apontam, também, o foco em entretenimento e expressões emocionais do fluxo de comunicação de todas as redes como um desses problemas (DI GRAZIA *et al.*, 2013). O chamado *digital divide*, um fenômeno que ocorre porque parte da população mundial não tem (ou tem pouco) acesso pleno à internet ou mesmo a recursos que possibilitam uma conexão online no meio digital, também é mencionado, gerando fortes questões sobre a real presença da população do planeta, de forma representativa, no ambiente online. Esse problema específico pode ser atenuado quando a metodologia do estudo trabalha com *big data* (FAN; BIFET, 2013). Porém, a amostra de pessoas em determinada rede social pode ser enviesada, já que todo serviço online, por sua natureza, atrai determinados segmentos populacionais como consumidores. Por consequência, podemos identificar perfis de público distintos e que podem não representar amostras adequadas para estudos científicos certos tipos (DI GRAZIA *et al.*, 2013).

Todos os fatores anteriormente mencionados levam a uma ideia de que a escolha da mídia para a realização de estudos de previsão eleitoral pode ser determinante, principalmente considerado as diferenças técnicas e estruturais de cada uma, o que por vezes pode impor mudanças ou formas de funcionamento de metodologias de estudo ou mesmo influenciar seus resultados. Kristensen (2017), por exemplo, defende que estudos na área de previsão eleitoral que se baseiam no Facebook e no Twitter mostram melhores resultados do que estudos que usam outras mídias. No campo, há pesquisas que mostram o potencial de previsão eleitoral baseados em dados digitais de plataformas como o YouTube, Google e até o Wikipedia (YASSERI; BRIGHT, 2016).

Metodologia

O presente estudo analisou a comunicação política no Facebook dos aspirantes ao cargo de Presidente da República durante o período de campanha do 1º turno, para identificar se há um modelo de previsão eleitoral baseado nas interações do público com os candidatos no Facebook. Assim, temos a nossa hipótese (H₁):

H₁: O grau de engajamento ao candidato no Facebook prevê a intenção de voto no mesmo, no 1º turno das eleições presidenciais do Brasil em 2018.

Com o fim de entender se o engajamento ao candidato no Facebook é consistente como preditor eleitoral, mediante a plataforma Atlas Político e Datafolha, foram utilizadas duas fontes de dados. A primeira se refere ao engajamento diário do candidato no Facebook, que será a nossa variável independente, obtida mediante

a *Dashboard* de *Big Data* do Atlas político, que trabalha com a API da rede social. A segunda fonte de dados, como variável dependente, é a pesquisa sobre intenção de voto, realizado diariamente mediante a plataforma Atlas Tracking do Atlas Político.

No que se refere ao Facebook foram coletadas cerca de 10 mil postagens, provenientes das páginas dos nove candidatos à Presidência da República estudados, entre 1 de junho e 7 de outubro de 2018. Os candidatos monitorados foram: Jair Bolsonaro (PSL), Fernando Haddad (PT), Ciro Gomes (PDT), Geraldo Alckmin (PSDB), João Amoedo (Novo), Henrique Meirelles (MDB), Marina Silva (Rede), Alvaro Dias (Podemos) e Guilherme Boulos (PSOL).

Cada postagem foi classificada pelo seu índice de engajamento, dado pela soma das reações, comentários e compartilhamentos. Em seguida, foi obtido o engajamento total diário de cada candidato, realizando a somatória do engajamento de cada sua postagem no período estudado.

Respeito às pesquisas de intenção de voto, o Atlas Tracking é uma plataforma que realiza *surveys* diários de opinião online. O questionário é publicado e recolhe entre 800 e 2000 respostas diariamente em todo o território nacional. Uma vez encerrado o tempo de resposta, e realizada a coleta definitiva dos dados, é feita uma pós-estratificação com balanceamento amostral, baseada em sete variáveis chave, além da verificação geográfica mediante endereço IP como forma de controle adicional.

No que tange mais especificamente o desenho amostral, trata-se, portanto, de um estudo pós-estratificado, representativo da população eleitoralmente ativa, com seleção aleatória em todos os estados do Brasil. As respostas finais têm um peso amostral ajustado por um algoritmo de ranqueamento para garantir a representatividade mediante sete variáveis chave de estratificação, como gênero, idade, renda mensal familiar, local de residência, uso das redes sociais e voto no 1º e 2º turno das últimas eleições para presidente da república. A margem de erro é de $\pm 2\%$ e o intervalo de confiança é de 95%.

Os dados do Datafolha foram utilizados somente para fins de validação. De fato, se compararmos o Atlas Tracking com o Datafolha, teremos somente uma discrepância de até 2% em 95% dos dados, o que, de certa forma, valida os dados utilizados no presente artigo. A comparação pode ser observada no GRAF. 1 a seguir:

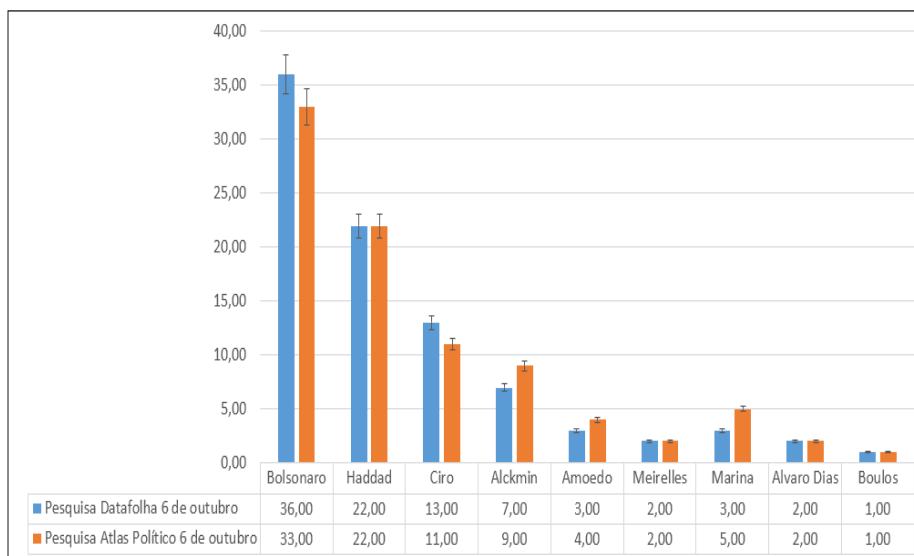


Gráfico 1: Comparação entre Atlas Tracking e Datafolha em 6 de outubro de 2018.
Fonte: Elaboração dos autores. (2019)

Com uma série final de 115 unidades de análises, relativas a cada dia estudado, temos os dados diários de engajamento ao candidato e intenção de voto ao mesmo. Com ditos dados, foi realizada uma matriz de correlação de Pearson. Em seguida, concentrando-nos exclusivamente nos coeficientes de correlação relativos ao engajamento e intenção de voto de cada candidato, foram realizadas regressões lineares múltiplas para entender se é possível afirmar que há um modelo de previsão eleitoral baseado no engajamento no Facebook. O software utilizado para a análise foi o SPSS.

Antes de realizar as regressões lineares, foram garantidas todas as condições prévias para o uso de modelos paramétricos como:

- distribuição normal e simétrica das variáveis dependentes;
- relação linear entre as variáveis dependentes e independentes;
- $n > 30$;
- independência dos valores;
- distribuição normal dos resíduos;
- homocedasticidade (independência dos resíduos).

Uma vez garantidas todas as condições ideais para a realização de estatísticas paramétricas, foram testados ao todo 90 modelos de previsão e selecionado, ao final, aquele que oferecia um valor de R^2 ajustado maior para cada variável dependente, indicando maior nível de previsibilidade

Já que se buscou entender como a performance comunicativa dos candidatos no Facebook impactou a opinião pública, gerando um efeito sobre a intenção de voto, optou-se por estudar o impacto mútuo entre diversas campanhas entre si.

Por esta razão foram realizadas regressões lineares múltiplas em lugar das lineares simples. É natural pensar que uma comunicação política que engaja eleitores pelos seus argumentos e símbolos evocados, possa desencajar os mesmos eleitores para outras campanhas. Porém, o panorama é ainda mais complexo, já que, se o que engaja o eleitor, entre outros fatores, é o discurso político, a adesão a certas pautas pode também favorecer a outros candidatos. Também por esta razão optamos por um modelo regressivo linear múltiplo, pois a comunicação de um candidato pode tanto afetar positivamente como negativamente o engajamento do público a outros candidatos.

Apresentação dos resultados

A seguir pode-se ler a Tabela 1 que mostra a matriz de correlação de Pearson. Como pode-se observar, há valores altos, acima de 0,7 e valores baixos, de 0,2. Tratou-se de não descartar os valores pequenos (inferiores entre -0,3 e 0,3), pois somente a regressão dirá se as correlações são significativas ou não. Assim, por exemplo, poder-se-ia ter um uma correlação de 0,1 com maior significância que outra de 0,7.

		Engajamento no Facebook								
		Bolsonaro	Haddad	Ciro	Alckmin	Amoedo	Meirelles	Marina	Alvaro Dias	Boulos
Intenção de Voto	Bolsonaro	0,772668	0,82148	0,690386	0,468526681	0,613643	0,559523	0,580131	0,4024948	0,43104
	Haddad	0,771551	0,79389	0,669228	0,502908484	0,608433	0,573569	0,558526	0,3858573	0,44742
	Ciro	0,677502	0,76349	0,616228	0,448435305	0,434198	0,514023	0,488581	0,3408383	0,31619
	Alckmin	0,524813	0,58916	0,487984	0,337442535	0,357502	0,384803	0,393336	0,204047	0,30749
	Amoedo	0,588843	0,59382	0,434798	0,488296189	0,70473	0,425705	0,526545	0,269933	0,26031
	Meirelles	0,435283	0,57391	0,418292	0,313698742	0,248332	0,401896	0,32302	0,286139	0,17791
	Marina	-0,46807	-0,43935	-0,449121	-0,108871222	-0,21669	-0,357207	-0,343242	-0,288283	-0,3803
	Alvaro Dias	-0,65136	-0,6455	-0,579714	-0,581487091	-0,648525	-0,488304	-0,534023	-0,26238	-0,31694
	Boulos	0,116386	0,13737	0,063556	0,075855564	0,11612	0,103853	0,006013	0,08412	0,03963

Tabela 1: Matriz de correlação de *Pearson*.

Fonte: Elaboração dos autores. (2019)

Já que correlações são somente indicativos iniciais que não indicam necessariamente um fenômeno de causalção, realizamos regressões múltiplas mediante modelagem linear no SPSS. Utilizamos como preditores de entrada todas as variáveis sobre engajamento, com nível de confiança de 95%, com método de seleção dos modelos em base aos melhores subconjuntos, usando como parâmetro comparativo, de entrada e saída, o R² ajustado. Ao todo foram testados 90 modelos preditivos e selecionados os que obtiveram os mais altos índices para cada variável dependente.

Tem-se assim a Tabela 2, que representa a sumarização dos modelos

de regressão para a variável de intenção de voto por candidato. Em dita tabela reportamos somente o melhor modelo para cada variável dependente. Na horizontal vemos as variáveis dependentes “intenção de voto” e na vertical temos as variáveis independentes “data” e “engajamento aos candidatos no Facebook”. As marcações indicam a presença ou ausência de cada variável independente no modelo de previsão da dependente. A linha relativa à R² ajustado indica o grau de previsibilidade de cada modelo, expresso nas colunas. As marcações em verde mostram as variáveis que têm mais peso no modelo ($p < 0,05$ e na maioria dos casos $p < 0,01$). As marcações em vermelho indicam variáveis de engajamento que têm um peso muito pequeno no modelo se considerarmos somente graus de significância de $p > 0,05$.

		Intenção de Voto								
		Bolsonaro	Haddad	Ciro	Alckmin	Amoedo	Meirelles	Marina	Alvaro Dias	Boulos
R² ajustado		,653	,521	,624	,641	,600	,342	,262	,814	,101
Engajamento	Bolsonaro	✓	✓	✓		✓	✓	✓		✓
	Haddad	✓	✓	✓				✓	✓	✓
	Ciro	✓	✓	✓	✓	✓	✓	✓	✓	
	Alckmin	✓						✓	✓	
	Amoedo		✓	✓	✓	✓	✓	✓		✓
	Meirelles	✓				✓		✓		✓
	Marina		✓			✓				✓
	Alvaro Dias				✓	✓			✓	
	Boulos	✓	✓	✓		✓		✓		✓
	Data	✓	✓	✓	✓	✓	✓		✓	✓

Tabela 2: Sumarização dos modelos de regressão múltipla.

Fonte: Elaboração dos autores. (2019)

Análise

A metodologia de análise consistiu em realizar dois tipos de coleta de dados. O índice de engajamento diário no facebook e o tracking diário de intenção de voto. No que tange à correlação entre engajamento e intenção de voto, foi evidenciado que há correlações discrepantes entre os dados de cada candidato. De modo geral há graus de correlação suficientemente altos, como pode-se contemplar na Tabela 1. Respeito às regressões, o primeiro ponto a salientar é o relativo ao teste de significância total e dos coeficientes do modelo.

O R² ajustado indica o quanto cada modelo de regressão linear múltipla consegue prever a variável dependente, ou seja, quais variáveis de engajamento no facebook conseguem explicar a variável dependente intenção de voto.

De modo geral, vemos pela TAB. 3 a seguir, que se obtiveram resultados que

confirmam a hipótese de previsão eleitoral mediante o engajamento. Não obstante a percentual pequena de previsibilidade para Boulos (10%), Marina (26%) e Meirelles (34%), os dados possuem alta significância. Todos os demais candidatos apresentam índices elevados de previsibilidade eleitoral pelo engajamento no Facebook, a exceção de Alvaro Dias, que obteve uma forte presença do elemento “data” no seu modelo, o que o invalida quase por completo neste caso específico.

	Candidatos								
	Bolsonaro	Haddad	Ciro	Alckmin	Amoedo	Meirelles	Marina	Alvaro Dias	Boulos
R2 ajustado (%)	65,3%	52% ¹	62,4%	64,1%	60%	34,2%	26,2%	81,4%	10,1%

Tabela 3: rau de previsibilidade dos modelos por candidato.

Fonte: Elaboração dos autores. (2019)

Por outro lado, pode-se observar que não sempre a variável de engajamento do próprio candidato explica de forma mais adequada o crescimento ou decréscimo de sua intenção de voto. Como é possível ver no caso de Boulos, o engajamento dos públicos ao Haddad explica boa parte da performance eleitoral do pessoalista. Este indício faz muito sentido, pois ambos compartilham uma parcela de eleitorado à esquerda. Quando se compartilha dos mesmos públicos, o crescimento de um significa o decréscimo do outro.

Outro elemento interessante a ser levado em consideração é o peso que a variável ‘data’ possui para diversos candidatos. Esta variável foi testada como independente em todos os modelos para entender se o fator de proximidade ao dia da votação mostraria algum impacto sobre a intenção de voto. Em alguns casos a data teve um peso enorme sobre o modelo, como no caso de Alckmin e Alvaro Dias, em outros o peso foi moderado ou marginal. Este fator também é importante na hora de considerar o engajamento um bom preditor.

Por outro lado, pode-se apresentar outra abordagem aos dados para realizar uma previsão eleitoral. Não obstante para tal abordagem não seja possível realizar uma regressão, pode-se observar claramente no gráfico 3, a seguir, como o total ponderado (a 100%) dos valores também representa indícios de que o engajamento no Facebook pode ser usado como abordagem indicativa à previsão eleitoral.

O Gráfico 2 representa o resultado do 1º turno, comparando-o ao Datafolha, ao Atlas Tracking e à média de engajamento total ponderado no Facebook. O gráfico demonstra que o engajamento consegue oferecer resultados bastante parecidos ao resultado da votação e às pesquisas de opinião, inclusive mais precisos em alguns

casos.

Tanto a abordagem regressiva linear múltipla quanto a de dados agregados ponderados ofereceram bons indícios de que o engajamento no Facebook pode ser utilizado como preditor eleitoral, embora não ofereça uma confiabilidade ao nível das tradicionais pesquisas de opinião. Em especial, a abordagem de dados agregados foi a que melhor previu a votação no 1º turno das eleições presidenciais de 2018. O gráfico 3 a seguir nos ilustra a comparação entre Datafolha, através Tracking e Engajamento ponderado no Facebook. Como pode-se contemplar, o engajamento ponderado, se aprimorado em futuros trabalhos, pode representar um índice bastante confiável para a previsão eleitoral.

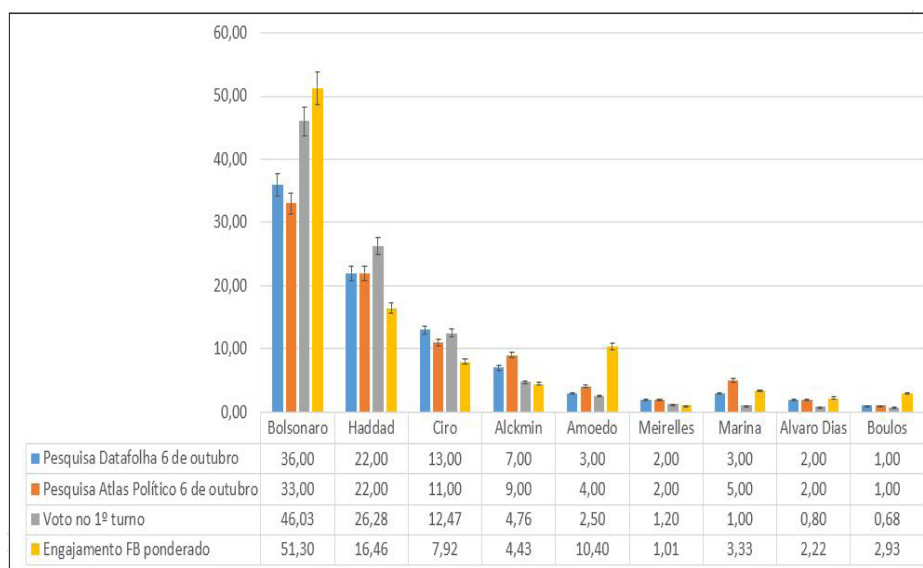


Gráfico 2: Comparação entre preditores eleitorais (%)
Fonte: Elaboração dos autores (2019).

Considerações Finais

O presente estudo teve como objetivo identificar na comunicação política no Facebook, índices capazes de reforçar teorias que defendem a representatividade dos dados digitais do comportamento comunicacional dos públicos em relação aos candidatos como preditor eleitoral confiável. Mais precisamente, a pesquisa buscou identificar nas métricas de engajamento no Facebook algum potencial capaz e auxiliar na previsão de votos em disputas eleitorais no Brasil, que pudesse ser comparado com os meios tradicionalmente usados para este tipo de pesquisa, realizadas mediante survey ou modelos estruturais, estes últimos mais comuns nos Estados Unidos. Com base na leitura dos dados digitais extraídos, analisados e dos resultados demonstrados nos gráficos e tabelas acima, o estudo concluiu que embora a média dos valores

pudesse refletir ótimos indícios sobre o resultado final do pleito em questão e, em alguns casos, até mesmo o percentual aproximado alcançado individualmente por cada campanha, as regressões realizadas ajudam a reforçar as teorias que corroboram com a existência deste tipo de modelo de previsão eleitoral.

De maneira geral podemos identificar nos dados vários aspectos interessantes a serem aprofundados em futuros trabalhos, como por exemplo, a clara presença de influência mútua entre candidatos que compartilham o mesmo eleitorado, ou que buscaram se apropriar dos mesmos temas na agenda ao longo da campanha eleitoral. Embora os valores de R2 ajustado não ficassem acima de 80% para todos os candidatos, justificando o caráter focal e experimental dessa pesquisa, o mesmo torna-se relevante para futuros trabalhos com intuito de comprovar a existência de índices que, de fato, venham a constituir uma nova fase de estudos de previsão eleitoral, tão ou mais precisos que os modelos tradicionalmente utilizados.

Buscar-se-á em futuros trabalhos cruzar outras variáveis ao engajamento, como as relativas às pautas políticas tratadas, análise de sentimento para diferenciar o engajamento positivo do negativo, assim como realizar um estudo intermídia, integrando dados de outras mídias sociais como o Twitter o Instagram, plataformas de conteúdo como o YouTube e de mensagens como o WhatsApp.

Notas

[1] O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001

Referências

BIFET A., FAN W.: Mining Big Data: Current Status, and Forecast to the Future. *SIGKDD Explorations*, v. 14, n.2, p.1-5, 2013.

BORA, N.N. Summarizing public opinions in tweet. *Journal Proceedings of CICLing*, 2014.

DIGRAZIA J, MCKELVEY K, BOLLEN J, ROJAS F. More Tweets, More Votes: Social Media as a Quantitative Indicator of Political Behavior. *PLoS ONE*, v.8, n.11, DOI <https://doi.org/10.1371/journal.pone.0079449>, 2013.

DAYAN, D.; KATZ, E. *Media events: the live broadcasting of history*. Cambridge: Harvard University Press, 1992.

GAYO-AVELLO, D. “I wanted to predict elections with twitter and all I got was this lousy paper” – A balanced survey on elections prediction using twitter data. *CoRR*, 2014, Online. Disponível em: <<http://arxiv.org/abs/1204.6441>>. Acesso em: 5 nov. 2019.

GAYO-AVELLO, D.; METAXAS, P. AND MUSTAFARAJ, E..Limitis of election predictions using twitter. *Int. Conf. on Weblogs and Social Media (ICWSM)*, p. 490-493, 2011. INSERIR LINK E

DATA DE ACESSO

GHOSH, S.; VISWANATH, B.; KOOTI, F.; SHARMA, N.K.; KORLAM, G.; BENEVENUTO, F.; GANGULY, N.; GUMMADI, K.P. Understanding and combating link farming in the twitter social network. *Int. Conf. on World Wide Web, WWW' 12*, p.61-70, 2012.

GOMES, N. *Formas persuasivas de comunicação política* - Propaganda Política e Publicidade eleitoral. Porto Alegre: EDUPUCRS, 2004.

GUO, L.; MCCOMBS, M. The Power of Information Networks: New Directions for Agenda Setting. [s.l.] : Routledge, 2015. JUNGHERR, A.; JÜRGENS, P.; SCHOEN, H. Why the pirate party won the german election of 2009 or the trouble with predictions: A response to tumasjan, A., sprenger, T. O., sander, P. G., & welppe, I. M. "predicting elections with twitter: What 140 characters reveal about political sentiment". *Social Science Computer Review*, [s. l.], 2012.

KRISTENSEN J.B; ALBRECHTSEN T; DAHL-NIELSEN E; JENSEN M; SKOVRIND M, BORNAKKE T. Parsimonious data: How a single Facebook like predicts voting behavior in multiparty systems. *PLoS ONE*, v.12, n.9, p.2017.

LUMEZANU, C.; FEAMSTER, N.; AND KLEIN, H. #bias: Measuring the tweeting behavior of propagandist. *Int. conf. on Weblogs and Social Media (ICWSM)*, p. 210-217, 2012.

MUKHERJEE, S.; MALU, A.;Bhattacharyya, P.;BHATTACHARYYA, P. Twisent: a multistage system for analyzing sentiment in Twitter.*Int. conf. on Information and knowledge management, CIKM' 12*, pagens 2531-2534, 2013.

NORRIS, P. *Democratic divide?* The impact of the Internet on parliaments worldwide. Harvard: Harvard University, 2000.

SALGADO, S. Campanhas eleitorais e cobertura mediática: abordagens teóricas e contributos para a compreensão das interações entre política e media. *Revista Brasileira de Ciência Política*, n.9, p. 229- 253.

TRUMPER, D. S.; MEIRA, W., & ALMEIDA, V. From Total Hits to UniqueVisitorsModel for Election'sForecasting. *Proceedings of the ACM WebSci'11*. Koblenz, 2011. TUMASJAN A; SPRENGER T. O.; SANDNER P. G.;WELPE I. M.. Election Forecasts With Twitter: How 140 Characters Reflect the Political Landscape. *Social Science Computer Review*, 2010.

TUMASJAN, A.; SPRENGER, T.O.; SANDER, P.G., & WELPE, I.M. Predicting elections with twitter: What 140 characters reveal about political sentiment. *Int. conf. on Weblogs and Social Media (ICWSM)*, p. 178-185, 2010.

WEIMANN, G.; BROSIUS, H.-B. . A New Agenda for Agenda-Setting Research in the Digital Era. In HENN, P.; VOWE, G. (Eds.), *Political Communication in the Online World: Theoretical Approaches and Research Designs*. Nova York: Routledge, 2016, p. 26-44.

YASSERI T.; BRIGHT J. Wikipedia traffic data and electoral prediction: towards theoretically informed models. *EPJ Data Sci*, 2016.