10, 11, 12, 13 e 14 de novembro de 2025
ISSN 2177-3866

SEGMENTATION STRATEGIES FOR POLITICAL MARKETING ON SOCIAL MEDIA: A cluster analysis

VINICIUS CARVALHO
CENTRO UNIVERSITÁRIO DO INSTITUTO DE EDUCAÇÃO SUPERIOR DE BRASÍLIA - IESB (IESB)

SIMONE DE ARAÚJO GÓES ASSIS
CENTRO UNIVERSITÁRIO IESB - INSTITUIÇÃO DE ENSINO SUPERIOR DE BRASÍLIA

SERGIO CORTES
CENTRO UNIVERSITÁRIO DO INSTITUTO DE EDUCAÇÃO SUPERIOR DE BRASÍLIA - IESB (IESB)

SEGMENTATION STRATEGIES FOR POLITICAL MARKETING ON SOCIAL MEDIA: A cluster analysis

Keywords: Voter segmentation, Digital Political Marketing and Clustering in Data Science.

1. Introduction

The rise of digital social networks has drastically reconfigured the contemporary political arena, transforming not only communication channels but also the very nature of interaction between candidates, voters, and public debate. The transition from a unidirectional mass communication model to a "many-to-many" paradigm — in which multiple senders and receivers interact simultaneously — has consolidated what Manuel Castells, in 1999, defined as the "Network Society," a new social morphology where power shifts from traditional institutions to the information flows that circulate within these networks. Iconic electoral campaigns, from Barack Obama to Trump, have demonstrated the effectiveness of using these platforms to establish a direct channel with the electorate, bypassing the filters of traditional media and mobilizing supporter bases with unprecedented agility. However, this new communication ecosystem, characterized by a massive and unstructured volume of data generated every second, poses a fundamental challenge to political marketing strategies. Traditional voter segmentation approaches, predominantly based on demographic and geographic criteria, prove insufficient to capture the complexity of identities and behaviors in the digital environment. Such methods offer a static and often superficial view, failing to identify the causes that motivate political affinities, as already pointed out by Haley's (1968) benefit segmentation theory. The heterogeneity of the virtual electorate and the fluidity of the issues discussed online demand a more dynamic and precise approach using data science. The evolution of political communication strategies necessarily involves the incorporation of advanced data analysis methodologies (De Slegte et al., 2024). The field of data science, particularly unsupervised machine learning techniques such as cluster analysis, offers a robust solution to this challenge (Papakyriakopoulos et al., 2020; Bessi & Ferrara, 2016).

In this context, the following research question emerges: how can the application of clustering techniques to behavioral data extracted from social networks contribute to improving segmentation in digital political marketing, allowing the identification of groupings of electoral profiles?

To this end, this research aims to improve segmentation strategies in digital political marketing by applying clustering techniques to behavioral data extracted from social networks, aiming to identify electoral groups with greater precision and analytical relevance.

By proposing the application of clustering techniques based on behavioral data extracted from social media, this research seeks to contribute to the improvement of segmentation strategies in digital political marketing. Its relevance lies in overcoming static electoral classification models by integrating data science and political communication into a replicable methodological framework. This provides support for more effective, ethical, and evidence-driven decisions, strengthening strategic intelligence in the connected public sphere.

3. Segmentation, social media and data analysis in political marketing

This chapter underpins the theoretical and methodological choices of the research by articulating the concepts of political segmentation, communication transformations on social media, and data science techniques applied to political marketing. It is based on the premise that, given the complexity of the digital environment, understanding voter behavior requires going beyond traditional segmentations, incorporating computational approaches capable of revealing latent patterns in online interactions. The main conceptual frameworks supporting this perspective are discussed below.

3.1 Segmentation strategies in political marketing and social networks

The evolution of the marketing discipline from a product-centric logic to a human-centric approach in a digitally connected world — a trajectory systematized by Kotler et al. (2017) — makes market segmentation a strategic pillar, integrating traditional and digital marketing to guide customers throughout the entire purchasing journey, generating a complex and unprecedented data ecosystem. Although established classical literature defines geographic, demographic, psychographic, and behavioral segmentation as the basis for segmentation, the most fundamental approach is benefit segmentation, proposed by Haley in 1968 and tested by several authors, including Crittenden & Muzyka (2002) and Kenney & Wenstein (2010). Its relevance lies in focusing on the determinants of behavior, identifying segments by causal factors, rather than descriptive ones, based on the premise that the benefits sought are the true reason for the existence of market segments.

The application of these principles in Brazilian politics is not a new phenomenon, but a continuum of professionalization that adapts to the technologies of each era, as demonstrated by Almeida (2007). This process ranges from Prudente de Moraes's use of voter lists in 1894 to Jair Bolsonaro's social media-based campaigns in 2018, confirming that the objective of segmenting voters to win votes has remained the same, with only the tools changing. However, in the current context, the effectiveness of demographic and geographic segmentation is limited in the face of a heterogeneous electorate, and the great challenge is to identify at scale the beneficial segments — that is, voters who "buy" candidates who offer the solutions, representation, or values they seek — based on massive and unstructured social media data, a complexity that traditional methods such as opinion polls cannot capture. To this end, the literature has explored cluster analysis, a technique used by Data Science that, by grouping individuals by their digital behaviors and interactions, reveals real affinities that transcend classical categories, enabling the application of benefit segmentation theory on a large scale for a deeper and more actionable understanding of the electorate (De Slegte et al., 2024, Papakyriakopoulos et al., 2020; Bessi & Ferrara, 2016).

The transition to a digital public sphere has reconfigured political communication, shifting the power of mass media to a hybrid media ecosystem where social networks have become central arenas of dispute, reflecting an academic debate that oscillated between initial optimism about the internet's democratizing potential and the skeptical view of pioneers like Norris (2001), who warned of the risk that technology could widen inequalities and create democratic exclusion. This theoretical tension materialized in Barack Obama's " *Politics 2.0* " community mobilization campaigns in 2008, and in the aggressive models of Donald Trump in 2016 and Jair Bolsonaro in 2018, who moved toward increasingly direct and confrontational dialogue with their bases, culminating in the use of closed networks like *WhatsApp* (Silva, 2016). At the heart of this transformation, Zuboff (2019) defines surveillance capitalism as an economic model that extracts behavioral data to fuel psychographic *microtargeting techniques* , enabling the delivery of persuasive messages that exploit psychological vulnerabilities, as observed in the *Cambridge Analytica scandal* , documented in the work " *Understanding Mass Influence : a case study of Cambridge Analytica as a contemporary mass influence campaign* " (Webb et al, 2021). Furthermore, this system is fueled by viral culture, where political memes function as effective rhetorical artifacts that use humor and emotion as informal campaign pieces (Chagas, 2017).

This dynamic is fueled by organized structures such as "digital militias," which undermine debate with disinformation, highlighting a structural transformation of politics that demands new forms of governance and prompts distinct regulatory responses, such as the judicial-led model in Brazil, in contrast to the European Union's legislative framework (Gomes, 2020). Thus, the trajectory of digital political communication can be understood as a shift from a promise of expanded participation to a reality of surveillance and manipulation, requiring democracies to create new accountability mechanisms to defend the public sphere.

3.2 Political marketing on social networks

The transition of political marketing to the digital environment represents a fundamental reconfiguration of power, aligned with Castells's (1999) theory of the network society, where the ability to manage information flows becomes central. This shift, driven by algorithms, favors personalization and segmentation, as seen in the use of Big Data and psychographic microtargeting, a practice that gained notoriety with the Cambridge Analytica scandal and was already highlighted as a trend in Brazil by researchers such as Antoniutti (2015). Brazilian strategists such as Torquato (2004) advocate the need for a "single message" and professionalism, while Vitorino (2020) highlighted the importance of a tripod that integrates politics, communication, and technology, prioritizing long-term reputation building. However, international literature, represented by authors such as Bimber and Davis (2003), demonstrates that the internet functions less as a space for converting opponents and more as an echo chamber for reinforcing beliefs and mobilizing existing bases, which Kakutani (2018) defines as a phenomenon enhanced by "filter bubbles" and "echo chambers".

The impact of these strategies was evident in the 2018 Brazilian elections, which marked the rise of disruptive digital campaigns, and was consolidated in 2022 with the professionalization and massive increase in investment in online advertising. Phenomena such as the "Hate Cabinet" illustrate the instrumentalization of disinformation as a deliberate tactic to discredit opponents and build the image of enemies (Melo, 2020). In response, Brazil adopted a proactive and interventionist regulatory stance through the Superior Electoral Court (TSE), banning *deepfakes* and non-consensual mass shootings, a model that contrasts with the primacy of freedom of expression in the US and the European Union's focus on data transparency. This tension reflects what Castells (2013) described as a symptom of the crisis of representative democracy, where distrust in traditional institutions opens space for the emergence of networked movements. The evolution of the threat, from textual disinformation to synthetic audiovisual manipulation, demonstrates that the challenge transcends technological regulation, pointing to a deeper crisis of trust that demands solutions focused on media literacy and the strengthening of democratic institutions themselves.

3.3 Data Science, Machine Learning, and Data Mining

The technical-scientific framework of this research distinguishes Artificial Intelligence (AI) as a broad field, Machine Learning (ML) as its subset that enables systems to learn from data, and Data Science as the interdisciplinary discipline that extracts knowledge. The central process adopted is Data Mining, which corresponds to the "knowledge discovery from data"

(KDD) stage and uses algorithms to find patterns in large datasets (Han et al., 2011). For profile segmentation, cluster analysis, an unsupervised learning technique that groups data with similar characteristics, is used. The effectiveness of this multivariate and behavioral approach, superior to simplistic segmentations, was demonstrated in a retail case study by Falqueto (2022), whose methodological model is proposed for transfer to the political domain. Automation of this process, known as *AutoML*, for clustering, however, faces the challenge of lacking ground truth labels, making validation difficult.

To overcome this barrier, the literature suggests the use of internal Clustering Validation Indices (CVIs), which measure the quality of clusters based on their intrinsic properties (Rozgonjuk, 2023). Advanced *AutoML* approaches , such as *AutoCluster*, solve the algorithm selection and hyperparameter optimization (*CASH*) problem by combining multiple CVIs , such as the *Silhouette Score* , the *Davies- Bouldin Index* , and the *Calinski-Harabasz Index*, in an evaluation committee. Inspired by these strategies, this work developed its own implementation that also uses an ensemble of CVIs to ensure robust and generalizable optimization in the search for the best segmentation.

4. Methodology

To answer the research question — "How can the application of clustering techniques to behavioral data extracted from social media contribute to improving segmentation in digital political marketing, enabling the identification of groupings of electoral profiles?" — this study adopts a quantitative, applied approach with a deductive methodological orientation. The choice of the quantitative approach is justified by the use of statistical and machine learning techniques to measure and analyze structured numerical data, as argued by Gil (2008). Its applied nature stems from the intention to offer a concrete solution to a real-world problem, promoting the practical use of scientific knowledge in the field of digital political marketing (Thiollent , 2018). The research focuses on the automated collection of public data from the official profiles of the 27 governors of the federative units of Brazil on the Instagram platform, during the period from June to July 2025. The analysis was conducted through clustering techniques using the *AutoCluster framework* , which combines meta-learning and internal validation to automate the selection and combination of clustering algorithms (Liu et al., 2021).

To this end, data collection in this research reflects a systematic process of data acquisition through web scraping, that is, data scraping from virtual environments. It offers a characterization of the textual corpus that constitutes the empirical basis of this study, addressing its source, structure, and scale. Data collection was performed programmatically using the APIFY platform, a tool specialized in automated data extraction via scrapers and application programming interfaces (APIs). Three different actors were used within this tool: "Instagram Profile Scraper," "Instagram Reels Scraper," and "Instagram Posts Scraper." The scripts, as well as all request configurations, are in the "Data Collection" section of the "analises.ipynb" file attached to this work. Data related to profile links were collected manually and saved in the "governadores.xlsx" spreadsheet, also attached. The scope of the collection covered data from the last 30 feed and reel publications , between June and July 20025, ensuring national coverage of the phenomenon under study.

The raw data was extracted and organized into three distinct JSON files, each corresponding to the output of a specific actor, with the file name serving as the actor's identifier. The files were: profiles.json , posts.json , and reels.json . The appendix to this work contains the file "Dicionário de valores.xlsx" (Variable Dictionary.xlsx) with information regarding the source, name, and description of all attributes collected and used by the API, separated by source file, along with the developed codes, as well as the files extracted from the API for analysis.

The clustering approach was divided into two levels of analysis: the strategic unit (the governors' profiles) and the content unit (their individual publications).

To define user profiles, four distinct clusterings were performed. The first two focused on performance and volume metrics (likesSum, commentsSum) and efficiency (% engagement), justified by the need to measure direct resonance with the audience and, crucially, normalize performance by audience size to differentiate volume from efficiency, avoiding the pitfalls of focusing on single metrics (Falqueto et al., 2022). The following analysis incorporated a temporal dimension with " recency" and "frequency," chosen to capture the dynamics and consistency of the communication strategy, essential in modern marketing that aims to track the entire audience journey (Kotler et al., 2017). Finally, the fourth profile analysis used structural variables (followersCount , followsCount , postsCount) to segment actors based on their "posture" or growth philosophy on the network, distinguishing organic attraction strategies from those focused on active community building. The decision to shift the focus of the analysis from profiles to individual posts was justified by the high volatility in profile-level engagement data, which could mask the performance of different types of content. The first clustering of posts used videoPlayCount , likesCount , and commentsCount, with the inclusion of video views being essential for assessing the reach of formats like Reels , which are central to contemporary political communication. This combination allows us to differentiate content with broad reach but shallow engagement from high-impact "blockbusters." The latest analysis, focused exclusively on likesCount and commentsCount, is justified by the search for a more granular understanding of the nature of audience interaction. The goal was to isolate the type of response generated, regardless of reach, allowing, for example, to distinguish content that is massively " liked " from content that, although less popular, incites deeper debate through comments, thus offering a more refined layer of strategic intelligence on the message's impact.

5. Analysis of Results

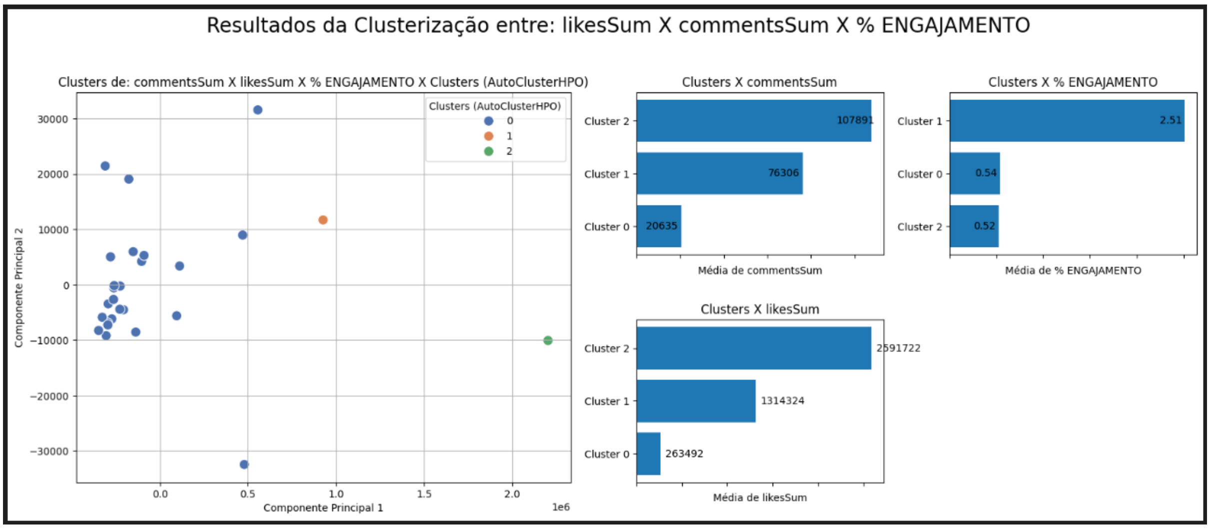
5.1 Profile-level clustering

In this section, we will present the detailed results of iterative cluster analyses using the AutoCluster framework and conducted with the aim of proposing this specific methodology as a method for segmenting voters in the context of political marketing, using a methodology that combines automated segmentation algorithms with visualization techniques, such as Principal Component Analysis (PCA) and direct scatter plots.

5.1.1 Clustering by likes, comments and engagement percentage

The initial clustering analysis revealed three distinct performance profiles for the governors. Cluster 0 groups the vast majority, with 25 profiles, forming a block of average performance, with an average engagement rate of 0.54% and moderate volumes of interactions. In stark contrast, Clusters 1 and 2 represent outliers. The main difference between them is the nature of their performance: Cluster 1 stands out for its efficiency, with an engagement rate (2.51%) almost five times higher than the average, while Cluster 2 stands out for its massive volume, accumulating a number of likes and comments drastically higher than the others, despite maintaining an engagement rate similar to that of the main group. The only notable similarity is the percentage engagement rate between Cluster 0 and Cluster 2, indicating that, although their audiences have completely different scales, the proportion of followers who interact is similar. Figure 1 provides a visual representation of the clusters.

Figure 1. Clustering of Profiles by Volume Metrics and Percentage of Governor Engagement



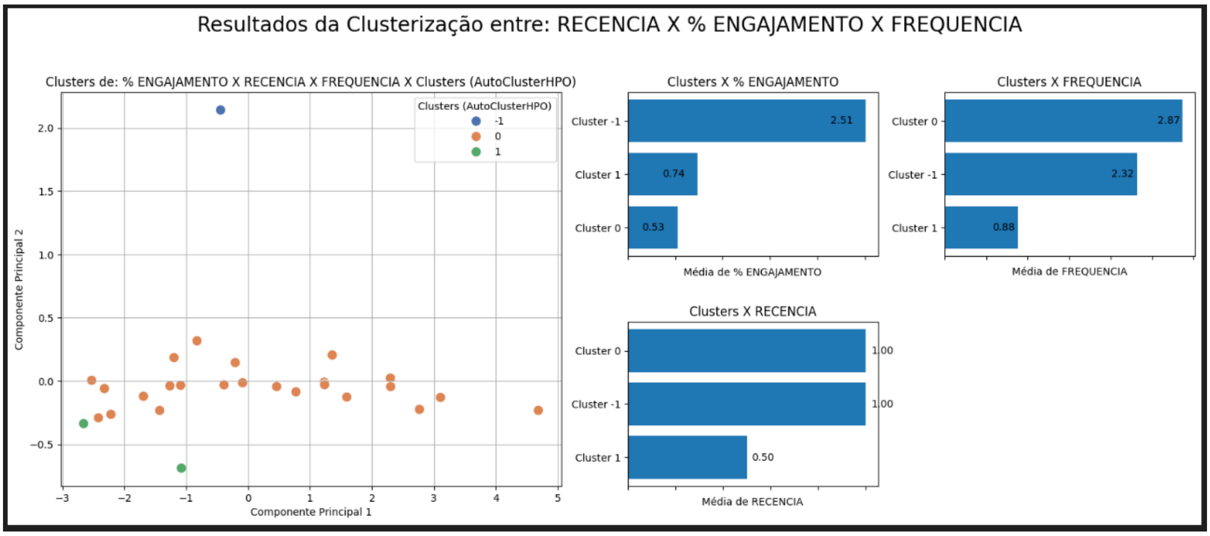
Based on the analysis, we can personify the three groups distinctly to understand their digital strategies. Cluster 0 can be characterized as "The Institutional Manager": it represents the archetype of the governor whose digital presence is formal and consistent, maintaining average and predictable performance, similar to most of their peers, focusing on communicating government actions without necessarily generating peaks in interaction. Cluster 1, composed solely of Jorginho Mello, governor of Santa Catarina, is "The Engagement Phenomenon": this unique profile demonstrates an exceptional ability to mobilize its base, justified by an engagement rate almost five times higher than the others, indicating that its content creates a deep and highly resonant connection with the public. Finally, Cluster 2, composed exclusively of São Paulo governor Tarcísio de Freitas, is "The Audience Giant": justified by his overwhelming volume of likes and comments, this governor operates on a massive communication scale; Although his engagement rate is average, the size of his audience is so vast that it makes him a powerhouse of reach, capable of amplifying his message to millions of people.

Academic literature supports the existence of three digital political personas with distinct theoretical foundations. The "Institutional Manager" finds support in the traditional communication model, which conceives communication as an instrument for legitimizing a standardized logic and ensuring the efficiency of the governmental process (Silva, 2020). This persona archetypically aligns with the figure of the "Father" or "Caregiver," who projects an image of security, experience, and responsibility, seeking to calm and guide the population rather than mobilize them for conflict (Schwartzenberg, 1978). In contrast, the "Engagement Phenomenon" is anchored in the concept of digital populism, which uses platforms to create a form of "direct representation" with the grassroots, bypassing traditional institutions (Urbinati, 2019). This strategy exploits the affective economy of networks to nurture "echo chambers" and mobilize emotions, which finds an "elective affinity" with populist logic (Gerbaudo, 2018) and often manifests itself through the "Hero" and "Outlaw" archetypes, which promote a narrative of rupture and "Cultural War" against an established system (Silva, 2020). Finally, the "Audience Giant" operates under the logic of the "Spectacle State" (Schwartzenberg, 1978), where politics becomes a performance and the leader becomes a "superstar." In this approach, reach and volume metrics, often classified as "vanity metrics" in marketing literature (Ries, 2011), are transmuted into political capital, functioning as massive social proof and a demonstration of power aligned with the "Ruler" archetype.

5.1.2 Clustering by recency, frequency and engagement percentage

Based on the data, the second cluster analysis allowed the governors to be segmented into three groups with distinct digital strategies. The main difference lies in the relationship between posting frequency and engagement performance: Cluster 0, the largest group with 24 profiles, adopts a high-volume strategy, with the highest posting frequency for standard engagement; Cluster 1, with two profiles, represents a "quality over quantity" approach, with low frequency and less recent activity, but which achieves above-average engagement when it does post; and Cluster -1, a single outlier, is the very high-impact profile, which, with moderate frequency, achieves an exceptional engagement rate, much higher than the others. The most notable similarity is that the vast majority of governors (Clusters 0 and -1) remain highly active, suggesting that, regardless of performance, recency is a common communication practice for most state executives.

Figure 2. Clustering of Profiles by Frequency, Recency and Engagement (%) in Publications



Based on data analysis and visualization, we can personify the clusters to reflect their distinct communication strategies. Cluster 0 is the "Ubiquitous Communicator": grouping most governors, this persona is justified by the higher frequency of posts and constant activity (maximum recency), adopting a volume strategy to ensure continuous presence in followers' feeds, even if this results in only average engagement. In contrast, Cluster 1, composed of Acre Governor Gladson Cameli and Rondônia Governor Marcos Rocha, represents "The Sharpshooter": justified by the low frequency and lower recency, this profile relies on precision, as when they post, they achieve above-average engagement, suggesting a "quality over quantity" tactic, where each post is a well-planned, accurate shot. Finally, Cluster -1, composed of Governor Jorginho Mello, is the "Strategic Influencer": this exceptional profile is justified by dominating all fronts, combining high frequency and constant activity with an extraordinarily high engagement rate, which characterizes him as a master of the platform, capable of maintaining volume and, at the same time, generating a massive and influential impact.

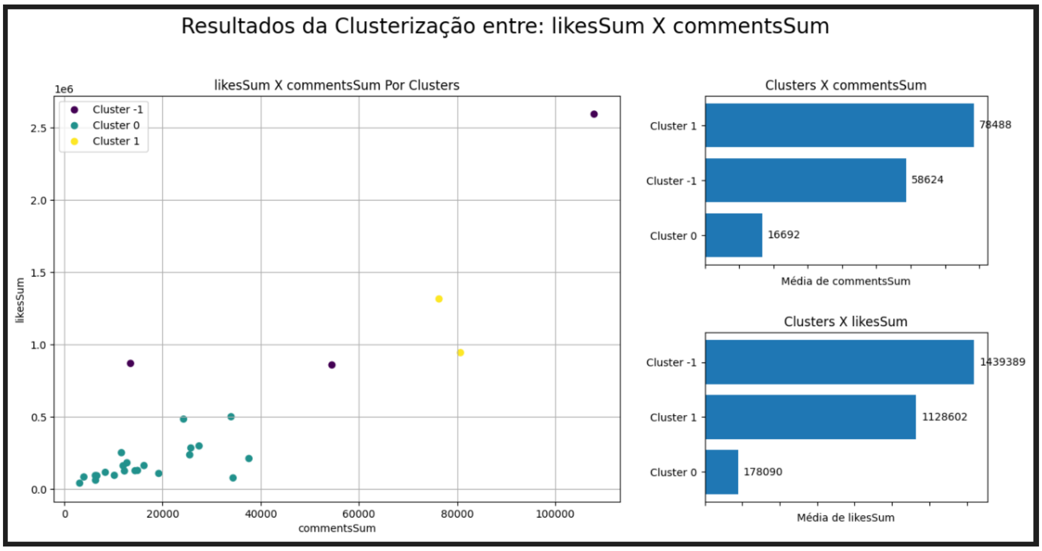
Academic literature offers theoretical support for the existence of three digital political communication personas, each anchored in different strategies and power logics. The "Ubiquitous Communicator" is grounded in the "permanent campaign" theory, which describes the need for continuous, high-volume communication to maintain visibility and public support, treating governance as an extension of the electoral campaign (Domalewska , 2018). In contrast, the "Sharpshooter" operates under the logic of "quality over quantity," prioritizing efficiency measured by engagement rate over frequency, where each post is planned to generate maximum impact and resonance with a specific audience, a distinction validated by analyses that differentiate the total volume of interactions from their relative effectiveness (Pinheiro, 2024). Finally, the "Strategic Influencer" combines volume and impact, finding support in the concept of digital populism, which explores direct and affective communication to bypass traditional media (Urbinati, 2019; Gerbaudo, 2018), and in the theory of the "Spectacle State", where the leader's personality and performance overlap with political substance (Schwartzenberg, 1978), often using moral and "Cultural War" agendas to maximize mobilization (Silva, 2020).

5.1.3 Clustering by likes and comments

The clustering analysis by likes and comments divided the governors into three groups based strictly on the volume of interactions, as illustrated in Figure 3.

Based on the analysis of interaction volume, clusters can be personified to highlight their different forms of digital impact. Cluster 0 is "The Regional Communicator." This persona represents the vast group of governors with standard performance, justified by a solid but moderate volume of likes and comments, characteristic of institutional communication with strong reach within their own state but lacking national viral reach. At an elite level, we have two distinct personas: Cluster -1, composed of governors Tarcísio de Freitas, Marcos Rocha, and Ronaldo Caiado, is "The Popularity Champion," justified by having the highest average number of likes (likesSum), indicating a communication strategy with broad and mass appeal, extremely effective in generating approval and reach. In turn, Cluster 1, composed of governors Romeu Zema and Jorginho Mello, is "The Catalyst of Debates", a persona justified by having the highest average number of comments (commentsSum), which demonstrates a unique ability to create content that provokes dialogue, engages the public in discussions and fosters a deeper level of interaction.

Figure 3. Content Clustering by Volume of Likes and Comments



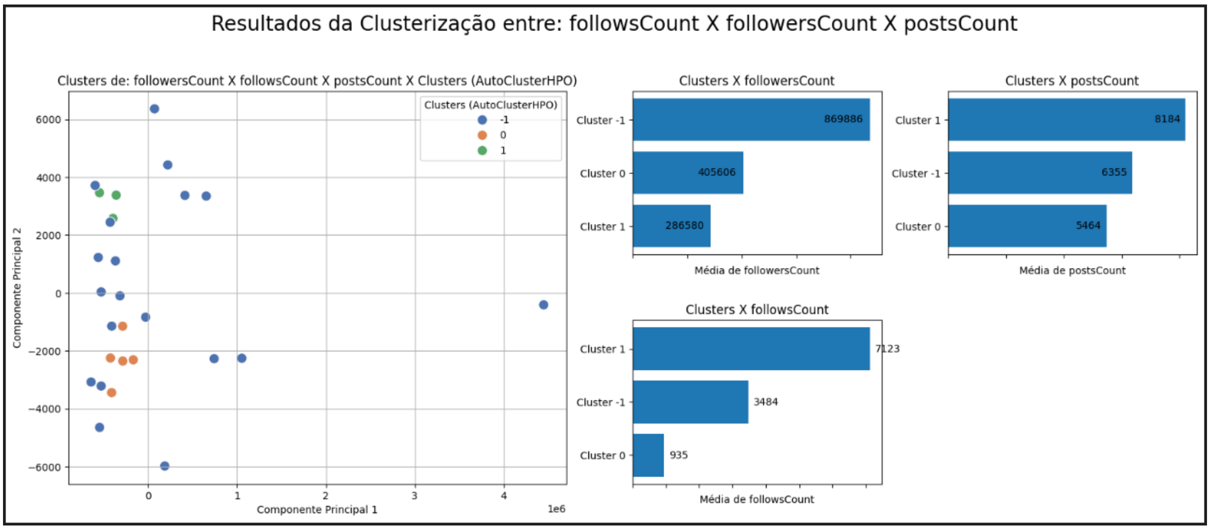
Academic literature provides a solid theoretical foundation for the three digital communication personas, differentiating them by their strategic models and the type of engagement they prioritize. The "Regional Communicator" is anchored in the traditional institutional communication model, which uses digital platforms to fulfill transparency and accountability mandates, resulting in a solid but geographically limited reach. In contrast, the "Popularity Champion" operates under the logic of digital populism, seeking a direct and affective connection with the public (Urbinati, 2019). Their strategy consists of maximizing "thin" engagement. through a massive volume of likes, which function as mass approval signals and validate power archetypes such as the "Hero" or the "Father" (Leighninger & Nabatchi, 2015; Schwartzenberg, 1978). Finally, the "Debate Catalyst" employs polemical discourse to generate "thick" engagement, measured by the high volume of comments (Leighninger & Nabatchi, 2015). This tactic transforms the comments section into an arena of debate and controversy, whose ambivalent nature, which can indicate both support and attack, requires qualitative analysis to be fully understood (Araújo et al., 2017).

5.1.4 Clustering by followers, following and publications

Based on the data, clustering revealed three distinct profile strategies among the governors. Figure 4 shows that the main difference between the groups lies in the "following" policy: Cluster 1 groups "aggressive" profiles, with the fewest followers but who follow the most accounts and post most frequently, suggesting a growth tactic; Cluster 0 represents "curated" profiles, with a solid follower base but who follow very few accounts, adopting a more formal stance; and Cluster -1, the largest group, is the most heterogeneous, containing profiles ranging from medium-sized audiences to the most popular in the country. The most notable similarity is that all clusters demonstrate a high volume of publications (with average postCounts between 5.4k and 8.2k), indicating that intense and constant content production is a common practice for all, regardless of growth strategy or audience size.

Based on the analysis of the profiles in Figure 4, the clusters reveal three personas with distinct positioning strategies. Cluster 0 is "The Official Channel": justified by the low number of accounts it follows in contrast to a solid follower base, this persona represents a unilateral and formal communication profile, positioning itself as a source of information, without employing follow-to-growth tactics.

Figure 4. Profile Clustering by Network Structure: Followers, Following, and Publication Volume



In contrast, Cluster 1, composed of governors Gladson Cameli, Elmano Freitas, and Rafael Fonteles, is "The Audience Builder." This persona is justified by having the smallest follower base but being the most active, with the highest volume of posts and, crucially, the largest number of followed accounts, indicating an aggressive growth strategy to expand its visibility and attract new followers. Finally, Cluster -1, the largest and most diverse , is "The Established Politician." This persona represents the mainstream scenario, grouping together most governors who already have a consolidated and diverse audience, maintaining a strong digital presence without the need to adopt the niche strategies of the other groups.

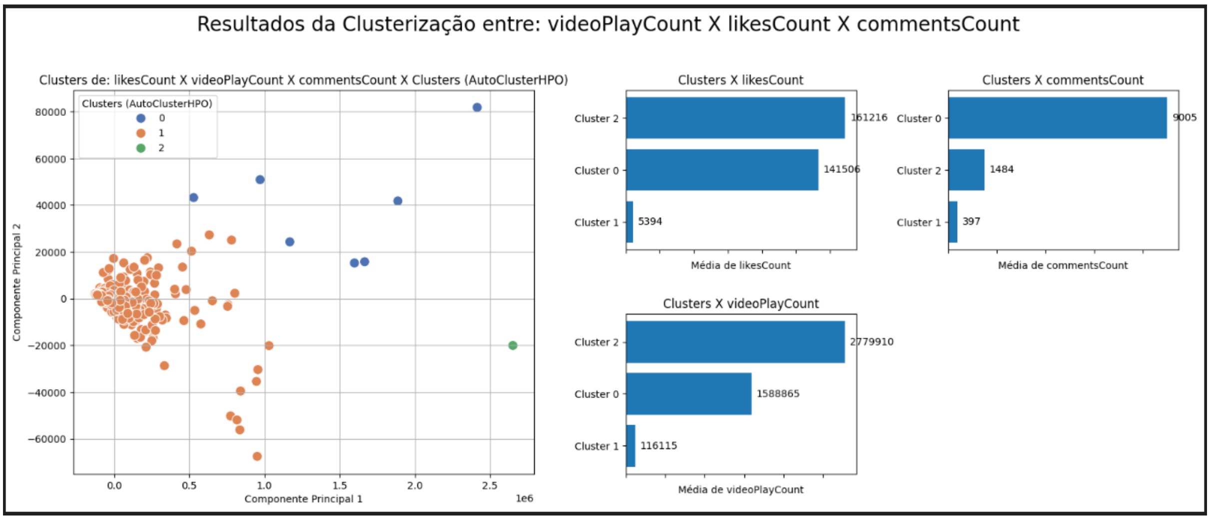
The academic literature on digital political communication provides a robust theoretical foundation for the three positioning personas. "The Official Channel" embodies the Normalization Hypothesis, in which offline power is replicated online (Margolis & Resnick , 2000), using unilateral and authoritative communication to project status, evidenced by a low number of followed accounts in contrast to a large follower base. In contrast, "The Audience Builder" embodies the Equalization Hypothesis, employing aggressive growth tactics to level the playing field (Petrova et al., 2021); their strategy is defined by an active pursuit of reciprocity, such as following a high number of profiles to attract followers, and a high frequency of posts to maximize visibility (Lewis et al., 2014). Finally, "The Established Politician" operates under the logic of the "permanent campaign" (Domalewska , 2018), using a hybrid communication model to maintain the relevance of its already consolidated audience, frequently resorting to digital populism tactics and the mobilization of affective engagement to keep its base cohesive and active (Urbinati, 2019; Brady et al., 2017).

5.2 Publication-level clustering

5.2.1 Clustering by views, likes and comments

The clustering analysis by views, likes, and comments, presented in Figure 5, separated the posts into three well-defined performance levels. The main difference between them is the scale and type of success: Cluster 1 represents the vast majority of posts (802 posts), with standard to moderate performance across all metrics, forming the baseline. In contrast, the other two clusters are high-performing but distinct in nature: Cluster 0 is a group of seven highly successful posts, balanced, with excellent average views, likes, and comments, while Cluster 2 is a single "viral" post that, despite having the highest number of video views and many likes, generated fewer comments than the average for Cluster 0. The similarity, therefore, lies between Clusters 0 and 2, as both represent a level of "successful" content, clearly distinguishing themselves from the standard performance of Cluster 1, functioning as the "hits" of the sample.

Figure 5. Clustering of Publications by: Views, Likes, and Comments

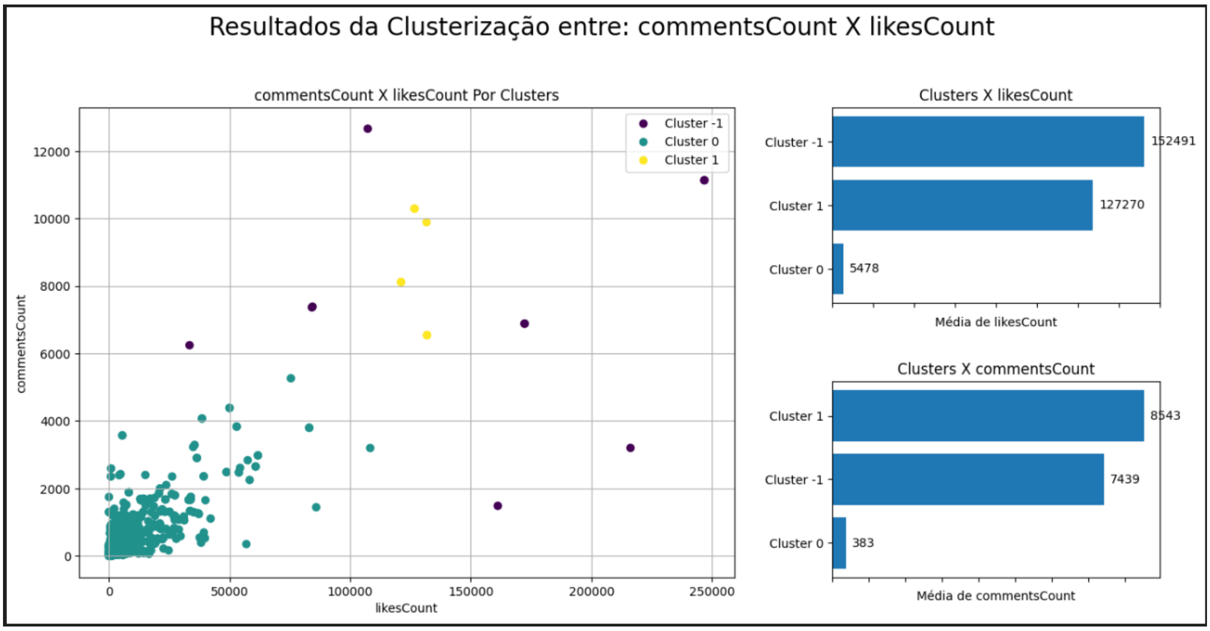


Based on the performance analysis of posts, clusters can be personified to illustrate the different types of content. Cluster 1, which encompasses the overwhelming majority of posts, is "Routine Content": justified by its modest and consistent performance metrics, this persona represents everyday posts that keep the profile active but without generating performance spikes. At an elite level, two successful personas can be distinguished: Cluster 0 is "Engagement Success," justified by its high performance across all metrics, but mainly by having the highest average number of comments, characterizing content that not only reaches many people but also stimulates deep debate. Finally, Cluster 2 is "The Viral Phenomenon," a single outlier post justified by astronomical numbers of views and likes, but with fewer comments than the "Engagement Success," representing content that went viral massively in terms of reach but with less depth of interaction.

5.2.2 Clustering by likes and comments

The clustering analysis by likes and comments, shown in Figure 6, divided the posts into a large group of standard performance and two distinct groups of high performance. The main difference lies in the scale of engagement and internal consistency of the successful clusters: Cluster 0 represents the vast majority of posts (1,603) with modest and routine performance. In contrast, Clusters -1 and 1 are both high performers, but Cluster 1 (7 posts) is an extremely consistent group, with low internal variation and the highest average number of comments, while Cluster -1 (10 posts) is a more heterogeneous group of success, with a much greater variation in results. The fundamental similarity, therefore, is that Clusters -1 and 1 function as the "upper floor" of performance, representing the highly successful posts that clearly stand out from the standard performance observed in Cluster 0.

Figure 6. Clustering of Publications by Interactions: Comments and Likes (commentsCount × likesCount)



Based on performance analysis, publications can be personified into three distinct categories that reveal their strategic functions. Cluster 0, comprising the vast majority of posts, is "Routine Communication": justified by its modest engagement numbers, it represents everyday content that keeps the profile active and fulfills an institutional role without generating spikes in interaction. At an elite level, we have two successful personas: Cluster 1 is "Consistent Success," justified by its high engagement, leading comment count, and, crucially, low internal variation, which characterizes it as a type of content with a predictable and reliable formula for generating debate. In contrast, Cluster -1 is "The Unpredictable Hit," a persona justified by its also high average performance, leading in likes, but with very large variation among its members. It represents high-impact publications, but whose level of success is more volatile and less replicable.

6. Conclusion/Contribution

This study demonstrated that cluster analysis enhances political marketing segmentation, overcoming the obsolescence of purely demographic methods in the "Network Society" era. The research validated clustering as a robust methodological solution, replacing predefined categories with behavioral archetypes empirically validated based on online voter behavior. The methodology revealed previously invisible segments. These distinctions offer a practical framework for optimizing resources and personalizing campaign communication. Although the analysis is limited to the Instagram platform and quantitative metrics, future research is encouraged to expand the methodology to other networks. The incorporation of Natural Language Processing (NLP) is suggested for sentiment analysis, inauthentic behavior detection, and topic modeling, seeking a more complete, qualitative, and strategic understanding of the digital electorate and its perceived benefits. Regarding impacts, the research contributes to the advancement of political targeting strategies by proposing a model based on empirical evidence extracted from digital behavior, overcoming the limitations of fixed typologies. By integrating data science into political marketing, it offers a replicable framework for communication personalization in campaigns. Its findings can support more effective, ethical, and data-driven decisions, strengthening strategic intelligence in the digital public sphere.

7. Bibliographic References

ALMEIDA, Débora VLC Rezende de. Political Marketing, History and Evolution in Republican Brazil. *In* : BRAZILIAN CONGRESS OF COMMUNICATION AND POLITICAL MARKETING, 2007, Belo Horizonte. Proceedings [...]. Belo Horizonte: Compolítica , 2007. Available at: http://compolitica.org/novo/anais/2007_gt_jmp-debora.pdf . Accessed on: July 10, 2025.

ARAÚJO, Ronaldo Ferreira; TRAVIESO-RODRIGUEZ, Crispulo ; SANTOS, Sarah Rúbia de Oliveira. Communication and political participation on Facebook: analysis of comments on Brazilian parliamentarians' pages. Information & Society: Studies, v. 27, n. 2, p. 279-290, 2017. Available at: http://eprints.rclis.org/38910/1/comunica%C3%A7%C3%A3o%20pol%C3%ADtica.pdf Accessed on: July 10, 2025.

BESSI, A., & FERRARA, E. (2016). Social bots distort the 2016 US Presidential online election discussion . First Monday , 21(11). https://doi.org/10.5210/fm.v21i11.7090

BIMBER, Bruce; DAVIS, Richard. Campaigning Online: The Internet in US Elections . New York: Oxford University Press, 2003. Available at: https://books.google.com.br/books/about/Campaigning_Online.html?id=pRtCg2T-1c8C&redir_esc=y Accessed on: July 10, 2025.

BRADY, William J. et al. Emotion shapes the diffusion of moralized content in social networks. Proceedings of the National Academy of Sciences , v. 114, n. 28, p. 7313-7318, 2017. Available at: https://www.pnas.org/doi/10.1073/pnas.1618923114 Accessed on: July 10, 2025.

CASTELLS, Manuel. The network society. São Paulo: Paz e Terra, 1999. Available at: https://csociais.wordpress.com/wp-content/uploads/2018/05/castellsm-a-sociedade-em-rede-cap-6.pdf Accessed on: July 10, 2025.

CASTELLS, Manuel. Networks of indignation and hope: social movements in the internet age. Translated by Carlos Alberto Medeiros. Rio de Janeiro: Zahar, 2013. Available at: https://www.scielo.br/j/ha/a/YSF98JjdMdSkwD8Khcs9sdp/?lang=pt# Accessed on: July 10, 2025.

CHAGAS , Viktor, et al. “The Politics of Memes and the Memes of Politics: Methodological Proposal for Content Analysis of Memes from the 2014 Electoral Debates”. Intexto , nº 38, January 2017, p. 173-96, doi:10.19132/1807-8583201738.173-196. Available at: https://seer.ufrgs.br/index.php/intexto/article/view/63892 Accessed on: July 10, 2025.
CRITTENDEN , VL, Crittenden , WF, & Muzyka , DF (2002). Segmenting the business- to - business marketplace by product attributes and the decision process . *Journal of Strategic Marketing* , *10* (1), 3–20. Available at: https://www.tandfonline.com/doi/abs/10.1080/09652540110100448 Accessed on: July 10, 2025.
DE SLEGTE, J., VAN DROOGENBROECK, F., SPRUYT, B., VERBOVEN, S., & GINIS, V. (2024). The Use of Machine Learning Methods in Political Science: An In-Depth Literature Review. Political Studies Review, 0(0). https://doi.org/10.1177/14789299241265084
DOMALEWSKA, Dorota . The permanent campaign in social media: A case study of Poland . *In* : PROCEEDINGS OF THE CEE E|DEM AND E|GOV DAYS, 2018. Available at: https://ejournals.facultas.at/index.php/ocgcp/article/view/1669 Accessed on: July 10, 2025.

GERBAUDO, Paolo. Social media and populism : an elective affinity ? Media, Culture & Society, v. 40, n. 5, p. 745-753, 2018. Available at: https://journals.sagepub.com/doi/10.1177/0163443718772192 Accessed on: July 10, 2025.
GIL, Antonio Carlos. Social Research Methods and Techniques. 6th ed. São Paulo: Atlas, 2008. Available at: https://ayanrafael.com/wp-content/uploads/2011/08/gil-ac-mc3a9todos-e-tc3a9cnicas-de-pesquisa-social.pdf Accessed on: July 10, 2025.
GOMES, Wilson. Chronicle of a Tragedy Foretold : How the Far Right Came to Power. São Paulo: Summus , 2020. Available at: https://www.amazon.com.br/Cr%C3%B4nica-uma-trag%C3%A9dia-anunciada-extrema-direita/dp/658655506X Accessed on: July 10, 2025.
HALEY, RI Benefit Segmentation : A Decision-Oriented Research Tool. Journal of Marketing, v. 32, n. 3, p. 30–35, Jul. 1968. Available at: https://journals.sagepub.com/doi/10.1177/002224296803200306 Accessed on: July 10, 2025.
HAN, J.; KAMBER, M.; PEI, J. Data Mining: Concepts and Techniques . 3rd ed. Waltham : Morgan Kaufmann, 2011. Available at: https://www.amazon.com.br/Data-Mining-Techniques-Jiawei-Han/dp/0123814790 Accessed on: July 10, 2025.
KAKUTANI, Michiko. The Death of Truth: Notes on Falsehood in the Age of Trump. Translated by André Czarnobai . Rio de Janeiro: Intrínseca, 2018. Available at: https://www.amazon.com.br/MORTE-VERDADE-Notas-sobre-mentira/dp/855100364X Accessed on: July 10, 2025.
KOTLER, Philip; KARTAJAYA, Hermawan ; SETIAWAN, Iwan. Marketing 3.0: The Forces Shaping the New Human-Centered Marketing. Rio de Janeiro: Elsevier, 2010. Available at: https://biblioteca.unisced.edu.mz/bitstream/123456789/2103/1/PHILIP%20KOTLER-AS%20FORCAS%20QUE%20ESTAO%20DEFININDO%20NOVO%20MERC..pdf Accessed on: July 10, 2025.
KOTLER, Philip; KARTAJAYA, Hermawan ; SETIAWAN, Iwan. Marketing 4.0: from traditional to digital. Rio de Janeiro: Sextante, 2017. Available at: https://www.drnishikantjha.com/booksCollection/Marketing%204.0_%20Moving%20from%20Traditional%20to%20Digital%20(%20PDFDrive%20).pdf Accessed on: July 10, 2025.
LEIGHNINGER, Matt; NABATCHI, Tina. Public Participation for 21st Century Democracy . San Francisco: Jossey -Bass, 2015. Available at: https://books.google.com.br/books/about/Public_Participation_for_21st_Century_De.html?id=A3fknAEACAAJ&redir_esc=y Accessed on: July 10, 2025.

LEWIS, Seth C.; HOLTON, Avery E.; CODDINGTON, Mark. Reciprocal journalism . Journalism Practice , v. 8, n. 2, p. 229–241, 2014. Available at: https://www.tandfonline.com/doi/abs/10.1080/17512786.2013.859840 Accessed on: July 10, 2025.
FALQUETO, AA; CEZAR, RTA Segmentation via machine learning: proposal for clustering e-commerce consumers of a multinational sports retail company. Holos , year 38, v. 4, e12032, 2022. Available at: https://www2.ifrn.edu.br/ojs/index.php/HOLOS/article/view/12032 Accessed on: July 10, 2025.
LIU, Y.; LI, S.; TIAN, W. AutoCluster : Meta-learning Based Ensemble Method for Automated Unsupervised Clustering . *In* : KARLAPALEM, K. et al. (Eds.). Advances in Knowledge Discovery and Data Mining. Call : Springer International Publishing , 2021. v. 12714, p. 246–258. (Lecture Notes in Computer Science). Available at: https://link.springer.com/chapter/10.1007/978-3-030-75768-7_20 Accessed on: July 10, 2025.
MARGOLIS, Michael; RESNICK, David. Politics as Usual: The Cyberspace "Revolution". Thousand Oaks: SAGE Publications , 2000. Available at: https://sk.sagepub.com/book/mono/politics-as-usual/toc Accessed on: July 10, 2025.

MELLO, Patrícia Campos. The Hate Machine : A Reporter's Notes on Fake News and Digital Violence. São Paulo: Companhia das Letras, 2020. Available at: https://www.amazon.com.br/m%C3%A1quina-%C3%B3dio-rep%C3%B3rter-viol%C3%AAncia-digital/dp/853593362X Accessed on: July 10, 2025.
NORRIS, Pippa . Digital Divide: Civic Engagement , Information Poverty , and the Internet Worldwide . Cambridge: Cambridge University Press, 2001. Available at: https://www.cambridge.org/core/books/digital-divide/20EFED5574695AC79D1BBC6E295B1EC0 Accessed on: July 10, 2025.
PAPAKYRIAKOPOULOS, O., HEGELICH, S., SHAHREZAYE, M., & MEDINA SERRANO, J. C. (2018). Social media and microtargeting: Political data processing and the consequences for Germany . *Big Data & Society, 5* (2), 1–15. https://doi.org/10.1177/2053951718811844
PETROVA, Maria; SEN, Arijit ; YILDIRIM, Pinar. Social Media and Political Contributions : The Impact of New Technology on Political Competition . Management Science, 2021. Available at: https://faculty.wharton.upenn.edu/wp-content/uploads/2016/11/10_Social-Media-and-Political-Contributions-The-Impact-of-New-Technology-on-Political-Competition.pdf Accessed on: July 10, 2025.
RIES, Eric. The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Build Wildly Successful Companies. São Paulo: Leya, 2011. Available at: https://s-inova.ucdb.br/wp-content/uploads/biblioteca/a-startup-enxuta-eric-ries-livro-completo.pdf Accessed on: July 10, 2025.
ROZGONJUK, D. Towards Automated Machine Learning: Hyperparameter Optimization in Online Clustering . 2023. Dissertation (Master in Data Science) - Institute of Computer Science, University of Tartu, Tartu, 2023. Available at: https://dspace.ut.ee/server/api/core/bitstreams/080dee2d-ca30-4d20-87b0-9559daaf7df4/content Accessed on: July 10, 2025.
SCHWARTZENBERG, Roger-Gérard. The Spectacle State. Rio de Janeiro: Difel, 1978. Available at: http://dspace.sti.ufcg.edu.br:8080/xmlui/bitstream/handle/riufcg/34380/O%20ESTADO%20ESPET%C3%81CULO%20-%20A%20OBRA%20DE%20ROGER-G%C3%89RARD%20-%20EVENTO%20II%20COLOQUIO%20INT.%20DE%20HIST%C3%93RIA%20GT%2021%20%202010.pdf?sequence=1&isAllowed=y Accessed on: July 10, 2025.

SILVA, Rafael Rocha Alves da. Bolsonarism, elections and social networks: an analysis of the discursive strategies of PSL deputies with super votes on Facebook. 2020. 130 p. Dissertation (Master's in Communication) – Federal University of Ceará, Fortaleza, 2020. Available at: https://repositorio.ufc.br/handle/riufc/57119 Accessed on: July 10, 2025.
SILVA, Sivaldo Pereira da; BRAGATTO, Rachel Callai; SAMPAIO, Rafael Cardoso (Orgs .). Digital Democracy, political communication and networks: theory and practice. Rio de Janeiro: Folio Digital, 2016. Available at: https://abcpublica.org.br/wp-content/uploads/2020/10/Democracia-Digital-comunica%C3%A7%C3%A3o-politica-e-redes-2-_compressed-2.pdf Accessed on: July 10, 2025.
TORQUATO, Gaudêncio. Political and Government Marketing: A Roadmap for Campaigns and Strategies. São Paulo: Summus , 2004. Available at: https://www.amazon.com.br/Marketing-Pol%C3%ADtico-Governamental-Francisco-Gaud%C3%AAncio/dp/8532302297 Accessed on: July 10 , 2025
THIOLLENT, Michel. Action research methodology . 18th ed. São Paulo: Cortez, 2018. Available at: https://www.amazon.com.br/Metodologia-pesquisa-a%C3%A7%C3%A3o-Michel-Thiollent/dp/8524917164 Accessed on: July 10, 2025.
URBINATI, Nadia . Me the People: How Populism Transforms Democracy . Cambridge: Harvard University Press, 2019. Available at: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-8675.12510 Accessed on: July 10, 2025.
VITORINO, Marcelo; MENDONÇA, Natália. Political campaigns on social media: how to communicate and win elections. Brasília, DF: Superior Electoral Court, 2020. Available at: https://www.amazon.com.br/Campanhas-pol%C3%ADticas-nas-redes-sociais/dp/858230613X Accessed on: July 10, 2025.
KENNEY, M. G. and Weinstein, A. (2010), " Psychographic segmentation of the self-employed : An exploratory study ", New England Journal of Entrepreneurship , Vol. 13 No. 2, pp. 47-56. Available at: https://digitalcommons.sacredheart.edu/neje/vol13/iss2/6?utm_source=digitalcommons.sacredheart.edu%2Fneje%2Fvol13%2Fiss2%2F6&utm_medium=PDF&utm_campaign=PDFCoverPages Accessed on: July 10, 2025.
WEBB, Michael; DOWLING, Melissa-Ellen; FARINA, Matteo. Understanding Mass Influence : a case study of Cambridge Analytica as a contemporary mass influence campaign . Adelaide: University of Adelaide, 2021. Technical report.
Available at: https://www.unsw.edu.au/content/dam/pdfs/unsw-canberra/dri/2023-02-research/2023-02-Understanding-Mass-Influence---A-case-study-of-Cambridge-Analytica.pdf Accessed on: July 10, 2025.
ZUBOFF, Shoshana . The Age of Surveillance Capitalism : The Fight for a Human Future at the New Frontier of Power. New York: PublicAffairs , 2019. Available at: https://www.amazon.com/Age-Surveillance-Capitalism-Future-Frontier/dp/1610395697 Accessed on: July 10, 2025.