Analyzing Patterns of Political Polarization in Contemporary Digital Society through Unsupervised Machine Learning and Data-Driven Clustering

Hendro Budiyanto[1,*], Sundring Pantja Djati[2]

[1]Faculty of Business Universitas Multimedia Nusantara, Scientia Boulevard, Tangereng, Banten, Indonesia

[2]Faculty of Business Universitas Esa Unggul, Jakarta, Indonesia

ABSTRACT

This study investigates the spatial and algorithmic dimensions of political polarization in the contemporary digital society by analyzing U.S. county-level election data through unsupervised machine learning. Using K-Means clustering combined with Principal Component Analysis (PCA), the research identifies four distinct voting clusters that represent structural variations in partisan alignment across the United States. The Elbow Method determined that k = 4 offers the optimal balance between model simplicity and interpretability, while the PCA results indicate that the first two principal components explain approximately 34.97% of the total variance in the dataset. The clustering outcomes reveal that counties exhibit strong spatial coherence in political behavior, forming ideologically homogeneous regional blocs that reflect persistent partisan divisions. The normalized voting heatmap further demonstrates that polarization is predominantly organized along the Republican (REP) and Democratic (DEM) axis, with minor parties contributing minimally to aggregate voting patterns. These findings suggest that digital communication technologies amplify existing regional ideological identities through algorithmic filtering and selective exposure, thereby reinforcing the two-party dominance rather than diversifying political participation. The study concludes that polarization in the digital age is not only a matter of ideological difference but also a function of spatial and algorithmic reinforcement, where online and offline environments jointly shape political homogeneity. The results underscore the need for inclusive digital infrastructures, improved media literacy, and algorithmic transparency to mitigate polarization and promote more deliberative democratic engagement in the digital public sphere.

Keywords Political Polarization, Digital Society, Machine Learning, Spatial Clustering, Algorithmic Reinforcement

Submitted 14 January 2026
Accepted 10 February 2026
Published 1 March 2026

Corresponding author
Hendro Budiyanto,
hendro.budiyanto@umn.ac.id

Additional Information and Declarations can be found on page 12

INTRODUCTION

Political polarization has emerged as one of the defining characteristics of democratic societies in the twenty-first century, reshaping patterns of political engagement, social interaction, and information exchange. In an era dominated by digital communication, the boundaries between online and offline political behavior have become increasingly blurred. Social media platforms, online news aggregators, and algorithmic recommendation systems have redefined how the rapid expansion of digital technologies has profoundly reshaped how individuals' access, interpret, and share political information. While these technologies were initially expected to expand democratic participation and diversify public discourse, they have also contributed to the consolidation of

How to cite this article: H. Budiyanto and S. P. Djati, "Analyzing Patterns of Political Polarization in Contemporary Digital Society through Unsupervised Machine Learning and Data-Driven Clustering," *J. Digit. Soc.*, vol. 2, no. 1, pp. 1-14, 2026.

ideological enclaves [1]. Algorithms that personalize content according to user preferences often expose individuals to information consistent with their prior beliefs, reinforcing confirmation bias and deepening partisan divisions [2]. As a result, political communication in the digital sphere has become both more immediate and more fragmented, producing a communication environment where ideological reinforcement outweighs deliberative engagement [3].

Although scholars have widely examined the behavioral and psychological aspects of digital polarization, empirical studies that integrate spatial and structural dimensions remain limited [4]. Much of the existing research focuses on user-level data from social media or opinion surveys, providing valuable insights into individual attitudes but often neglecting the broader systemic patterns that connect geography, demography, and digital influence. The spatial dimension of polarization is particularly important because political affiliation and ideological preferences often exhibit regional clustering. Communities that share socioeconomic conditions, cultural values, and media ecosystems are likely to develop similar voting behaviors. However, little is known about how these spatial tendencies interact with digital infrastructures that amplify or attenuate partisan information. Furthermore, the application of machine learning techniques in political science has largely centered on prediction and classification, while their potential to uncover latent structures and hidden groupings within electoral data has not been fully realized [5]. Exploring these structures through unsupervised learning methods provides an opportunity to understand polarization not as a set of individual choices but as an emergent property of interconnected social, spatial, and algorithmic systems.

In response to these gaps, this study applies unsupervised machine learning methods, specifically K-Means clustering combined with PCA, to examine county-level voting data in the United States. The aim is to identify underlying patterns of political polarization and to visualize how ideological alignment varies across regions. This methodological approach enables the detection of natural groupings within the data, revealing how local and regional political identities align with broader national trends. By integrating quantitative data analysis with theoretical insights from digital sociology and political communication, this research seeks to demonstrate that polarization in the digital era is not only ideological but also structurally and spatially embedded. The findings contribute to ongoing discussions about the impact of digital technologies on democratic processes, showing how algorithmic personalization and online information flows interact with long-standing geographic and social divisions. Ultimately, this study provides an empirical framework for understanding political polarization as a multidimensional process shaped by the convergence of digital systems, spatial environments, and collective behavior.

Literature Review and Related Works

Political polarization has long been recognized as a central concern in democratic societies, particularly due to its impact on social cohesion, civic participation, and institutional trust. Traditional studies conceptualize polarization as ideological distance between parties or voters, often measured through voting records and public opinion surveys [6], [7]. However, with the rise of digital media, the nature of polarization has shifted from a primarily political phenomenon to a digitally mediated one. Online environments now serve as spaces where individuals construct, reinforce, and contest ideological identities [8]. The emergence of social networking platforms has enabled users to selectively expose themselves to politically congruent information, strengthening confirmation bias and reducing ideological diversity [9], [10]. This process has contributed to the formation of echo chambers and filter bubbles that limit exposure to opposing viewpoints and increase cognitive isolation within digital communities [11], [12].

Algorithmic personalization has further deepened these divisions by shaping what users see, share, and engage with on digital platforms. Recommendation systems such as those used by Facebook, YouTube, and X (formerly Twitter) tend to prioritize engagement over informational diversity [13], [14]. Research shows that emotionally charged and partisan content achieves higher interaction rates, incentivizing algorithmic amplification of such material [15]. This dynamic not only reinforces existing beliefs but also contributes to affective polarization, where political opponents are perceived as social or moral adversaries rather than legitimate competitors [16], [17]. As digital communication becomes increasingly mediated by algorithms, polarization is no longer merely a reflection of ideological differences but an outcome of technologically driven information architectures [18], [19]. Consequently, political discourse in digital spaces is shaped by feedback loops between human behavior, algorithmic curation, and networked communication.

The growing integration of computational methods in social science has allowed researchers to study polarization from new perspectives. Machine learning techniques, for example, have been applied to classify political sentiment, detect misinformation, and predict electoral outcomes [20], [21]. While these studies have advanced predictive accuracy, many rely on supervised models that depend on labeled datasets and predefined categories, which may obscure latent ideological structures. Unsupervised methods, such as clustering algorithms, offer a more exploratory approach by revealing hidden relationships in high-dimensional political data [22]. Recent studies have employed K-Means and hierarchical clustering to identify ideological communities and voting blocs in both digital and geographic contexts [23], [24]. Dimensionality reduction techniques, including PCA, have also been used to visualize political divisions and map regional voting behavior [25]. However, most of these works focus either on social media interactions or aggregate electoral outcomes without explicitly connecting spatial and algorithmic dimensions of polarization.

This study builds upon these prior contributions by integrating unsupervised machine learning with spatial analysis to explore patterns of political polarization in the United States. Unlike research that focuses solely on online discourse or behavioral metrics, this study employs K-Means clustering combined with PCA to identify natural groupings in county-level voting data. The approach allows for a simultaneous examination of ideological alignment, spatial proximity, and structural differentiation, offering a data-driven representation of polarization as both a social and computational phenomenon. By framing polarization as a multidimensional process influenced by geography and digital systems, this research contributes to a more comprehensive understanding of how algorithmic reinforcement interacts with regional political identities in the digital society [26], [27], and [28].

Methods

This study employed a quantitative and computational approach using unsupervised machine learning to identify structural patterns of political polarization in the digital era. The analytical workflow consisted of five major stages: data preprocessing, feature scaling, dimensionality reduction, clustering, and visualization. Each stage was designed to transform raw electoral data into meaningful representations of ideological similarity and spatial coherence, as summarized in figure 1, which illustrates the sequential steps of the research pipeline.

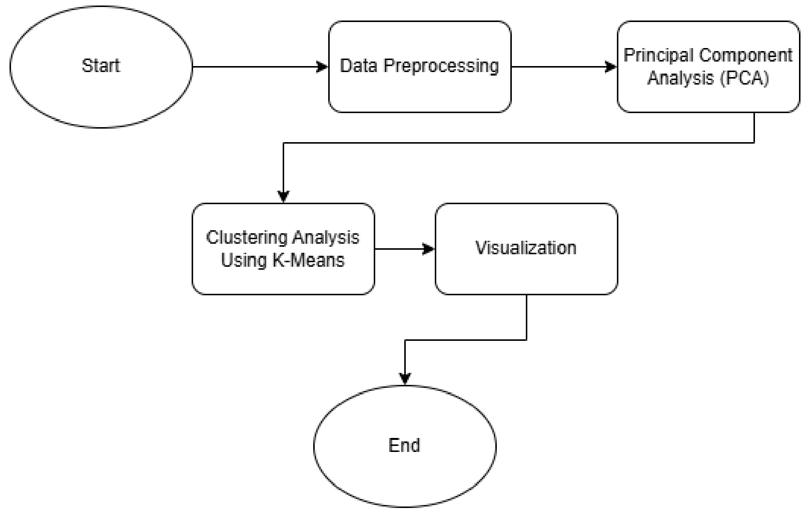

Figure 1 Research Steps

The dataset used in this research was derived from official United States county-level presidential election results stored in president_county_candidate.csv. Each observation represented a combination of state and county, with corresponding vote counts for multiple political parties. The dataset contained four main attributes: state, county, party, and total_votes. These variables provided a comprehensive basis for understanding regional political alignment. To structure the data for computational analysis, a pivot transformation was performed so that each row represented a county and each column represented a political party, with cell values corresponding to total votes. Missing values were replaced with zeros to ensure completeness, while non-numerical attributes were excluded. This transformation produced a multidimensional feature space where each county was characterized by its distribution of votes across different political parties.

Before clustering, the data were normalized using z-score standardization to eliminate disparities in vote magnitudes among counties. The normalization process was defined as:

$$z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j} \quad (1)$$

x_{ij} represents the vote count for party j in county i, μ_j is the mean vote count for party j across all counties, and σ_j is the corresponding standard deviation. This transformation ensured that each feature contributed equally to the clustering process by rescaling all variables to zero mean and unit variance.

To reduce the dimensionality of the dataset while preserving essential variance, PCA was applied. PCA transforms correlated variables into a set of orthogonal components that capture the maximum variance in the data. Mathematically, PCA decomposes the standardized data matrix Xinto a set of eigenvectors and eigenvalues according to the equation:

$$Z = XW \tag{2}$$

$Z$is the transformed data in the principal component space, and $W$is the matrix of eigenvectors corresponding to the covariance matrix $X^T X$. The first two principal components (PC_1 and PC_2) explained approximately 34.97% of the total variance, with PC_1contributing 20.81% and PC_2contributing 14.16%. The two-dimensional representation provided by PCA facilitated clearer visualization of inter-county relationships and enhanced the separability of clusters in the feature space.

The reduced data were subsequently clustered using the K-Means algorithm, which partitions the dataset into $k$clusters by minimizing the Within-Cluster Sum of Squares (WCSS) objective function:

$$J = \sum_{k=1}^{K} \sum_{i \in C_k} \| x_i - \mu_k \|^2 \tag{3}$$

$J$is the total within-cluster variance, $x_i$represents an individual data point, $\mu_k$is the centroid of cluster C_k, and $K$is the total number of clusters. The optimal number of clusters was determined using the Elbow Method, which evaluates how WCSS decreases as Kincreases. The inflection points of the curve, observed at $K = 4$, indicated that four clusters provided an optimal trade-off between model simplicity and interpretability. The final K-Means configuration used $K = 4$, a random seed of 99, $n_{\text{init}} = 25$, and a maximum of 500 iterations to ensure stability and reproducibility.

Following the clustering process, visualization techniques were employed to interpret and validate the results. The scatter plot illustrated the spatial distribution of counties in the two-dimensional PCA space, where color-coded groups represented distinct clusters. This visualization highlighted the natural grouping of politically similar counties and revealed the presence of transitional and homogeneous regions. To further understand inter-party relationships, a heatmap was generated to depict the normalized average vote share of major political parties across the four clusters. Normalization was performed by dividing the mean vote of each party in a given cluster by its maximum value across clusters, producing a normalized scale between 0 and 1. The heatmap, visualized using the RdBu_r color palette, effectively illustrated partisan dominance and variation, emphasizing that polarization primarily follows the Republican–Democratic axis with limited third-party influence.

All visualizations, including the Elbow plot, were created using Matplotlib and Seaborn libraries in Python, ensuring high-quality graphical output suitable for publication. Each figure was saved in high resolution (400 dpi) to maintain clarity in printed form. This methodological design provides both analytical rigor and visual transparency, enabling reproducible and interpretable insights into political polarization patterns.

In summary, the methodological framework integrates statistical normalization, dimensionality reduction, and unsupervised learning to identify emergent patterns of political alignment. The complete analytical pipeline—comprising preprocessing, PCA transformation, K-Means clustering, and visualization is depicted schematically in figure 2, which presents the overall research steps from raw data acquisition to the generation of interpretable political clusters and visual heatmaps. This systematic workflow ensures computational efficiency and interpretive clarity, offering a scalable model for future studies examining political behavior in the context of digital and spatial analysis.

Algorithm 1: Political Polarization Clustering

Input:

County-level election dataset $D = \{(s_i, c_i, p_i, v_i)\}_{i=1}^{N}$

Output:

Clustered political regions and visualized heatmap of normalized voting patterns

Process:

Start

Load the United States county-level election dataset from president_county_candidate.csv.

Transform the dataset into a matrix form $X = [x_{ij}]_{n \times m}$, where each row i represents a county and each column j represents a political party. Missing values are replaced with zero.

Standardize each variable using z-score normalization to ensure comparability across features.

$$z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j}$$

Apply Principal Component Analysis (PCA) to reduce dimensionality while preserving the highest possible variance.

$$Y = ZW \text{ where } W \text{ is the eigenvector matrix of } Z^T Z$$

Retain two principal components PC_1 and PC_2, which together explain approximately 34.97% of the total variance.

Initialize K-Means parameters with $K = 4$, random seed = 99, $n_{init} = 25$, and maximum iterations = 500.

Assign each observation y_i to the nearest cluster centroid using the Euclidean distance criterion.

$$C_i = \arg\min_k \| y_i - \mu_k \|^2$$

Update the centroid of each cluster until convergence.

$$\mu_k = \frac{1}{|C_k|} \sum_{y_i \in C_k} y_i$$

Minimize the Within-Cluster Sum of Squares (WCSS) objective function.

$$J = \sum_{k=1}^{K} \sum_{y_i \in C_k} \| y_i - \mu_k \|^2$$

Determine the optimal number of clusters using the Elbow Method by identifying the inflection point in $J(K)$. The optimal cluster number is $K^* = 4$.

Compute the average vote per cluster for each political party.

$$M_{kj} = \frac{1}{|C_k|} \sum_{i \in C_k} x_{ij}$$

Normalize the cluster means across parties to a scale between 0 and 1.

$$N_{kj} = \frac{M_{kj}}{\max_k M_{kj}}$$

Generate visualizations including: (a) a PCA scatter plot showing cluster distribution, (b) a heatmap of normalized voting averages, and (c) an Elbow plot of $J(K)$.

Export visual results and cluster summary for interpretation.

End

Result

To identify the most appropriate number of clusters for representing variations in county-level voting behavior, the Elbow Method was employed using the within-cluster sum of squares (WCSS) as the evaluation criterion. The WCSS metric quantifies the degree of compactness within each cluster by calculating the sum of squared distances between data points and their corresponding cluster centroids. As the number of clusters (k) increases, the WCSS value typically decreases because each cluster captures a smaller and more homogeneous subset of the data. However, beyond a certain point, additional clusters lead to only marginal improvements in compactness, suggesting diminishing returns. This inflection point, often referred to as the "elbow," provides an empirical basis for selecting the optimal number of clusters that balances model simplicity and explanatory power.

The Elbow plot generated from this analysis revealed a clear and stable inflection at k = 4, where the reduction in WCSS begins to plateau. This observation indicates that four clusters provide an optimal compromise between minimizing intra-cluster variance and maintaining meaningful group differentiation. In practical terms, this means that the voting behavior of U.S. counties can be effectively represented by four major patterns or groupings. Each cluster likely reflects a distinct combination of partisan preferences, demographic structures, and regional voting tendencies. The adoption of k = 4 ensures that the clustering model captures the underlying heterogeneity in political alignment while avoiding overfitting that might obscure broader trends across the national electoral landscape.

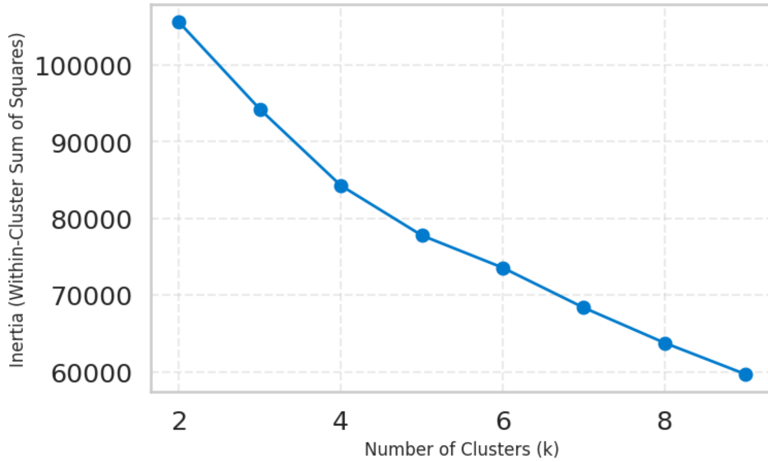

Figure 2 Elbow Method to Determine Optimal Number of Clusters

After determining the optimal number of clusters, PCA was applied to the standardized dataset to reduce its dimensionality and facilitate visualization of inter-county political similarities. PCA transforms a high-dimensional dataset into a smaller set of uncorrelated components that retain most of the original variance. This approach allows complex patterns of voting behavior across multiple political parties to be summarized in a more interpretable two-dimensional space. In this analysis, the first two principal components jointly explained approximately 34.97% of the total variance, with Principal Component 1 (PC1) accounting for 20.81% and Principal Component 2 (PC2) contributing 14.16%. The relatively high explanatory power of these two components suggests that they effectively capture the dominant trends in political preference across U.S. counties, including major ideological divides and variations in party strength.

The two-dimensional scatter plot produced from the PCA results, as illustrated in figure 3, reveals a clear and balanced separation among the four clusters. This visualization indicates that the K-Means model effectively distinguishes between counties based on their voting characteristics. Clusters 0 and 1 are concentrated near the center of the plot, reflecting regions with mixed or moderate political tendencies where both major parties maintain significant support. In contrast, Clusters 2 and 3 appear as smaller, more distinct groupings positioned further from the plot's origin, signifying counties with more extreme or homogeneous ideological alignments. The spatial arrangement of these clusters suggests that political polarization in the United States exhibits a geographically structured pattern, where counties with similar partisan preferences tend to cluster together, forming identifiable regional blocs of ideological homogeneity.

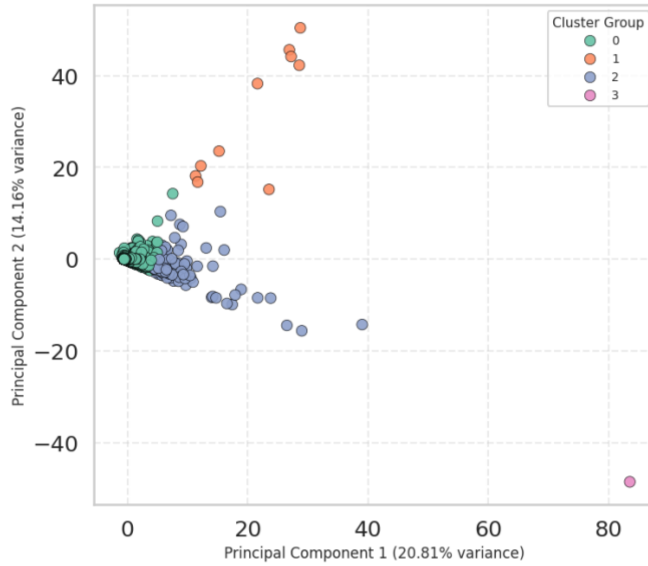

Figure 3 Balanced and Stable Visualization of Political Polarization Clusters

To further interpret the meaning of each cluster and identify distinctive voting characteristics, the average normalized vote distribution across political parties was calculated and represented through a heatmap, as shown in figure 4. This visualization provides an intuitive overview of how party preferences vary across the four identified clusters. The analysis indicates that the two dominant political parties, the REP and DEM parties, overwhelmingly shape the overall voting structure at the county level. Both parties consistently exhibit normalized averages close to or equal to 1.00 across all clusters, confirming their central role in defining electoral behavior. However, the relative balance between these two parties varies meaningfully across clusters, revealing nuanced regional patterns. Cluster 0 demonstrates the strongest Republican alignment with values of REP = 1.00 and DEM = 0.83, corresponding to counties that historically lean conservative. Cluster 1 presents a more competitive configuration with REP = 0.69 and DEM = 1.00, suggesting swing regions where electoral outcomes may fluctuate between the two major parties depending on contextual factors such as demographics or campaign intensity.

In contrast, Cluster 2 maintains a moderate degree of polarization with REP = 0.60 and DEM = 1.00, while Cluster 3 reveals the weakest Republican presence at REP = 0.38 and the strongest Democratic dominance with DEM = 1.00. These latter clusters are likely to represent urbanized or socially progressive regions where Democratic support is highly consolidated. The analysis of minor parties, including Libertarian (LIB), Green (GRN), and Constitution (CST), shows normalized averages consistently below 0.05, underscoring their limited influence on aggregate voting behavior. The overall heatmap pattern demonstrates that political polarization in the digital age remains fundamentally organized around the two-party system. The absence of distinct third-party clustering suggests that ideological competition continues to be structured primarily along the traditional Republican–Democrat axis, rather than diversifying into new or alternative political alignments. This outcome reflects both the institutional strength of the two-party system and the reinforcing role of digital media environments that amplify established partisan divisions.

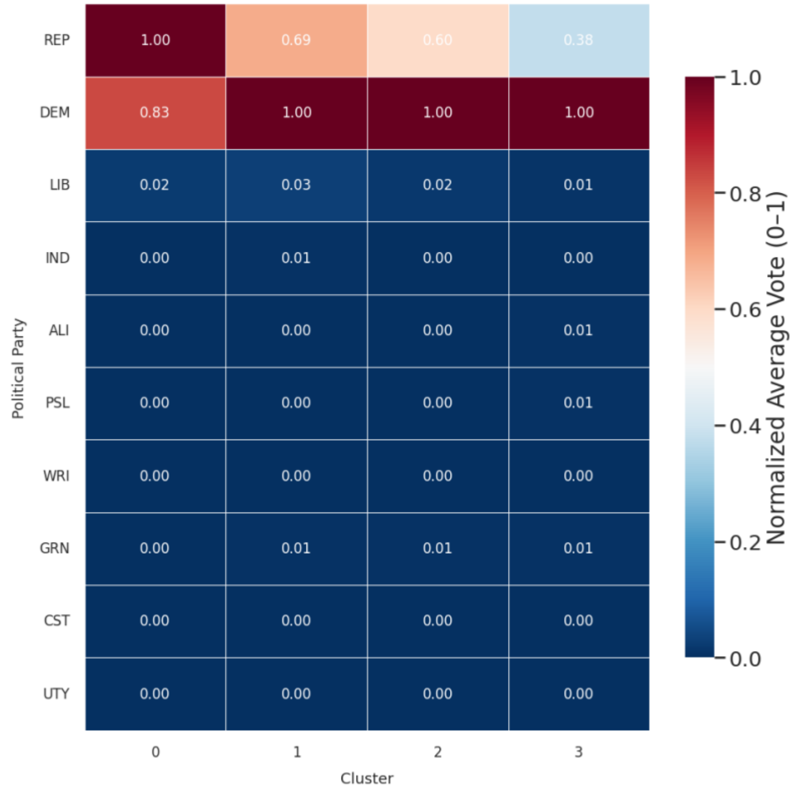

Figure 4 Normalized Voting Distribution Across Political Clusters

Overall, the clustering results reveal a spatially uneven but structurally coherent pattern of political polarization. Counties with similar ideological orientations tend to cluster geographically, suggesting that local sociopolitical environments and digital communication ecosystems may reinforce ideological alignment. The existence of transitional clusters, such as Cluster 1, highlights regions of ideological fluidity that may serve as “bridging zones” between polarized blocs. These findings align with theories of digital echo chambers and algorithmic reinforcement, emphasizing that political identity in the digital age is both regionally localized and digitally amplified through social media and networked communication.

Discussion

The results of the clustering analysis reveal that political polarization in the digital age is not merely an abstract ideological phenomenon but a spatially embedded social reality. The emergence of four distinct clusters suggests that political identity at the county level in the United States is increasingly structured by both geographic proximity and digital interaction. Counties that share similar political leanings tend to form cohesive regional blocs, reflecting the persistence of spatially anchored partisanship. This spatial clustering aligns with theories of “geographic sorting,” which posit that individuals are more likely to reside in communities where the prevailing political attitudes reflect their own values and worldviews. However, the digital layer of interaction appears to amplify this effect, as individuals in similar ideological spaces are also more likely to engage with the same online media sources, join politically homogenous networks, and consume content that reinforces their pre-existing beliefs. Consequently, digital communication technologies act as a multiplier of offline polarization, transforming localized political tendencies into entrenched ideological ecosystems that extend across both physical and virtual environments.

Moreover, the distribution of clusters and the corresponding heatmap patterns illustrate the deep entrenchment of the two-party system within this digitalized political landscape. The analysis shows that the Republican and Democratic parties dominate nearly all clusters, while minor parties remain politically marginal, underscoring the resilience of traditional political structures despite the transformative power of digital platforms. Rather than decentralizing influence, digital communication appears to reinforce the authority of established parties by amplifying partisan narratives through algorithmic curation and engagement-based visibility. This dynamic contributes to the reproduction of echo chambers and the reduction of exposure to alternative viewpoints, further solidifying ideological boundaries. Transitional clusters, particularly those with mixed partisan compositions, represent important sites for understanding potential pathways of depolarization or discourse diffusion, as they embody the few remaining intersections between competing digital publics. However, the growing precision of data-driven political advertising and the increasing personalization of digital media consumption risk eroding these shared spaces. In this sense, the findings highlight the paradox of digital society: while it enhances access to information and participation, it simultaneously entrenches divisions by fostering self-reinforcing political microcosms that challenge deliberative democratic engagement.

Conclusion

The findings of this study demonstrate that political polarization in the digital age is both a spatial and algorithmic phenomenon, deeply embedded within the structure of social interaction and information exchange. Through the application of unsupervised machine learning methods, particularly K-Means clustering combined with PCA, this research identified four distinct county-level voting clusters that collectively capture the underlying geography of political identity in the United States. These clusters reveal that electoral behavior is not randomly distributed but instead reflects coherent regional alignments shaped by ideological consistency and digital communication dynamics. The predominance of the REP and DEM parties across all clusters underscores the enduring influence of the two-party system, while the marginalization of minor parties such as the Libertarian, Green, and Constitution parties indicates the limited impact of alternative movements within the broader political ecosystem. The spatial coherence observed in the clustering results suggests that digital platforms, rather than bridging ideological divides, have amplified them through algorithmic personalization, targeted political messaging, and the creation of ideologically insulated online communities. This pattern reinforces the notion that digital society simultaneously democratizes access to information and entrenches partisan segmentation, thereby intensifying echo chambers and reducing opportunities for cross-ideological discourse. From a broader perspective, these insights emphasize the need for multidisciplinary strategies to address digital polarization, including the redesign of recommendation algorithms, the promotion of media literacy, and the development of inclusive digital infrastructures that encourage exposure to diverse perspectives. Future research should extend this analytical framework by integrating temporal dynamics, social media activity, and demographic variables to better understand how digital ecosystems evolve in shaping political identities and sustaining ideological divisions over time.

Declarations

Author Contributions

Conceptualization: H.B. and S.P.D.; Methodology: S.P.D.; Software: H.B.; Validation: H.B. and S.P.D.; Formal Analysis: H.B. and S.P.D.; Investigation: H.B.; Resources: S.P.D.; Data Curation: S.P.D.; Writing Original Draft Preparation: H.B. and S.P.D.; Writing Review and Editing: S.P.D. and H.B.; Visualization: H.B.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] C. R. Sunstein, *#Republic: Divided Democracy in the Age of Social Media*. Princeton, NJ: Princeton University Press, 2017. doi: 10.1086/696988.

[2] E. Pariser, *The Filter Bubble: What the Internet Is Hiding from You*. New York: Penguin Press, 2011.

[3] W. L. Bennett and S. Livingston, “The disinformation order: Disruptive communication and the decline of democratic institutions,” European Journal of Communication, vol. 33, no. 2, pp. 122–139, 2018, doi: 10.1177/0267323118760317.

[4] M. Gentzkow, J. M. Shapiro, and M. Taddy, “Measuring group differences in high-dimensional choices: Method and application to congressional speech,” *Econometrica*, vol. 87, no. 4, pp. 1307–1340, 2019, doi: 10.3982/ECTA16566.

[5] A. T. J. Barron, J. Huang, R. L. Spang, and S. DeDeo, “Individuals, institutions, and innovation in the debates of the French Revolution,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 115, no. 18, pp. 4607–4612, 2018, doi: 10.1073/pnas.1717729115.

[6] N. McCarty, K. T. Poole, and H. Rosenthal, *Polarized America: The Dance of Ideology and Unequal Riches*, 2nd ed. Cambridge, MA: MIT Press, 2016.

[7] A. Abramowitz and S. Webster, “The rise of negative partisanship and the nationalization of U.S. elections in the 21st century,” *Electoral Studies*, vol. 41, no. March, pp. 12–22, 2016, doi: 10.1016/j.electstud.2015.11.001.

[8] Z. Papacharissi, *Affective Publics: Sentiment, Technology, and Politics*. Oxford, UK: Oxford University Press, 2015.

[9] W. H. Dutton, G. Blank, and K. Dubois, “Is there a filter bubble or echo chamber? A cross-national panel study of internet users,” *Journal of Information Technology & Politics*, vol. 18, no. 1, pp. 1–24, 2021, doi: 10.1080/19331681.2020.1799588.

[10] S. Flaxman, S. Goel, and J. M. Rao, “Filter bubbles, echo chambers, and online news consumption,” *Public Opinion Quarterly*, vol. 80, no. S1, pp. 298–320, 2016, doi: 10.1093/poq/nfw006.

[11] D. M. J. Lazer et al., “The science of fake news,” *Science*, vol. 359, no. 6380, pp. 1094–1096, 2018, doi: 10.1126/science.aao2998.

[12] E. Bakshy, S. Messing, and L. A. Adamic, “Exposure to ideologically diverse news and opinion on Facebook,” *Science*, vol. 348, no. 6239, pp. 1130–1132, 2015, doi: 10.1126/science.aaa1160.

[13] R. K. Garrett, “Selective exposure: New directions and perspectives,” *Current Opinion in Psychology*, vol. 44, pp. 112–117, 2022, doi: 10.1016/j.copsyc.2021.08.004.

[14] F. Ribeiro, K. Zubiaga, M. Liakata, and S. C. H. Hoi, “The role of Twitter in the political polarization of the United States,” *Proceedings of the International AAAI Conference on Web and Social Media (ICWSM)*, vol. 12, no. 1, pp. 729–738, 2018.

[15] M. Brady, J. Wills, J. Jost, J. Tucker, and J. Van Bavel, “Emotion shapes the diffusion of moralized content in social networks,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 114, no. 28, pp. 7313–7318, 2017, doi: 10.1073/pnas.1618923114.

[16] S. Iyengar, G. Sood, and Y. Lelkes, “Affect, not ideology: A social identity perspective on polarization,” *Public Opinion Quarterly*, vol. 76, no. 3, pp. 405–431, 2012, doi: 10.1093/poq/nfs038.

[17] L. Mason, *Uncivil Agreement: How Politics Became Our Identity*. Chicago, IL: University of Chicago Press, 2018.

[18] T. W. Goolsby, “Algorithmic governance and the politics of personalization,” *Policy & Internet*, vol. 12, no. 3, pp. 314–334, 2020, doi: 10.1002/poi3.234.

[19] Z. Tufekci, “Algorithmic harms beyond Facebook and Google: Emergent challenges of computational agency,” *Colorado Technology Law Journal*, vol. 13, no. 1, pp. 203–218, 2015, doi: 10.2139/ssrn.2655716.

[20] D. Grimmer and B. M. Stewart, “Text as data: The promise and pitfalls of automatic content analysis methods for political texts,” *Political Analysis*, vol. 21, no. 3, pp. 267–297, 2013, doi: 10.1093/pan/mps028.

[21] R. Bond and D. Messing, “Social and emotional predictors of political polarization on Twitter,” *PLoS ONE*, vol. 16, no. 9, p. e0257433, 2021, doi: 10.1371/journal.pone.0257433.

[22] K. D. Galesic et al., “Unsupervised machine learning for analyzing political texts: Discovering latent ideology,” *Journal of Computational Social Science*, vol. 4, no. 3, pp. 741–763, 2021, doi: 10.1007/s42001-021-00115-4.

[23] R. W. Bond, D. J. Fink, and J. A. Wagner, “Clustering ideological communities in social media networks,” *Social Network Analysis and Mining*, vol. 9, no. 1, pp. 1–12, 2019, doi: 10.1007/s13278-019-0587-4.

[24] M. D. Conover, J. Ratkiewicz, M. Francisco, B. Gonçalves, F. Menczer, and A. Flammini, “Political polarization on Twitter,” in *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media (ICWSM)*, Menlo Park, CA: AAAI Press, pp. 89–96, 2011, doi: 10.1609/icwsm.v5i1.14126.

[25] M. E. J. Newman, “Modularity and community structure in networks,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 23, pp. 8577–8582, 2006, doi: 10.1073/pnas.0601602103.

[26] P. Barberá, “How social media reduces mass political polarization: Evidence from Germany, Spain, and the US,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 117, no. 37, pp. 22784–22791, 2020, doi: 10.1073/pnas.2018836117.

[27] D. Freelon, “Discourse architecture, ideology, and democratic norms in social media debates,” *New Media & Society*, vol. 17, no. 5, pp. 772–791, 2015, doi: 10.1177/1461444813513259.

[28] J. Tucker et al., “Social media, political polarization, and political disinformation: A review of the scientific literature,” *Political Science Quarterly*, vol. 133, no. 4, pp. 707–739, 2018, doi: 10.1002/polq.12881.