

CLUSTERIZACION ELECTORAL DE LOS MUNICIPIOS DE ANTIOQUIA MEDIANTE ALGORITMO DE CLASIFICACION NO SUPERVISADA KMEANS

José S. Díaz-Arguello¹

¹Economista y Magister en Ciencias Económicas de la universidad Santo Tomás, Especialista en Analítica de datos del Politécnico Jaime Isaza Cadavid. Correo electrónico: sebas_isa@hotmail.com, <https://orcid.org/0009-0007-9995-9674>

RESUMEN

Este estudio tiene como objetivo aplicar el algoritmo de clasificación no supervisada K-Means para identificar patrones de votación de los municipios del departamento de Antioquia según su comportamiento electoral en las elecciones al Senado de 2022. Para esto se seleccionaron los partidos políticos con una participación superior al 5 % del total de votos en el departamento. Los resultados permitieron identificar cuatro clústeres con patrones electorales claros: una tendencia liberal predominante, un grupo equilibrado entre partidos, una agrupación de orientación hacia el Centro Democrático y un grupo de predominio conservador. La metodología demostró ser eficaz para detectar patrones electorales territoriales, evidenciando la utilidad del aprendizaje automático en el análisis electoral y su potencial aplicación en estudios políticos.

Palabras clave: Electoral, Democracia, Política, Agrupamiento, Estadística.

Recibido: 20 de octubre de 2025. Aceptado: 11 de noviembre de 2025

Received: October 20, 2025. Accepted: November 11, 2025

ELECTORAL CLUSTERING OF THE MUNICIPALITIES OF ANTIOQUIA THROUGH THE UNSUPERVISED CLASSIFICATION ALGORITHM K-MEANS

ABSTRACT

This study aims to classify the municipalities of the department of Antioquia according to their electoral behavior in the 2022 Senate elections, through the application of the unsupervised classification algorithm K-Means. Political parties with a participation greater than 5% of the total votes in the department were selected. The results allowed the identification of four clusters with clear electoral patterns: one with a predominant liberal tendency, another showing a balanced distribution among parties, a group oriented toward the Centro Democrático party, and a group with a conservative predominance. The methodology proved effective in detecting territorial electoral patterns, demonstrating the usefulness of machine learning techniques in electoral analysis and their potential application in comparative political studies.

Keywords: Electoral, Democracy, Politics, Clustering, Statistics

Cómo citar este artículo: J. Díaz. "Clusterización electoral de los municipios de Antioquia mediante algoritmo de clasificación no supervisada K-Means." Revista Politécnica, vol.21, no.42 pp.88-98, 2025. DOI:10.33571/rpolitec.v21n42a6

1. INTRODUCCIÓN

Los algoritmos de clasificación pueden ser herramientas importantes para analizar y comprender los datos electorales. Su aplicación puede ayudar a políticos y a los partidos en la toma de decisiones informadas y estratégicas, y a mejorar su capacidad para conectarse con los votantes de manera más efectiva. Al examinar datos masivos se pueden identificar patrones y tendencias emergentes que ayuden a los actores políticos a comprender el panorama electoral y anticipar posibles escenarios. Esto proporciona una ventaja al momento de diseñar estrategias, asignar recursos y tomar decisiones tácticas en tiempo real.

El objetivo de esta investigación es clasificar a los municipios del departamento de Antioquia de acuerdo a su comportamiento electoral aplicando el algoritmo de inteligencia artificial no supervisada “K-means”, para que, con base en los resultados electorales de los diferentes partidos políticos en las elecciones a Senado de 2022, este algoritmo realice un proceso de agrupamiento de los municipios buscando similitudes en los resultados porcentuales de cada partido político.

La metodología de agrupamiento Kmeans, fue presentada por Stuart Lloyd en 1982 en el artículo “Least Squares Quantization in PCM” [1], en el que desarrolla un método para asignar niveles cuantificados a señales analógicas minimizando el error cuadrático medio, y que posteriormente se desarrollaría por otros autores para ser aplicado en otros ámbitos de investigación.

La literatura reciente sobre metodología aplicada al análisis electoral ha puesto énfasis en el desarrollo de herramientas que transformen los datos electorales en insumos reutilizables para la investigación y la toma de decisiones públicas. Dado que la inteligencia artificial y los algoritmos estadísticos son un tema álgido y creciente en el ámbito académico, existen varias investigaciones que aplican diferentes técnicas en temas electorales y políticos, la revisión de estos trabajos es importante para conocer e identificar las metodologías aplicadas y resultados encontrados.

El trabajo académico titulado “Analyzing Electoral Data Using Partitional and Hierarchical Clustering Algorithms” [2] es un estudio sobre el análisis de datos electorales utilizando algoritmos de agrupamiento jerárquico y particional, en este trabajo, los autores describen el uso de dos algoritmos de agrupamiento para investigar cómo se organizan los resultados de la segunda ronda de las elecciones presidenciales en Brasil de 2018, teniendo en cuenta los valores del Índice de Desarrollo Humano Municipal (IDHM).

La investigación tuvo como objetivo descubrir posibles relaciones entre indicadores que caracterizan perfiles de votantes brasileños y perfiles de municipios brasileños. El trabajo se centra en datos municipales descritos por subíndices IDHM, así como en la población municipal, relacionados con los 5.558 municipios del país. Los resultados del estudio muestran que los algoritmos de agrupamiento jerárquico (algoritmo DIANA) y particional (algoritmo k-Means) son herramientas eficaces para analizar datos electorales.

Con el fin de contribuir al principio de representatividad en las elecciones legislativas en la democracia de Taiwan, los autores del artículo Using Intelligent Clustering to Implement Geometric Computation for Electoral Districting [3], proponen una metodología para realizar de una manera más óptima el proceso de “distritación electoral”. En esta investigación aplicaron la metodología “Intelligent Clustering” compuesta también por el algoritmo K-Means, para identificar de una manera óptima la cantidad de distritos electorales del condado de Changhua. Como resultado, la investigación encontró una distribución mejor y más justa que la versión oficial, incluyendo errores de población inferiores y una forma más completa y simétrica de los distritos, evidenciando la aplicabilidad del algoritmo K-means para agrupar distritos electorales.

Por otro lado, Perdomo [4] plantea la pregunta ¿Se pueden predecir geográficamente los resultados electorales?, para responder a esto analiza las elecciones en Ciudad de México para el año 2000, dada la polarización de la ciudad y el conocimiento político del electorado. Como metodología utiliza una regresión múltiple, alimentada con el resultado electoral para jefe de gobierno de la ciudad como variable dependiente y los niveles de segregación espacial y el cambio de estos niveles como variables independientes. Como resultado de esta aplicación, Perdomo encuentra que la geografía electoral tiene una influencia significativa en varios aspectos, lo cual destaca su importancia en el ámbito electoral. Además de corroborar los resultados de

estudios previos, los cuales encontraron que en la Ciudad de México existe una división electoral y una segregación socioeconómica.

Otros estudios buscan crear nuevas metodologías de agrupamiento para las ciencias políticas, Nesterov, en el 2020 en su investigación “Soft clustering by convex electoral model” [5] sugiere una nueva técnica para el agrupamiento suave de datos multidimensionales. Se basa en un modelo de votación convexo, en el que los votantes elegirán su partido con una probabilidad que va a depender de la divergencia entre sus preferencias y la posición del partido, en este modelo los partidos podrán tomar decisiones según las encuestas para adaptar sus posiciones políticas a los resultados.

El estudio “Complex Clustering Using Constraint Programming: Modelling Electoral Map Creation” [6] propone un modelo de agrupamiento multinivel para la conformación de mapas electorales, formulado mediante programación de restricciones. Este enfoque aborda el problema como un caso de agrupamiento complejo, en el que intervienen múltiples condiciones espaciales, poblacionales y estructurales que exceden las capacidades de los métodos convencionales de aprendizaje automático. Los resultados evidencian que la programación de restricciones permite generar mapas electorales óptimos y consistentes con diversos criterios de equidad y coherencia territorial.

Por su parte, el trabajo “Redefinition of the Greek Electoral Districts through the Application of a Region-Building Algorithm” [7] propone un enfoque de agrupamiento espacial basado en restricciones, orientado a redefinir los distritos electorales de Grecia de manera más homogénea. El autor desarrolla el algoritmo SPiRAL (SPatial Integration and Redistricting ALgorithm), diseñado para integrar características geográficas y socioeconómicas en la delimitación de regiones. Este procedimiento realiza una construcción progresiva de grupos espaciales homogéneos, verificando en cada paso criterios de adyacencia, equilibrio poblacional y tamaño territorial. La aplicación del método en la Prefectura de Lakonia evidencia una mejora sustancial en la configuración del mapa electoral, tanto en su coherencia geométrica como en la representatividad poblacional, demostrando la utilidad del enfoque SPiRAL para abordar problemas reales de redistribución electoral.

Finalmente, el estudio “Constrained Clustering via Diagrams: A Unified Theory and Its Application to Electoral District Design” [8] propone una teoría unificada del agrupamiento restringido, fundamentada en el uso de diagramas geométricos para obtener particiones optimizadas bajo múltiples criterios. A través de distintos tipos de diagramas aplicados a espacios euclidianos y métricos discretos, los autores desarrollan una metodología que garantiza clústeres conectados y consolidados, capaces de preservar la estructura espacial original de los datos. En contextos electorales, esta técnica se traduce en la posibilidad de generar distritos coherentes tanto en su configuración territorial como en su composición demográfica, mostrando el potencial de la geometría computacional como herramienta para el diseño de mapas electorales justos y consistentes.

2. MATERIALES Y METODO

El proceso para la realización de esta investigación inició con el proceso de limpieza y organización de los datos, para posteriormente aplicar el modelo de clasificación K-Means y finalmente graficar los resultados. Para la ejecución de este proceso se utilizó el lenguaje de programación Python junto con las siguientes librerías: Pandas para realizar el proceso de limpieza, la transformación y la organización de los datos en un dataframe. La librería Sklearn para la aplicación del modelo Kmeans, con el que se realizará el agrupamiento de los municipios según los resultados electorales. Y matplotlib con el fin de generar los mapas del departamento y mostrar de manera gráfica la clasificación y características de votación de los municipios.

2.1 Metodología KMeans

El algoritmo de clasificación K-means es una técnica utilizada en el campo de la minería de datos y el aprendizaje automático para agrupar un conjunto de datos en diferentes categorías o clústeres basados en sus características y similitudes.

La metodología de ejecución de este algoritmo consiste en realizar la suma de las distancias al cuadrado entre los diferentes datos y el punto central más cercano. La cantidad de puntos centrales iniciales se seleccionan de manera aleatoria para definir cada dato a que centroide se acerca más, para después, realizar un

proceso iterativo optimizando el valor cuadrático de las distancias, es decir, cuando dejan de experimentar cambios significativos en su ubicación.

2.2 Fuente de los datos

Los datos recopilados para el análisis consisten en la votación al senado por municipio en el departamento de Antioquia para el periodo electoral 2022. Estos datos se obtuvieron de la Registraduría Nacional del Estado civil.

2.3 Tratamiento de los datos

Para obtener información relevante para el área de estudio (Antioquia), se aplicó un filtro en la variable departamento, dejando únicamente los datos correspondientes al área de estudio, Antioquia. Además, se filtró la circunscripción para obtener únicamente la circunscripción nacional. En cuanto al campo código_partido, se excluyeron los campos vacíos que corresponden a votos nulos, anulados y el voto en blanco.

La agrupación de los votos de cada partido por municipio permite identificar la participación porcentual del mismo en la votación, para lo que se calculó el porcentaje de votación de cada partido con respecto a los votos totales del municipio. Este proceso de normalización de la información busca compensar las diferencias en la cantidad de población de los municipios y de esta manera, se asegura que los datos estén representados de manera comparable evitando que el algoritmo identifique características similares únicamente en función del volumen de votos.

Con el objetivo de utilizar información relevante, se aplicó el modelo únicamente a los partidos políticos cuya votación total en el departamento representara al menos el 5% de la votación, como resultado de este proceso de selección, los siguientes partidos fueron los escogidos para ser empleados en la implementación del modelo: Partido Centro Democrático (CD), Partido Conservador Colombiano (PC), Coalicion Alianza Verde y Centro Esperanza (AVyCE), Partido Liberal Colombiano (PL) y el Pacto Historico (PH).

2.4 Definición y evaluación de K (Cantidad de clústeres)

La elección del número óptimo de clústeres es una parte fundamental en los modelos de clasificación no supervisada, ya que nos define los grupos en los que segmentaran los datos, en este caso, los municipios y su votación por partido. Para esta investigación se utilizó el método Elbow, que define gráficamente el valor "K" o cantidad de clústeres.

El método de Elbow, descrito por Kodinariya y Makwana [9], consiste en calcular la suma de los errores cuadráticos internos para distintos valores de K, observando el punto en el cual la reducción de la varianza intraclúster pierde su fuerza o inercia. Este punto es denominado "codo" y representa el número óptimo de clústeres. Una vez definida la cantidad de clústeres se utilizó el coeficiente de silueta y se revisó la Dispersión media Intra Clúster con el fin de evaluar el resultado de K.

El coeficiente de silueta, propuesto originalmente por Rousseeuw [10], mide la coherencia interna de cada clúster mediante la comparación entre la distancia media de un objeto con los elementos de su mismo grupo y la distancia con los objetos de otros grupos. Los valores próximos a 1 indican una buena separación entre clústeres, y valores cercanos a 0 o negativos reflejan asignaciones poco precisas.

Como complemento, se evaluó la dispersión media intraclúster, entendida como la distancia promedio entre las observaciones y el centroide de su grupo. Según Jain [11], este indicador refleja la homogeneidad interna de los conglomerados y permite identificar niveles de cohesión dentro del modelo.

3. RESULTADOS Y ANALISIS

3.1 Definición de los clústeres

La aplicación de la metodología Elbow dio como resultado un agrupamiento de los municipios de Antioquia en 3 clústeres como se observa en la Figura 1.

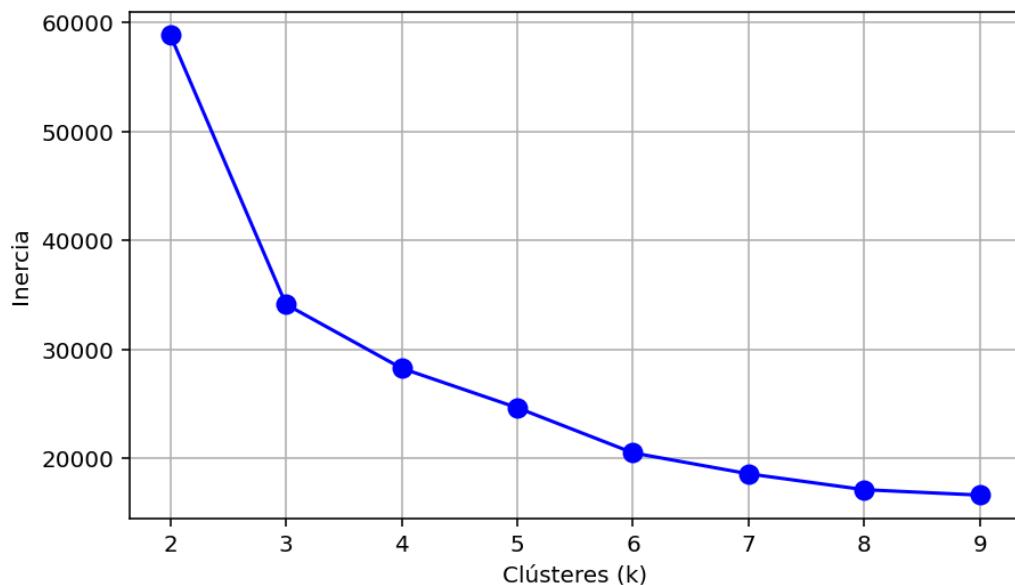


Figura 1. Gráfico metodología Elbow.

La evaluación del resultado de K con las metodologías Coeficiente de silueta y la Dispersión media intraclúster indican que el calculo $k=3$ es la cantidad de clústeres óptimos, sin embargo, la revisión empírica de los resultados de aplicar 3, 4 y 5 clústeres, indican la importancia de incluir un cuarto clúster en el modelo, como se referencia en la Tabla 1.

Tabla 1. Evaluación cantidad de clústeres.

k	Coeficiente Silueta	Dispersión media intraclúster	Justificación
3	0,36	14,9	Mejor separación global, captura los patrones principales.
4	0,30	13,7	Permite identificar un grupo adicional con tendencia clara al <u>partido que cuenta con la mayor votación en el departamento</u> .
5	0,29	12,1	Los clústeres adicionales generan fragmentación sin patrones claros (sobreajuste).

Aunque el coeficiente de silueta indica que el modelo con tres clústeres presenta la mejor separación global entre grupos, la exploración empírica del modelo con cuatro clústeres revela la aparición de un conjunto de municipios con una tendencia electoral particularmente concentrada hacia el Centro Democrático, este partido cuenta con la mayor votación a nivel departamental; también se observa que pasar de 3 a 4 clústeres evita que se fusionen dos patrones electorales distintos: el patrón equilibrado y el patrón orientado al Centro Democrático, generando un grupo adicional que aporta valor interpretativo al análisis, al distinguir territorios con comportamiento político homogéneo que no se diferenciaban con menor número de clústeres. En contraste, al incrementar el número de clústeres a cinco ($k=5$), los nuevos grupos generados no muestran patrones de votación definidos.

3.2 Clústeres generados

La información del resultado de los clústeres es presentada geográficamente en un mapa para conocer la dinámica general de la ideología predominante en cada municipio de Antioquia según las elecciones a congreso del 2022 tal como se muestra en la Figura 2, junto con la Tabla 2 con los municipios pertenecientes a cada clúster.

Aún más importante para presentar los resultados es conocer cuáles son las características de cada uno de estos clústeres, es decir, las votaciones predominantes para cada partido en cada municipio, y para mostrarlas se presentan las gráficas individuales de cada clúster con el mapa y el grafico de araña de cada uno.

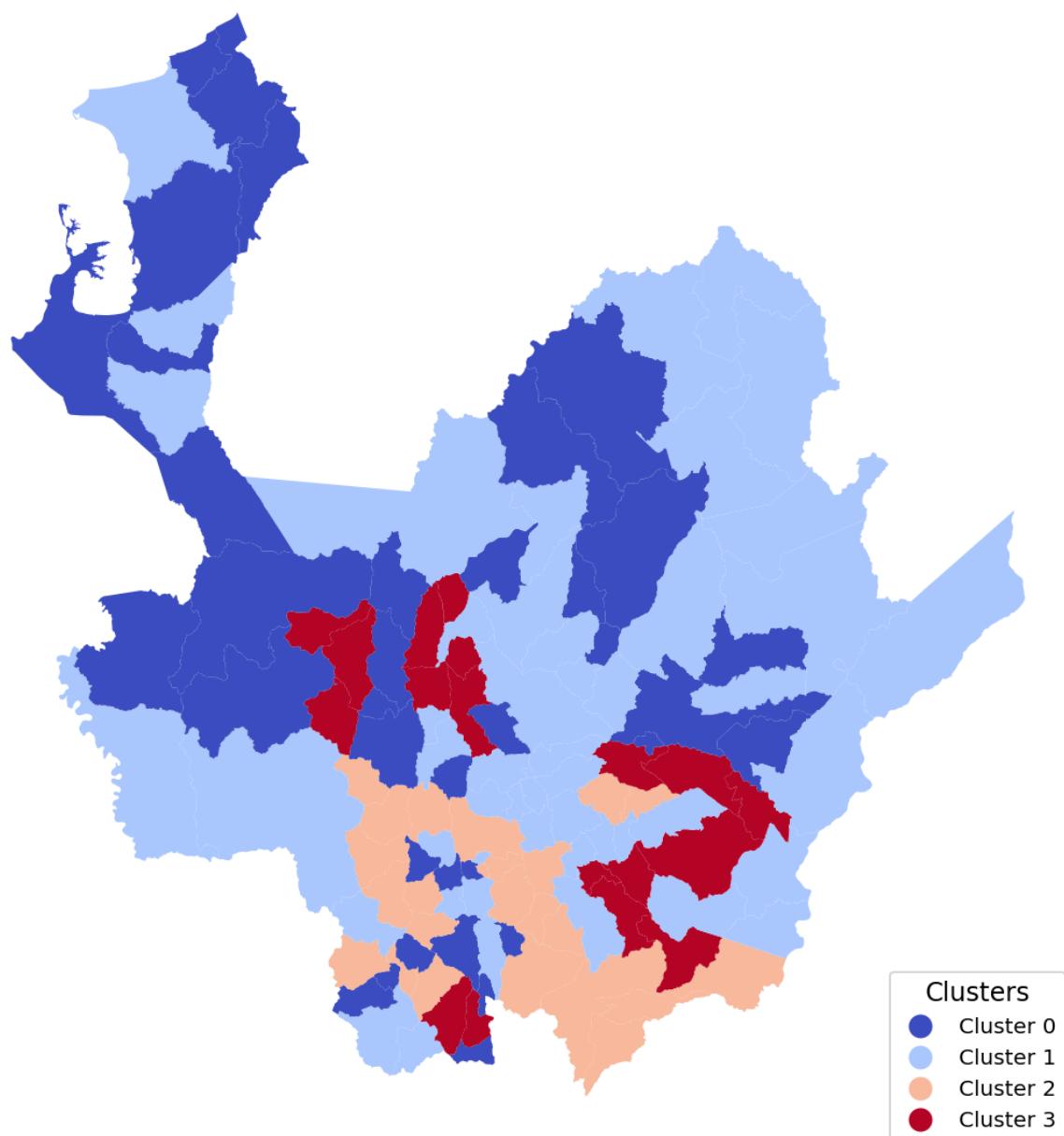


Figura 2. Mapa de municipios de Antioquia clasificados en clústeres.

Tabla 2. Cantidad y municipios por clúster.

Clúster	Cantidad	Municipios
0	35	Angelópolis, Anori, Arboletes, Armenia, Betania, Briceño, Buritica, Caceres, Campamento, Caramanta, Carepa, Cisneros, Dabeiba, Entrerríos, Fredonia, Frontino, Giraldo, Guadalupe, Hispania, La Estrella, La Pintada, Maceo, Montebello, Murindo, Mutata, Peque, San Jeronimo, San Juan De Uraba, San Pedro De Uraba, Santa Fe De Antioquia, Taraza, Tarsio, Turbo, Vegachi, Yolombo.
1	49	Amaga, Amalfi, Andes, Angostura, Apartado, Barbosa, Bello, Caldas, Carmen De Viboral, Carolina, Caucasia, Chigorodo, Copacabana, Donmatias, El Bagre, Girardota, Gomez Plata, Guarne, Guatape, Heliconia, Itagui, Ituango, Jardin, Marinilla, Nechi, Necocli, Olaya, Penol, Puerto Berrio, Puerto Nare, Puerto Triunfo, Remedios, Salgar, San Andres De Cuerquia, San Luis, San Pedro De Los Milagros, San Rafael, San Vicente, Santa Barbara, Santa Rosa De Osos, Segovia, Sopetran, Urrao, Valdivia, Vigia Del Fuerte, Yali, Yarumal, Zaragoza.
2	23	Abejorral, Alejandria, Anza, Argelia, Betulia, Caicedo, Ciudad Bolivar, Concepcion, Concordia, Ebejico, El Retiro, Envigado, Jerico, La Ceja, La Union, Medellin, Narino, Pueblorrico, Rionegro, Sabaneta, Sonson, Titiribi, Venecia.
3	18	Abriaqui, Belmira, Canasgordas, Caracoli, Cocorna, Granada, Liborina, Sabanalarga, San Carlos, San Francisco, San Jose De La Montana, San Roque, Santo Domingo, Santuario, Tamesis, Toledo, Uramita, Valparaiso.

3.3 Características por clúster

Una vez determinado el agrupamiento de los municipios según su patrón electoral, la información se representa mediante mapas temáticos y gráficos de telaraña. En el mapa se visualiza la distribución espacial de los municipios pertenecientes a cada grupo, mientras que el gráfico de telaraña ilustra el comportamiento porcentual de la votación por partido político. En este último, el área sombreada en color celeste corresponde al promedio de participación de cada partido dentro del clúster, y los puntos azules representan los valores individuales obtenidos en cada municipio para cada partido.

3.3.1 Clúster 0

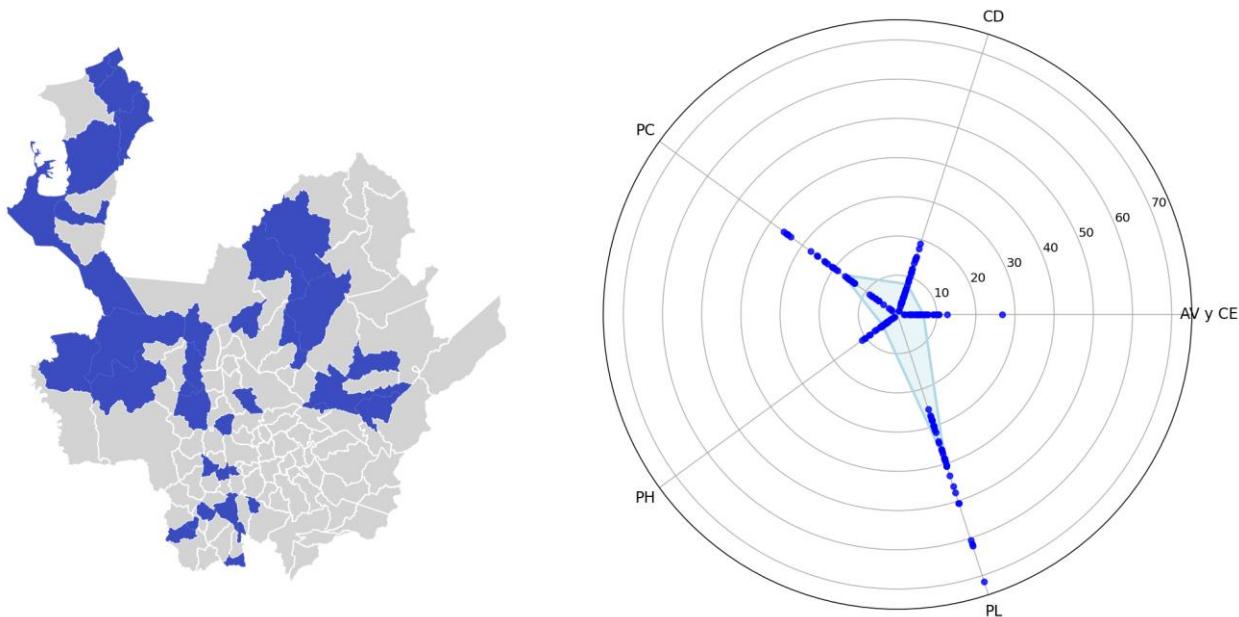


Figura 3. Mapa de municipios y porcentajes de votación del clúster 0.

El **Clúster 0** esta conformado por 35 municipio, con un patron electoral caracterizado principalmente por tener la votacion agrupada en el partido liberal, en el rango del 25% y el 72% de participacion electoral. Los otros partidos se agrupan por debajo del 30%, exceptuando algunos pocos municipios en los que el Partido conservador esta cerca del 35%. Estadisticamente se observa una media de 39,2 % para el Partido Liberal, acompañado de una desviación estandar de 11,3%, lo que confirma la predominancia liberal y, a la vez, la heterogeneidad interna de este grupo. **Este clúster puede describirse como de fuerte tendencia Liberal.**

3.3.2 Clúster 1

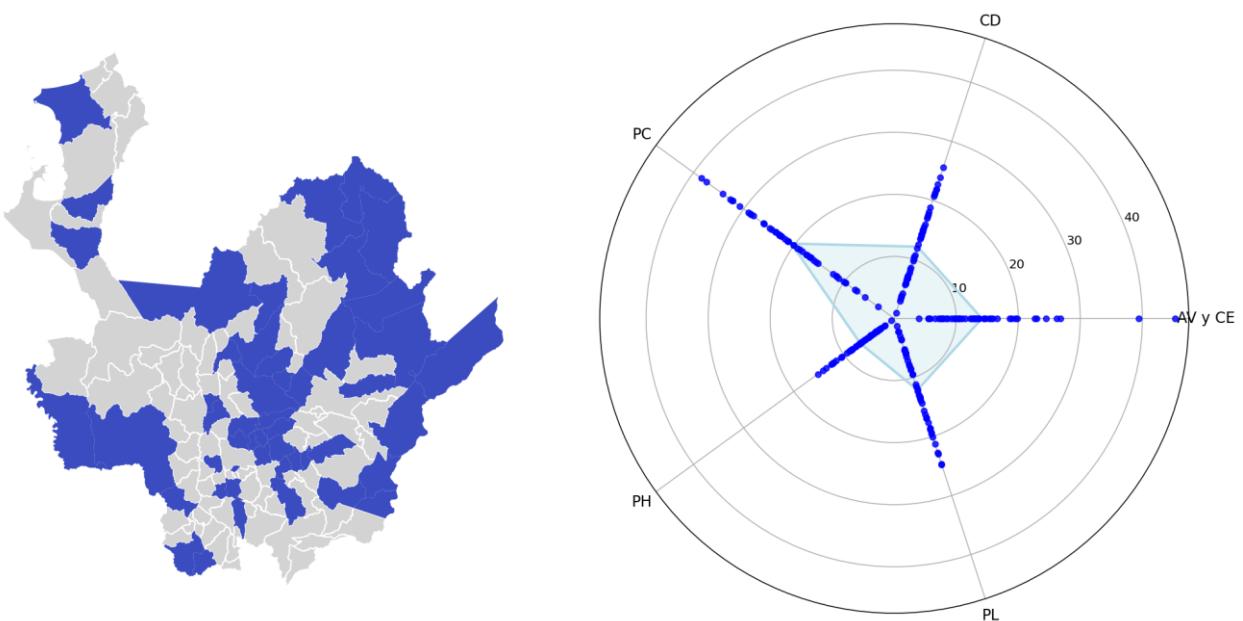


Figura 4. Mapa de municipios y porcentajes de votación del clúster 1.

En el **Clúster 1** encontramos 49 municipios dentro del cual encontramos un patron electoral mas equilibrado y heterogeneo entre partidos. Podemos observar que estos municipios se cocentran en el circulo del 30% para la mayoria de los partidos exceptuando el pacto historico, algunos municipios suben del 30% en votacion para el partido conservador y alianza verde y centro esperanza.

Las estadisticas presentan un patrón igualmente equilibrado, el Partido Conservador alcanza un promedio del 20,5 % y el Partido Liberal un 12,1 %, mientras que la Alianza Verde y Centro Esperanza (14,1 %) y el Centro Democrático 12,1%. Por lo anterior, podemos definir que el **patron electoral del clúster 1 es equilibrado y homogeneo.**

3.3.3 Clúster 2

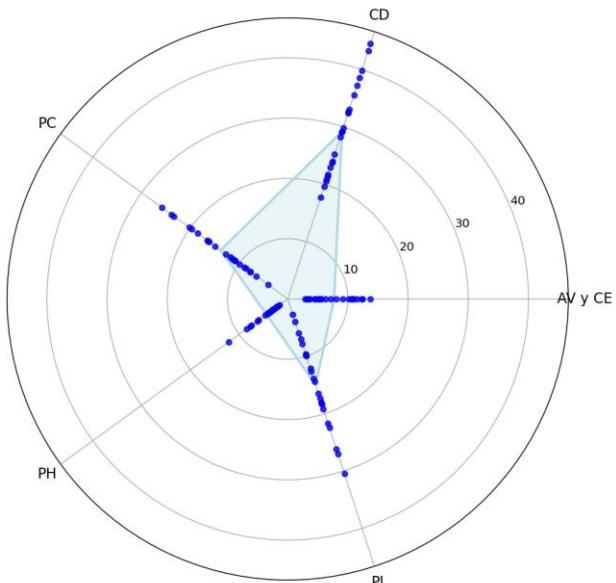
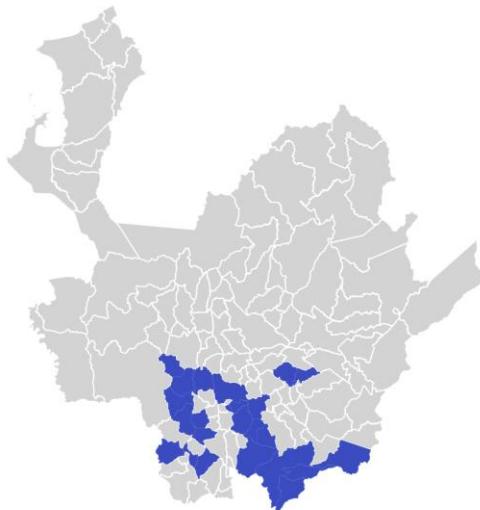


Figura 5. Mapa de municipios y porcentajes de votación del clúster 2.

El **Clúster 2** está conformado por 23 municipios con un patrón electoral bien definido por un porcentaje de participación en los votos de entre 20% y 45% en el Centro Democrático; seguido por los partidos tradicionales (conservador y liberal) cuentan con tendencia de votación a con máximos cercanos al 30%. Estadísticamente se evidencia una concentración notoria hacia el Centro Democrático, con un promedio de 29,2 % de votación y una desviación de 7,9 %, confirmando la cohesión ideológica en los municipios del sur del departamento. **Estos municipios del sur del departamento tienen una fuerte tendencia hacia el Centro Democrático.**

3.3.4 Clúster 3

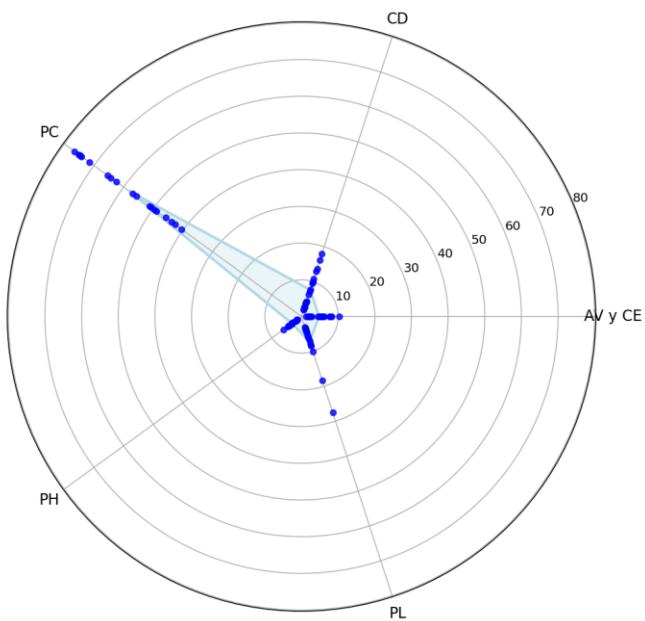
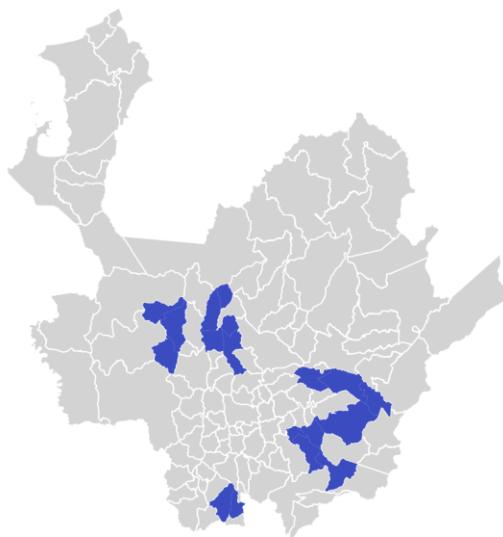


Figura 6. Mapa de municipios y porcentajes de votación del clúster 3.

El **Clúster 3** está conformado por 18 municipios, en los que en su totalidad la votación del partido conservador se ubicó por encima del 40% de la votación. Mientras que los otros partidos no suelen superar el 20% de la votación. Estadísticamente destaca por la marcada supremacía conservadora, con una media del 58,3

% y una desviación de 12,4 %, lo que demuestra uniformidad ideológica y bajo pluralismo electoral en esos territorios. **Podemos describir el patrón electoral en estos municipios como netamente conservador.**

4. CONCLUSIONES

De esta investigación es posible concluir que la aplicación del algoritmo K-Means permitió identificar agrupamientos de municipios con comportamientos electorales homogéneos que al observar la distribución en los gráficos de telaraña se evidencian las similitudes, demostrando que las técnicas de aprendizaje automático constituyen herramientas valiosas para el análisis político y territorial.

Los resultados también evidencian la existencia de cuatro patrones electorales diferenciados en Antioquia: un patrón electoral de fuerte tendencia liberal, un grupo de votación equilibrada, un grupo con concentración hacia el Centro Democrático y un bloque predominantemente conservador. En conjunto, estos patrones reflejan una configuración político-ideológica, con una inclinación mayoritaria hacia la derecha representada por los partidos Conservador y Centro Democrático y una orientación de centroizquierda asociada al Partido Liberal.

Mientras que el clúster 0 muestra una predominancia liberal con concentraciones de voto que superan el 60 %, el clúster 2 y 3 reflejan una polarización que tiende a consolidar territorios con alta cohesión ideológica hacia la derecha tradicional. En consecuencia, la segmentación obtenida mediante el algoritmo K-Means no solo diferencia patrones de voto, sino que también permite inferir el arraigo político y las posibles líneas de fractura dentro del mapa electoral departamental.

Por lo anterior, este estudio aporta una aproximación metodológica replicable para el análisis electoral en otras regiones, fortaleciendo la integración entre ciencia de datos y ciencias sociales.

5. REFERENCIAS BIBLIOGRÁFICAS

- [1] S. Lloyd, "Least Squares Quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, 1982. doi: 10.1109/TIT.1982.1056489.
- [2] P. Nietto, M. Nicoletti, and N. Sacco, "Analyzing Electoral Data Using Partitional and Hierarchical Clustering Algorithms," in *Intelligent Systems Design and Applications*, 2022. doi: 10.1007/978-3-031-27440-4_6.
- [3] Y.-C. Hung and L.-Y. Chen, "Using Intelligent Clustering to Implement Geometric Computation for Electoral Districting," *International Journal of Geo-Information*, vol. 8, no. 9, 2019. doi: 10.3390/ijgi8090369.
- [4] C. J. Perdomo, "¿Se pueden predecir geográficamente los resultados electorales?," *Estudios Demográficos y Urbanos*, vol. 23, no. 3, 2008. doi: 10.24201/edu.v23i3.1322.
- [5] Y. Nesterov, "Soft clustering by convex electoral model," *Soft Computing*, vol. 24, 2020. doi: 10.1007/s00500-020-05148-4.
- [6] L. Kotthoff, B. O'Sullivan, S. Ravi, and I. Davidson, "Complex Clustering Using Constraint Programming: Modelling Electoral Map Creation," unpublished manuscript, 2015.
- [7] Y. Photis, "Redefinition of the Greek Electoral Districts through the Application of a Region-Building Algorithm," *European Journal of Geography*, vol. 3, no. 2, 2012.

[8] A. Brieden, P. Gritzmann, and F. Klemm, “Constrained Clustering via Diagrams: A Unified Theory and Its Application to Electoral District Design,” *European Journal of Operational Research*, vol. 263, no. 1, 2017. doi: 10.1016/j.ejor.2017.04.018.

[9] T. Kodinariya and P. Makwana, “Review on Determining of Cluster in K-means Clustering,” *International Journal of Advance Research in Computer Science and Management Studies*, vol. 1, no. 6, 2013.

[10] P. Rousseeuw, “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis,” *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987. doi: 10.1016/0377-0427(87)90125-7.

[11] A. K. Jain, “Data clustering: 50 years beyond K-means,” *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651–666, 2010. doi: 10.1016/j.patrec.2009.09.011.