

MODELO DE MICROSEGMENTACIÓN ALGORÍTMICA PARA LA CAPTACIÓN DEL VOTO EN EL ECUADOR

ALGORITHMIC MICROSEGMENTATION MODEL FOR VOTE CAPTURE IN ECUADOR

José Ignacio Guaitoso Pilla ^{1*}

¹ Universidad Tecnológica Indoamérica, Facultad de Ciencias administrativas y económicas. Maestría en Administración de Empresas con mención en Innovación y Dirección Estratégica. Ecuador. ORCID: <https://orcid.org/0009-0002-5954-2042>. Correo: jguaitoso@indoamerica.edu.ec

Juan Carlos Suárez Pérez ²

² Universidad Tecnológica Indoamérica. Ecuador. ORCID: <https://orcid.org/0009-0000-5968-7669>. Correo: juansuarez@uti.edu.ec

* Autor para correspondencia: gpijose8@gmail.com

Resumen

La segmentación algorítmica del electorado consiste en procesar grandes volúmenes de datos con técnicas de aprendizaje automático para agrupar a los ciudadanos en colectivos más homogéneos, superando así el enfoque meramente demográfico. Objetivo: Construir un modelo de microsegmentación basado en k-means capaz de identificar perfiles diferenciados de votantes en Ecuador y facilitar la captación del sufragio. Método: Se adoptó un enfoque cuantitativo, transversal y de alcance descriptivo-explicativo. A partir del padrón electoral nacional para 2026 (N = 13.331.335), con un nivel de confianza del 95 % y un error del 5 %, se obtuvo una muestra de 385 encuestas. El instrumento presentó una adecuada validez de contenido (V de Aiken = 0,95) y una alta confiabilidad global (alfa de Cronbach = 0,881). Para el análisis se emplearon k-means, estandarización z-score, el método del codo y el coeficiente Silhouette. Resultados: El algoritmo permitió distinguir dos segmentos con diferencias significativas en las seis dimensiones psicográficas, así como en la intención de voto y la participación ($H = 86,851$; $p < 0,001$): el pasivo conservador (59,7 %) y el receptivo digital (40,3 %), con una asociación positiva y moderada con la captación del voto ($p = 0,561$). Conclusiones: El modelo ofrece un insumo práctico para diseñar mensajes, seleccionar canales y distribuir recursos de campaña, siempre dentro del marco ético y legal de protección de datos. De este modo, se rechaza la hipótesis nula y se confirma la existencia de perfiles electorales diferenciados y aprovechables para la captación del voto.

Palabras clave: Microsegmentacion algorítmica; marketing político; segmentación; captación; votación

Abstract

Algorithmic voter microsegmentation combines large-scale data processing with machine learning techniques to sort citizens into internally homogeneous collectives, moving beyond the conventional demographic approach. Objective: To build a microsegmentation model based on k-means that identifies differentiated voter profiles in Ecuador and supports vote capture. Method: A quantitative, cross-sectional and descriptive-explanatory design was adopted. The sample was set on the 2026 national electoral registry ($N = 13,331,335$), with 95 % confidence, 5 % error and 385 resulting surveys; the instrument reached content validity (Aiken's $V = 0.95$) and high overall reliability (Cronbach's $\alpha = 0.881$). The analysis relied on k-means, z-score standardization, the elbow method and the Silhouette coefficient. Results: The algorithm sorted voters into two segments differing significantly across the six psychographic dimensions and in vote intention and participation ($H = 86.851$; $p < 0.001$): passive-conservative (59.7 %) and receptive-digital (40.3 %), with a positive moderate-strength association with vote capture ($\rho = 0.561$). Conclusions: The model offers an actionable input for crafting messages, choosing channels and allocating campaign resources within the ethical and legal framework of data protection; the null hypothesis is rejected and differentiated, usable electoral profiles for vote capture are confirmed.

Keywords: *Algorithmic microsegmentation; political marketing; segmentation; campaigning; voting*

Fecha de recibido: 26/07/2026

Fecha de aceptado: 15/10/2026

Fecha de publicado: 19/10/2026

Introducción

En el campo del marketing político, la microsegmentación algorítmica se ha consolidado como una vía que suma el uso intensivo de datos a la inteligencia computacional para descomponer al electorado en bloques pequeños con rasgos en común. A raíz de la aceleración digital posterior a la pandemia, del torrente de información que los usuarios dejan en internet y de los avances del aprendizaje automático, hoy es factible trazar retratos del votante bastante más detallados que los que ofrecían los criterios demográficos clásicos (Ikotun et al., 2023; Waggoner, 2021; Ahmed et al., 2020). Por lo demás, la investigación experimental muestra que los mensajes confeccionados según el perfil psicográfico detectado por algoritmos logran una persuasión mayor que los mensajes genéricos, aunque el alcance de esa ventaja varía con el entorno social y con las normas de cada sistema político (Zarouali et al., 2022; Hackenburg & Margetts, 2024; Matz et al., 2017).

Antes de proseguir conviene separar tres conceptos que con frecuencia se mezclan. La predicción electoral procura adelantar resultados con modelos supervisados entrenados con series históricas; la persuasión política analiza cómo los mensajes diferenciados alteran las actitudes; y la microsegmentación constituye un procedimiento descriptivo de agrupación que expone la estructura subyacente del electorado sin prever ni inducir conductas. Este trabajo se ubica de manera consciente en el tercer campo: emprende una segmentación

descriptivo-explicativa que perfila grupos agregados y examina sus diferencias, sin trabajar con electores individuales. Dos disciplinas complementarias sostienen el argumento de este apartado: el marketing político y la ciencia de datos. Su encuentro ha disparado una discusión académica global en torno a los beneficios y los límites éticos de segmentar al electorado, discusión frente a la cual este estudio adopta una postura crítica: propone un uso únicamente analítico y agregado de los datos (Zuiderveen Borgesius et al., 2018; Seybold, 2025).

El origen de la microsegmentación algorítmica cabe situarlo en la convergencia del big data electoral con el aprendizaje automático orientado a esparcir mensajes políticos por medios digitales. Desde las campañas pioneras que operaron con datos masivos hasta la etapa actual, signada por la inteligencia artificial generativa, se han afinado diseños de segmentación que combinan la dimensión psicográfica y la demográfica del votante (Simchon et al., 2024). En ese recorrido, k-means destaca como la técnica de agrupamiento más usada en la ciencia de datos aplicada, tanto en el marketing de consumo como en los análisis de conducta política y en la clasificación de audiencias en línea (Ikotun et al., 2023; Waggoner, 2021; Sinaga & Yang, 2020).

En el plano de los efectos, Tappin et al. (2023) evidenciaron con estudios a gran escala que la microsegmentación rinde beneficios persuasivos medibles cuando el contenido se acomoda al perfil psicográfico del destinatario; el hallazgo amplía la senda abierta por Matz et al. (2017) acerca del direccionamiento psicológico de piezas hacia grandes audiencias. Aun así, Hackenburg y Margetts (2024) recuerdan que el impacto depende del tejido social e institucional de cada país y de las características de su régimen electoral, y Dobber et al. (2021) alertan sobre los límites que imponen los contenidos falsos del tipo deepfake. Tales precisiones invitan a interpretar aquella evidencia con prudencia en un Estado plurinacional como el ecuatoriano. Ante estos riesgos, la literatura aconseja una salida pragmática: administrar y atenuar el micro direccionamiento político sin llegar a su prohibición absoluta (Aagaard & Marthedal, 2023).

El linaje de estas estrategias se ubica por lo común en la campaña presidencial de Estados Unidos de 2012, reputada como un punto de inflexión por su uso intenso de datos y su segmentación dirigida a comunicarse con votantes de diversos territorios; no obstante, la prueba de ello procede en su mayoría de la prensa y de crónicas internas de campaña más que de valoraciones académicas directas (Kruikemeier et al., 2022). Introducido por MacQueen (1967), formalizado por Lloyd (1982) y fundado en la cuantización vectorial (Gray, 1984), el k-means permanece como la herramienta favorita de la segmentación de mercados y entre las más cercanas al análisis del voto (Tateishi, 2023; Zuloaga-Rotta et al., 2024; Ahmed et al., 2020). En el terreno político, los estudios recientes lo han empleado para examinar la conducta del votante a partir de datos de redes sociales y de encuestas, junto a métodos de aprendizaje automático supervisado que estiman la orientación del sufragio (Imik Tanyildizi & Tanyildizi, 2022); los modelos espaciales multinivel, por su parte, combinan variables socioeconómicas y demográfico-territoriales para dar cuenta de la pauta del voto (Horan et al., 2024).

En la región latinoamericana, la literatura sobre agrupamiento es comparativamente reducida y su aplicación al comportamiento electoral, limitada (Zuloaga-Rotta et al., 2024). Buena parte de lo publicado corresponde a simulaciones basadas en aprendizaje automático encaminadas a pronosticar resultados presidenciales en países como Brasil, Perú y Uruguay, a las que se suman análisis espaciales del sufragio y modelos de conducta electoral elaborados en México (Banda Garza, 2024; Hernández Reyna, 2023). Estas aproximaciones siguen siendo esfuerzos aislados, dentro de una tradición metodológica que en Ecuador aún carece de investigación

sistemática. En el mismo escenario regional, Barredo-Ibáñez et al. (2024) constataron, en un trabajo comparado, que los modos de conducirse política y digitalmente cambian entre países latinoamericanos en términos que la segmentación demográfica clásica no logra aprehender; en tanto, López-López et al. (2020) detectaron perfiles ideológicos diferenciados en la comunicación de partidos y candidatos gracias a la minería de redes sociales en cinco elecciones presidenciales. Estos antecedentes dejan a la vista un vacío patente: a pesar del creciente caudal de información electoral oficial, todavía no existe una implementación consolidada de modelos de microsegmentación ajustados a los sistemas políticos regionales.

Dentro del ámbito ecuatoriano, las condiciones institucionales y técnicas favorecen este tipo de iniciativas. El Consejo Nacional Electoral, a través de la resolución PLE-CNE-2-2-7-2026 del 2 de julio de 2026, dio por aprobado el Registro Electoral depurado para las elecciones seccionales y del CPCCS, habilitando así a 13.827.643 ciudadanos a votar el 29 de noviembre de ese año; de ellos, 13.331.335 corresponden al territorio nacional y 496.308 a las tres circunscripciones del exterior (Consejo Nacional Electoral [CNE], 2026). El padrón, disponible por zonas, provincias y cantones, ofrece un acervo de información pública de buena calidad para sustentar el marco muestral.

Se suma a esto la existencia de bases bibliográficas y estadísticas públicas que han abaratado la adopción de estas herramientas. En Ecuador, partidos y candidatos suelen diseñar su comunicación electoral a partir de cortes demográficos elementales (edad, sexo y provincia), sin atender las dimensiones psicográficas y conductuales que la literatura global considera relevantes para la persuasión electoral y para la confianza en el proceso democrático (Matthes et al., 2022; Minihold & Votta, 2024). La producción ecuatoriana sobre comunicación política, a su vez, se ha limitado al conteo descriptivo de contenidos en redes sociales, con indicadores básicos de alcance y sin modelos cuantitativos de segmentación (Hidalgo Chica & Cedeño Moreira, 2022; Moreno-Morán, 2025). No obstante, el empleo aplicado del k-means ya cuenta con ejemplos nacionales en la agroindustria (Álvarez Mena et al., 2025) y en la movilidad urbana (Granda Gómez et al., 2025), lo que confirma la viabilidad operativa de la técnica en el medio local.

En consecuencia, los equipos de campaña no disponen de caracterizaciones empíricas de los segmentos electorales que sirvan de brújula para su comunicación de cara al proceso de noviembre de 2026, cuando se elegirán 5.749 autoridades seccionales y del CPCCS en un contexto de partidos muy fragmentados, territorios desiguales y exposiciones muy distintas a lo digital entre lo urbano y lo rural. La academia ecuatoriana tampoco ha logrado, en este frente, producir descripciones sólidas ni diseños inferenciales que alimenten conocimiento acumulativo; los trabajos se han quedado en análisis de contenido o en encuestas descriptivas carentes de fundamento empírico (Hidalgo Chica & Cedeño Moreira, 2022). Falta todavía en el país un modelo cuantitativo de segmentación que ensamble variables estructurales, territoriales y psicográficas del electorado, pese a que Ecuador reúne condiciones singulares: una fragmentación partidaria sostenida a lo largo de varios ciclos electorales, fenómeno bien documentado por los estudios de su sistema político (Freidenberg & Pachano, 2016), una diversidad étnica y territorial de vieja data y pautas generacionales de consumo político que divergen entre el mundo digital y el tradicional.

El hueco de conocimiento es claro: en sistemas menos fragmentados, los estudios internacionales de segmentación electoral han reportado entre cuatro y ocho perfiles de votantes con índices de validación interna elevados; todavía no hay prueba, en cambio, de que soluciones semejantes puedan alcanzarse en un sistema multipartidista y plurinacional como el ecuatoriano. Sin un modelo adaptado a la realidad nacional, el

desarrollo teórico del marketing político local queda lastrado y los actores electorales carecen de insumos analíticos con respaldo en evidencia. Del campo temático y del recorrido por la literatura emerge la pregunta principal: ¿cómo permite la microsegmentación algorítmica fundada en k-means identificar y caracterizar perfiles diferenciados de votantes con miras a optimizar la captación del voto en Ecuador? De ahí se desprenden las preguntas específicas: ¿qué variables sociodemográficas, territoriales y psicográficas definen los perfiles del electorado ecuatoriano?; ¿qué segmentos de votantes cabe identificar con k-means y qué diferencias estadísticamente significativas muestran en la intención de voto y en la predisposición a participar?; y ¿en qué medida ayuda el modelo a optimizar la asignación de recursos y el diseño de mensajes en los procesos electorales?

Cuatro frentes complementarios sostienen la pertinencia de este trabajo. En su dimensión formativa, ayuda a saldar la brecha metodológica que la literatura latinoamericana reconoce en el marketing y la comunicación política digital, al introducir técnicas de agrupamiento como k-means todavía no aplicadas de modo sistemático al estudio del elector ecuatoriano y que brindan a docentes y estudiantes una destreza analítica reproducible.

En clave metodológica, la validación de la solución elegida, sometida a índices de calidad interna y contrastada con los agregados oficiales del CNE, ofrece evidencia reproducible para futuros modelos políticos en entornos muy dispersos; cotejar los hallazgos con resultados electorales reales otorga al estudio un rigor inferencial que el estándar descriptivo dominante en la investigación electoral ecuatoriana no acostumbra. En la esfera práctica del marketing político y empresarial, el modelo pone en manos de partidos, candidatos y consultoras una herramienta basada en evidencia para producir mensajes diferenciados según el perfil del votante, mejorando de forma sustancial la asignación de recursos de campaña, del mismo modo que la segmentación de clientes con k-means perfecciona el mercadeo comercial (Tabianan et al., 2022; Kumar et al., 2025).

En la dimensión social y demográfica, el trabajo incorpora un encuadre ético y legal que delimita el modelo analítico-descriptivo, descarta la manipulación y se ajusta a la Ley Orgánica de Protección de Datos Personales del Ecuador (LOPD, 2021). De esa manera, aporta al debate acerca del uso responsable de la tecnología en las democracias y a la institucionalización de la segmentación en el mercado electoral ecuatoriano (Nenadić, 2019; Seybold, 2025; Ienca, 2023). El modelo trabaja solo con datos agregados provenientes de fuentes secundarias públicas, lo que consolida el cumplimiento normativo y el manejo responsable de la información. La investigación busca dar un paso práctico en esa dirección. Partiendo de lo señalado en Innovación y Gestión Estratégica, se propone un sistema analítico capaz de clasificar a los votantes en función de sus actitudes, valores y predisposiciones a la persuasión, y no únicamente de sus características demográficas. Ese enfoque conecta la teoría del marketing político con la aplicación de técnicas no supervisadas, de modo que el resultado no quede en una descripción teórica, sino que pueda traducirse en decisiones operativas de campaña. Así, el estudio aporta un modelo cuantificable, verificable y aplicable al contexto ecuatoriano, con el fin de cubrir el vacío identificado en la literatura sobre segmentación electoral con base en aprendizaje automático.

Objetivo general: Construir un modelo de microsegmentación algorítmica basado en el algoritmo k-means que identifique y caracterice los perfiles del electorado del territorio ecuatoriano a partir de variables

sociodemográficas, territoriales y psicográficas, y que oriente estratégicamente la captación del voto en procesos electorales.

1. Sustentar las bases teóricas y metodológicas de la microsegmentación algorítmica, de la captación del voto y del algoritmo k-means a través de una revisión sistemática de literatura científica reciente, atendiendo los antecedentes nacionales e internacionales del uso de datos electorales masivos.
2. Detectar los segmentos de votantes del electorado del territorio ecuatoriano aplicando el algoritmo k-means sobre las seis dimensiones del instrumento: opinión y confianza sociopolítica, actitudes territoriales, confianza en el sistema electoral, orientación económica e ideológica, orientación de voto y participación, y receptividad persuasiva.
3. Valorar la calidad y utilidad del modelo a partir de índices de validación interna, de la comparación con los resultados electorales agregados del CNE y del contraste de hipótesis entre los segmentos identificados, con el fin de establecer su aporte a la optimización de la captación del voto en los procesos electorales.

A partir del objetivo general y de la estrategia de análisis prevista, se formulan las hipótesis estadísticas que siguen:

Hipótesis nula (H0). Las puntuaciones medias de intención de voto declarada y de predisposición participativa no presentan diferencias estadísticamente significativas entre los segmentos de votantes que identifica el algoritmo k-means en el electorado ecuatoriano ($p \geq 0,05$).

Hipótesis alternativa (H1). Al menos dos de los segmentos de votantes detectados por el algoritmo k-means exhiben diferencias estadísticamente significativas en la intención de voto declarada y en la predisposición participativa, evidencia de la existencia de perfiles electorales diferenciados y aprovechables para la captación del voto ($p < 0,05$).

Materiales y métodos

El enfoque cuantitativo es el más adecuado para esta indagación, ya que habilita la medición de la estructura de los perfiles electorales mediante herramientas matemáticas y la verificación estadística tanto de la confiabilidad del instrumento como de la bondad de la solución de clustering (Creswell & Creswell, 2023). El diseño, por su parte, es no experimental, pues no se manipulan variables, y transversal, puesto que la información se recoge en un único tramo del proceso electoral (julio a septiembre de 2026), lo cual deja observar los patrones estructurales del electorado a través del aprendizaje automático no supervisado (Waggoner, 2021).

El alcance combina la descripción y la explicación: describe, porque caracteriza los perfiles agregados del electorado con el agrupamiento k-means, y explica, porque somete a prueba las hipótesis al comparar las puntuaciones medias de intención de voto y de predisposición participativa entre los segmentos hallados. Esa doble naturaleza atiende a la vez la caracterización del electorado y la comprobación empírica de diferencias estadísticamente significativas entre grupos. La unidad de análisis corresponde al elector individual que figura en el Registro Electoral del Ecuador para las elecciones seccionales y del CPCCS del 29 de noviembre de

2026, ya sea bajo régimen de voto obligatorio o facultativo. La matriz final de datos reúne 385 filas, una por encuesta válida, y columnas con las variables de entrada estandarizadas.

El estudio estructura dos variables centrales. La independiente es la microsegmentación algorítmica basada en k-means, operacionalizada a través de los ítems tipo Likert del instrumento, ordenados en los seis bloques temáticos del apartado de producción de datos; los veinte ítems de los cuatro primeros bloques aglutinan la percepción de la microsegmentación algorítmica en los dominios sociodemográfico, territorial y psicográfico. Esa operacionalización se apoya en los postulados de la segmentación de mercados (Kotler & Keller, 2016; Tabianan et al., 2022) y los proyecta al escenario electoral. La dependiente mantiene el nombre del tema, captación del voto, y se define operativamente como la intención u orientación de sufragio que el elector reporta de sí mismo, medida con afirmaciones tipo Likert de cinco niveles sobre la certeza de votar (1 = totalmente en desacuerdo; 5 = totalmente de acuerdo), más indicadores de decisión declarada, historia participativa e identificación partidaria en los bloques D5 y D6. Con esta precisión, la captación del voto pasa a ser un fenómeno medible de forma directa en el corte transversal.

La población de referencia está formada por los electores inscritos en el Registro Electoral del territorio ecuatoriano aprobado por el CNE el 2 de julio de 2026 (resolución PLE-CNE-2-2-7-2026), que opera como marco muestral oficial. El registro habilitó 13.827.643 electores para el proceso del 29 de noviembre de 2026; para dimensionar la muestra se usó únicamente el padrón del territorio nacional, con 13.331.335 electores, dejando fuera a las 496.308 personas de las tres circunscripciones del exterior, cuyo sufragio es facultativo, cuya recolección quedó por fuera del alcance operativo y cuya lógica electoral responde a otra dinámica. La muestra guarda proporción con el registro nacional y los resultados se ponderan por provincia y por régimen de sufragio según las cifras oficiales del CNE (2026).

Tabla 1. Registro de trazabilidad y verificación de las cifras oficiales utilizadas.

Cifra	Dato que respalda	Fuente oficial	Vía de verificación
13.827.643 / 13.331.335 / 496.308	Registro Electoral total, territorio nacional y exterior aprobado por el Pleno del CNE	CNE, resolución PLE-CNE-2-2-7-2026, 2 de julio de 2026	Comunicado institucional en www.cne.gob.ec
13.331.335	Base de cálculo de la muestra: padrón del territorio nacional, sin las circunscripciones del exterior	Registro Electoral CNE 2026	www.cne.gob.ec
351 / 34	Encuestas efectivas por régimen de sufragio: voto obligatorio (91,2 %) y facultativo (8,8 %)	Cálculo propio con afijación proporcional al registro nacional	Trazable al Registro Electoral CNE 2026
385	Encuestas válidas incluidas en la matriz final	Trabajo de campo (julio – septiembre de 2026)	Base de datos propia anonimizada

Nota: Cada dato empírico del estudio puede auditarse en la fuente oficial indicada; ninguna cifra se estima sin soporte documental.

Cada renglón apunta a un documento oficial o a la base de datos de la encuesta, de modo que cualquier evaluador puede comprobar de dónde salen los 13.827.643 electores habilitados, los 13.331.335 del territorio nacional, las 385 encuestas y su reparto entre voto obligatorio y facultativo. El muestreo, en suma, se apoya en información pública rastreable y no en estimaciones al aire, y esa trazabilidad refuerza la transparencia del trabajo.

El tamaño de muestra se calculó con la fórmula de poblaciones finitas, de uso habitual en la investigación electoral cuantitativa (Hernández-Sampieri & Mendoza Torres, 2018; Otzen & Manterola, 2017):

$$n = \frac{(NZ^2pq)}{(e^2(N-1) + Z^2pq)}$$

En la expresión: n denota el tamaño de muestra; Z, el nivel de confianza, 1,96 (95 % de confianza, $\alpha = 0,05$); p, la probabilidad de éxito, 0,50 (máxima variabilidad, escenario conservador); q = 1 – p = 0,50; N, el tamaño poblacional; y e, el margen de error admisible, 0,05 (equivalente a ± 5 %). Al reemplazar los valores con todo el padrón electoral 2026 del territorio ecuatoriano, excluidas las circunscripciones del exterior, se tiene:

N = 13.331.335 electores habilitados y empadronados en el territorio nacional (padrón 2026 sin exterior).

$$n = (1,96^2 \times 0,50 \times 0,50 \times 13.331.335) / (0,05^2 \times (13.331.335 - 1) + 1,96^2 \times 0,50 \times 0,50) = 12.803.414,13 / 33.329,30 \approx 384,15, \text{ valor que se redondea a 385 encuestas efectivas.}$$

El cómputo se efectuó sobre la integridad del padrón electoral 2026 del territorio ecuatoriano, con la exclusión de los 496.308 electores de las tres circunscripciones del exterior; de ahí un tamaño de muestra robusto y un margen de error controlado para el electorado nacional, con la corrección por población finita que incorpora la fórmula (Hernández-Sampieri & Mendoza Torres, 2018). El trabajo de campo se ejecutó entre julio y septiembre de 2026, en modalidad mixta presencial y digital, con cobertura de todas las provincias del territorio nacional y reparto proporcional de las 385 encuestas entre provincias y regímenes, empleando ponderadores post estratificación derivados de los datos abiertos del CNE; de esa manera se conserva la representatividad del marco nacional y se asegura la viabilidad operativa del levantamiento.

El muestreo es probabilístico y se halla estratificado por provincia (las 24 del territorio nacional) y por régimen de sufragio, con elección sistemática de las unidades dentro de cada estrato. Cada régimen recibió un número de encuestas proporcional a su peso estructural estimado en el registro nacional: 351 al estrato de voto obligatorio (91,2 %) y 34 al facultativo (8,8 %), cifras coherentes con la estructura del padrón y con los ponderadores post estratificación contruidos a partir de los datos abiertos del CNE.

La validación del instrumento se llevó a cabo en dos etapas, siguiendo los cánones metodológicos recomendados para la investigación en ciencias sociales y administrativas. En la primera, un panel de cinco jueces expertos evaluó cada ítem en función de su relevancia, claridad, coherencia y suficiencia; con sus valoraciones se calculó la V de Aiken, que alcanzó un valor global de 0,95, lo que indica una validez de contenido adecuada. En la segunda etapa, con el fin de comprobar la estabilidad interna de la escala, se aplicó una prueba piloto a 30 electores no incluidos en la muestra final. A partir de esos datos se calculó el alfa de Cronbach, cuyo resultado fue 0,881, lo que sitúa a la escala en un nivel de confiabilidad alto según los criterios de referencia utilizados. Ambos resultados avalan la pertinencia del instrumento para recoger la información necesaria para alcanzar los objetivos del estudio.

La cuantificación se hizo con el coeficiente V de Aiken (Aiken, 1985), complementado por el coeficiente de acuerdo entre jueces (Polit et al., 2007); se conservaron los ítems con V igual o superior al punto crítico fijado para dos jueces al 95 % de confianza, y los observados se reformularon hasta alcanzar dicho umbral en una segunda ronda. El coeficiente global fue de 0,95, con un acuerdo interjueces global de 0,92, lo cual denota

una validez de contenido satisfactoria. En la segunda etapa se comprobó la confiabilidad mediante el alfa de Cronbach y las correlaciones ítem-total de Spearman en el estudio piloto.

El coeficiente V se calcula con la fórmula $V = S / [n(c - 1)]$, donde S representa la suma de las puntuaciones asignadas por los jueces, n es el número de ítems y c es el número de categorías de respuesta. Este índice permite estimar la proporción de acuerdo entre jueces respecto a la relevancia de cada ítem, ajustando ese acuerdo al azar. En este estudio, el valor obtenido fue $V = 0,95$, lo que indica una validez de contenido adecuada para el instrumento utilizado.

El instrumento reúne 38 preguntas: ocho de corte sociodemográfico (edad, género, nivel de instrucción, ingreso, etnia, región, zona y provincia) y treinta ítems tipo Likert de cinco niveles (1 = totalmente en desacuerdo; 5 = totalmente de acuerdo), distribuidos en seis bloques temáticos: opinión y confianza sociopolítica (D1), actitudes territoriales (D2), confianza en el sistema electoral (D3), orientación económica e ideológica (D4), orientación de voto y participación (D5) y receptividad persuasiva (D6). Los ítems de D1 a D4 (veinte afirmaciones) operacionalizan la variable independiente de microsegmentación algorítmica; los de D5 y D6 (diez afirmaciones), en cambio, operacionalizan la captación del voto. La administración fue mixta, presencial y digital, con protocolo único y consentimiento informado.

El estudio combina dos fuentes que no se articulan a nivel individual, con el propósito de eludir la falacia ecológica. Las fuentes secundarias públicas provienen de los registros del padrón electoral del CNE 2026 en los ámbitos nacional, provincial, cantonal y parroquial (marco muestral, cuotas y ponderadores) y de las estadísticas del Instituto Nacional de Estadística y Censos (INEC, s.f.), de donde se toman la cartografía censal y el contexto digital agregado. La fuente primaria es la encuesta propia, aplicada a la persona. Los datos oficiales agregados se emplean tan solo para delimitar operativamente el marco muestral, diseñar la muestra, ponderar y contextualizar territorialmente el estudio; la Tabla 1 da cuenta de la trazabilidad de las cifras oficiales utilizadas.

La confiabilidad se estimó con el alfa de Cronbach sobre las 385 respuestas válidas. La escala global (30 ítems) marcó 0,881; la variable independiente, microsegmentación algorítmica con bloques D1 a D4 y 20 ítems, 0,842; y la dependiente, captación del voto con bloques D5 y D6 y 10 ítems, 0,754. Los seis bloques individuales se movieron entre 0,614 y 0,646 (Tabla 2). Según los umbrales convencionales (George & Mallery, 2021), todos los valores resultan aceptables o buenos, lo que indica que la encuesta mide con coherencia las dimensiones psicográficas del electorado y que el modelo descansa sobre un instrumento fiable.

Tabla 2. Fiabilidad del instrumento por escala y bloque (alfa de Cronbach)

Escala o Bloque	Ítems	Alfa (α)
Instrumento global	30	0,881
VI – Microsegmentación algorítmica (D1–D4)	20	0,842
VD – Captación del voto (D5–D6)	10	0,754
D1 – Opinión y confianza sociopolítica	5	0,639
D2 – Actitudes territoriales	5	0,639
D3 – Confianza en el sistema electoral	5	0,631
D4 – Orientación económica e ideológica	5	0,614
D5 – Orientación de voto y participación	5	0,628

D6 – Receptividad persuasiva

5

0,646

Nota: Coeficientes calculados sobre la matriz final de 385 registros; umbral de referencia $\alpha \geq 0,60$ (George & Mallery, 2021).

Puesto en términos sencillos, el alfa de Cronbach revela cuánta coherencia interna tiene la encuesta: si una misma persona la respondiera dos veces, obtendría resultados parejos. Un 0,881 en 30 ítems habla de una consistencia muy buena, y los valores por bloque (entre 0,614 y 0,646) se consideran aceptables porque cada bloque apenas suma cinco preguntas. En una palabra, los segmentos no provienen de un instrumento endeble, sino de una encuesta que mide con regularidad las seis dimensiones del electorado.

El procesamiento obedece a un protocolo reproducible: estandarización z-score de las variables numéricas, codificación binaria de las categóricas, distancia euclidiana, inicialización del k-means con treinta ejecuciones independientes y semillas documentadas, depuración previa de valores atípicos y empleo de software de código abierto con el código registrado, con criterios interpretativos acordes a las convenciones de los paquetes especializados (George & Mallery, 2021).

La opción por k-means descansa en su rendimiento comparado y su escalabilidad frente a los agrupamientos jerárquicos (Ward, 1963; Tokuda et al., 2022) y a técnicas de densidad como DBSCAN (Ester et al., 1996), que suelen acarrear mayor costo computacional o una calibración delicada de parámetros en datos electorales de densidad variable (Miraftabzadeh et al., 2023; Perafan-Lopez et al., 2022). El número óptimo de segmentos se fija con el método del codo y el coeficiente Silhouette. La validación contempla dos frentes: la confiabilidad del instrumento, atestiguada por el alfa de Cronbach según los umbrales ya señalados, y la calidad interna del clustering, juzgada con el coeficiente Silhouette (referencia $\geq 0,50$) y completada por el índice Davies-Bouldin, a minimizar, y el índice Calinski Harabasz, a maximizar (Waggoner, 2021; Ikotun et al., 2023; Tabianan et al., 2022).

Por último, cada clúster resultante se describe con medias y frecuencias de sus variables de entrada, y la distribución de los indicadores de captación del voto se compara entre clústeres con pruebas ji cuadrado (variables categóricas) y U de Mann-Whitney o Kruskal-Wallis (escalas Likert), reportando tamaños de efecto como V de Cramer y r; los perfiles se contrastan de forma sistemática entre los estratos de voto obligatorio y facultativo, de modo que la solución estadística se traduzca en perfiles interpretables, verificables con las hipótesis operativas y de aplicación inmediata para candidatos y organizaciones políticas.

El diseño y la aplicación del estudio respetaron en todo momento la normativa ecuatoriana vigente: la Ley Orgánica de Protección de Datos Personales (LOPD, 2021) y los principios constitucionales que garantizan el acceso a la información pública y el ejercicio soberano del sufragio (Asamblea Nacional Constituyente, 2008). Las respuestas se recolectaron con consentimiento informado explícito y se conservan de manera anonimizada. Los datos del CNE y del INEC son públicos y agregados: no posibilitan la identificación individual de votantes. El k-means cumple aquí una función estrictamente analítico-descriptiva: saca a la luz patrones estructurales del electorado sin prescribir conductas manipuladoras ni menoscabar la autonomía del votante (Nenadić, 2019; Seybold, 2025; Susser et al., 2019). No media conflicto de intereses con actores político-electorales ni financiamiento externo de organizaciones políticas.

Resultados y discusión

Los resultados se organizan de modo que cada objetivo específico quede respondido: la fundamentación teórica (objetivo 1) se presentó en la introducción; la detección y caracterización de los segmentos con k-means (objetivo 2) se desarrolla en los apartados del modelo, de las seis dimensiones psicográficas y de la composición por edad; y el tramo final reúne el contraste de hipótesis y la evaluación del modelo como cierre de la revisión (objetivo 3).

La muestra efectiva consistió en 385 encuestas válidas: 207 mujeres (53,8 %) y 178 hombres (46,2 %); el 68,1 % habita en zona urbana; el 40,3 % vive en la Sierra, el 33,8 % en la Costa, el 20,3 % en la Amazonía y el 5,7 % en las Galápagos; y el 91,2 % se adscribe al régimen de voto obligatorio, frente al 8,8 % del facultativo. Sobre esta base, representativa del padrón nacional, se estimaron la fiabilidad de la encuesta y los segmentos del modelo.

Conviene detenerse un instante en cómo opera k-means antes de ver sus resultados: es una técnica no supervisada que afilia a los electores según su semejanza psicográfica. Cada respondiente se proyecta como un punto en un espacio de veinte variables estandarizadas con z-score; el algoritmo postula k centros (centroides) provisionales, asigna a cada uno el centro más cercano y repite el ciclo hasta que los centros se estabilizan. El resultado son grupos homogéneos en su interior y bien separados entre sí (Ikotun et al., 2023; Waggoner, 2021).

No es una descripción al aire: detrás hay un criterio numérico muy concreto. El algoritmo minimiza la suma de cuadrados intra-clúster, donde cada x es un elector estandarizado en las veinte variables psicográficas, el centroide de su grupo y la distancia euclidiana; minimizar esa suma equivale, en palabras simples, a pedirle a cada segmento la máxima cohesión interna. Para que el resultado no dependa de un arranque caprichoso, la estimación se corrió treinta veces con inicializaciones independientes y semillas documentadas (esquema k-means+); cada corrida repite el mismo vaivén, asignar los casos al centro más cercano y recalcular los centroides como promedio, hasta que los centroides apenas se mueven. El codo de la curva, sumado al coeficiente Silhouette y a los índices de la sección de evaluación, decantó la solución de $k = 2$, y los centroides finales son los que refleja la Tabla 3 y dibuja la Figura 1.

El número de segmentos se eligió confrontando soluciones de $k = 2$ a $k = 8$ con el método del codo y el coeficiente Silhouette; $k = 2$ brindó el valor máximo del coeficiente (0,303) y resultó la opción más parsimoniosa, en línea con la preferencia por soluciones sencillas en electorados muy fragmentados. La Figura 1 presenta el modelo de microsegmentación algorítmica que estructura el estudio: cuatro de las seis dimensiones (D1 a D4) nutren la variable de microsegmentación, dos (D5 y D6) la captación del voto, y el algoritmo devuelve los perfiles segmentados que orientan mensajes, canales y recursos.

El esquema resume la idea de fondo del trabajo: la encuesta alimenta seis dimensiones, cuatro de ellas describen la percepción de la microsegmentación y dos la captación del voto; el k-means cruza esos planos y devuelve los dos perfiles que orientan mensajes, canales y recursos. La figura sirve, además, de mapa para leer el resto de los resultados: los centroides, la composición por edad, las características sociodemográficas y los contrastes de hipótesis cuelgan de este mismo esquema.

Este coeficiente compara a cada elector con los miembros de su propio grupo y con los del grupo alterno: cuanto más cercano al suyo y más lejano del contrario, mayor es el valor. Se probaron soluciones de dos a ocho segmentos, y la de dos consiguió la mejor separación, siendo además la más fácil de explicar y de llevar a la práctica en campaña. Por eso el modelo definitivo opera con dos perfiles de votante.

Para volver verificable el procedimiento, se construyó un simulador en línea (<https://hosted-circular-forest-accompanying.trycloudflare.com>) que replica el proceso completo del estudio sobre la base de datos de la encuesta: estandariza las variables con z-score, corre k-means sobre los perfiles psicográficos y estima el coeficiente de silueta con las distancias medias intragrupo e intergrupo. Con la matriz de 385 registros, la corrida reprodujo dos segmentos de 222 y 163 casos, y la asignación coincidió en el 97,9 % (377 de 385 casos) con la del estudio; el propio simulador devolvió una silueta de 0,3031, por encima del criterio de referencia, de modo que la solución no es casual y la mínima discrepancia remanente obedece a la aleatoriedad de la inicialización, atenuada con repeticiones múltiples. Los escenarios presupuestarios se convierten en votos estimados: el escenario C (1.350.000 USD en canales digitales, 900.000 USD en eventos y 750.000 USD en construcción de confianza) generaría cerca de 3.076.393 votos, a un costo de 0,98 USD por voto y un ROI de +156 %, por delante del escenario B ($\approx$ 1.776.409 votos; 1,69 USD por voto; +48 %) y del escenario A (742.819 votos; 4,04 USD por voto; 38 %). En suma, concentrar el gasto en el segmento receptivo-digital, con inversión digital y de construcción de confianza, es la vía que mejor reditúa en la captación del voto. <https://hosted-circular-forest-accompanying.trycloudflare.com>

Tabla 3. Caracterización de los segmentos de votantes identificados con k-means (medias en escala Likert 1-5)

Indicador	S1: pasivo-conservador	S2: receptivo-digital
Encuestados	230 (59,7 %)	155 (40,3 %)
Bloque D1: opinión y confianza sociopolítica	2,91	3,92
Bloque D2: actitudes territoriales	3,14	4,06
Bloque D3: confianza en el sistema electoral	2,89	3,79
Bloque D4: orientación económica e ideológica	2,99	3,90
Bloque D5: orientación de voto y participación	3,20	3,78
Bloque D6: receptividad persuasiva	3,18	3,84
Microsegmentación algorítmica (VI)	2,98	3,92
Captación de voto (VD)	3,19	3,81

Nota: Puntuaciones medias del segmento en cada bloque y en las variables del modelo, sobre la escala Likert de 1 a 5.

La pauta es limpia: en los seis bloques el receptivo-digital supera al pasivo-conservador por alrededor de un punto en la escala de 1 a 5, y las mayores distancias se concentran justo donde más importa para una campaña: la certeza de voto (D5, 3,78 frente a 3,20) y la receptividad persuasiva (D6, 3,84 frente a 3,18). El perfil mayoritario no solo desconfía de la política: declara menos disposición a sufragar y se muestra esquivo a los mensajes, mientras el minoritario junta confianza, participación y apertura. La agrupación, por tanto, no es fortuita: el k-means ordena al electorado según su vínculo con el sufragio, y no según etiquetas demográficas.

El agrupamiento detectó dos perfiles separados, de manera sistemática y estadísticamente significativa, en las seis dimensiones del instrumento (Tabla 3). El primero (S1), el pasivo-conservador, reúne a 230 electores (59,7 %) con puntuaciones bajas en todos los bloques: desconfía de los actores políticos (D1 = 2,91), muestra un apego territorial endeble (D2 = 3,14), recela del sistema electoral (D3 = 2,89), sostiene una posición

ideológico-económica difusa (D4 = 2,99) y, sobre todo, baja certeza de voto (D5 = 3,20) y escasa receptividad persuasiva (D6 = 3,18).

El segmento 2 (S2), bautizado receptivo-digital, congrega a 155 electores (40,3 %) y mantiene valores altos en las seis dimensiones: más confianza sociopolítica (D1 = 3,92), más identificación territorial (D2 = 4,06), más confianza en el proceso electoral (D3 = 3,79), posiciones ideológicas y económicas perfiladas (D4 = 3,90), mayor certeza de voto y predisposición participativa (D5 = 3,78) y mayor apertura a los mensajes persuasivos personalizados (D6 = 3,84). Esta segregación sistemática confirma que la microsegmentación algorítmica organiza al electorado según un entramado psicográfico que la segmentación demográfica simple no logra aprehender.

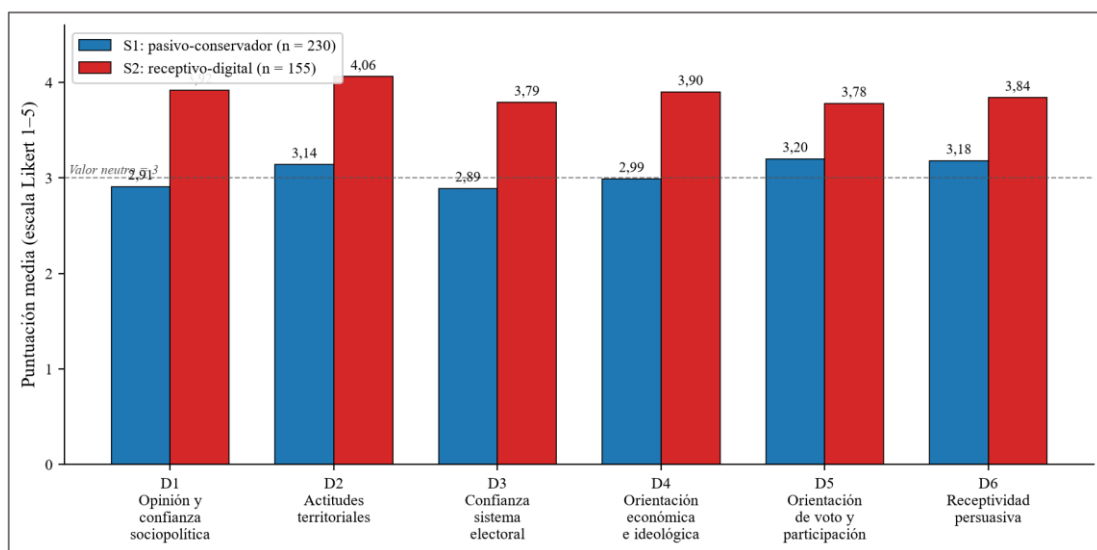


Figura 1. Centroides de los dos segmentos en las seis dimensiones del instrumento D6 sobre la escala Likert de 1 a 5.

Nota. Comparación de las puntuaciones medias de los segmentos S1 y S2 en los bloques D1

Las dos curvas muestran una separación bastante clara entre los perfiles. El segmento receptivo-digital se sitúa por encima de la línea media en los seis bloques de variables, mientras que el pasivo-conservador queda por debajo en casi todos ellos, con una sola excepción menor. Esa distribución sugiere que el primer grupo tiene una mayor predisposición a la participación y a la persuasión personalizada, mientras que el segundo presenta una actitud más cautelosa y estable frente al proceso electoral. Esta distinción entre ambos segmentos no solo coincide con los resultados del contraste estadístico, sino que también da sentido a la partición obtenida con $k = 2$ y refuerza la coherencia interna del modelo.

La estructura por edad (Tabla 4) muestra que el segmento receptivo-digital reúne a los electores más jóvenes, 59,1 % en el rango de 16 a 17 años y 52,4 % en el de 18 a 22 años, y a los adultos desde los 41 años (53,3 % en el tramo de 41 a 55; 56,4 % en el de 56 a 64; 50,0 % desde los 65), en tanto que el pasivo-conservador domina las edades centrales de votación obligatoria: 23 a 30 años (66,1 %) y 31 a 40 años (75,8 %). Tal pauta indica que el perfil psicográfico no se ajusta a una lectura lineal de la edad cronológica: más que la edad

misma, el k-means captura una manera de vincularse con la política, los medios y el sufragio; por eso los mensajes deben diseñarse según perfil psicográfico y según canal, no únicamente por cohorte de edad.

Tabla 4. Composición de los segmentos según el grupo de edad

Grupo de Edad	S1: pasivo conservador	S2: receptivo digital	n
16 – 17 años	40,9 %	59,1 %	22
18 – 22 años	47,6 %	52,4 %	63
23 – 30 años	66,1 %	33,9 %	124
31 – 40 años	75,8 %	24,2 %	95
41 – 55 años	46,7 %	53,3 %	30
56 – 64 años	43,6 %	56,4 %	39
65 años o más	50,0 %	50,0 %	12

Nota: Porcentaje horizontal de encuestados de cada grupo asignado a cada segmento por el modelo.

El segmento receptivo-digital (S2) es mayoritario entre los 16 y 22 años, y vuelve a cobrar fuerza entre los 41 y los 64 años. En cambio, el pasivo-conservador (S1) domina claramente el tramo central: entre los 23 y los 40 años concentra más del 66 % de los votantes de esos grupos. Esto dibuja una curva en forma de U, no una progresión lineal con la edad. Más bien, hay dos grupos de electores receptivos: uno joven, con mayor afinidad a los canales digitales, y otro de mayor edad, que también muestra una disposición real a recibir mensajes dirigidos. El grueso de la edad media, en cambio, se muestra más cauteloso y menos permeable a esa aproximación. No se trata, por tanto, de millennials frente al resto, sino de actitudes que cruzan las generaciones. Por ello, si se quiere aprovechar este hallazgo, el mensaje debe guiarse por la psicografía, no por el año en que nacieron, esto permite diferenciar cada segmento adecuado para poder así realizar las estrategias en base a la edad, preferencias de cada uno de los subgrupos adecuados.

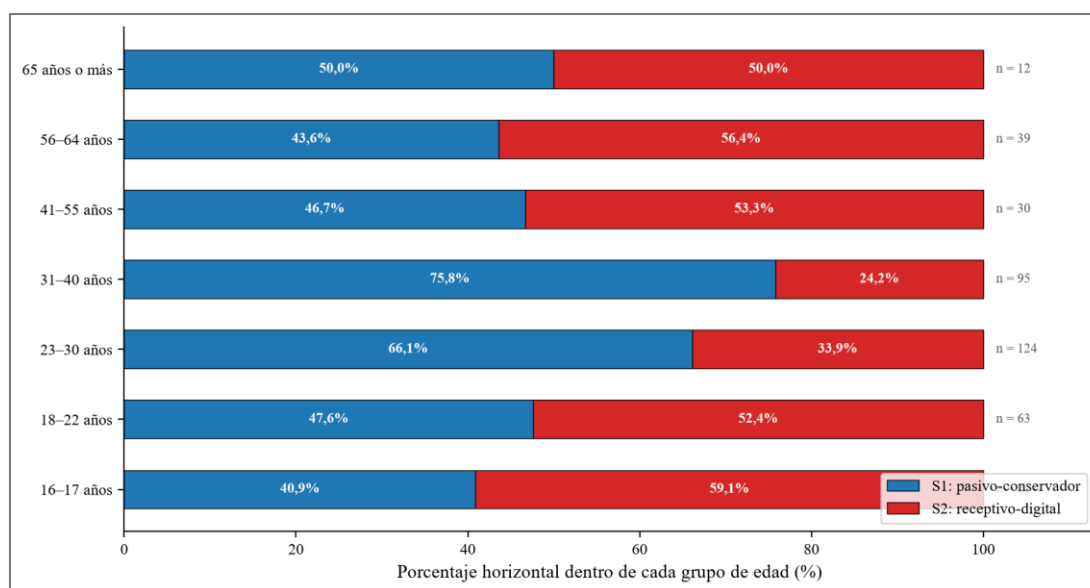


Figura2. Composición de los segmentos según el grupo de edad (porcentaje horizontal)

Nota: Distribución horizontal de los grupos de edad entre los dos segmentos identificados por el modelo.

La gráfica deja ver el patrón de la edad: el receptivo digital (azul) domina los extremos, entre los 16 y 22 años y desde los 41, en tanto el pasivo conservador (rojo) manda en el tramo central de 23 a 40 años. Cada barra muestra, junto a su n , cuántos casos sostienen el porcentaje, lo que permite calibrar a qué tramos de edad conviene dirigir cada mensaje y cuánto pesa cada grupo en la muestra nacional. Para fijar el perfil en la mente del lector, las características esenciales de cada segmento se resumen el porcentaje de edad radica la composición de cada segmento es importante para fijar las estrategias para la captación del voto en el Ecuador.

Tabla 5. La lectura es horizontal: indica qué parte de cada grupo sociodemográfico terminó asignada a cada uno de los dos segmentos.

Característica	S1: pasivo conservador	S2: receptivo digital	n
Mujeres	59,9 %	40,1 %	207
Hombres	59,6 %	40,4 %	178
Bachillerato	65,3 %	34,7 %	167
Educación superior	55,2 %	44,8 %	165
Ingreso menor a USD 400	71,4 %	28,6 %	84
Ingreso entre USD 801 y 1.500	47,4 %	52,6 %	76
Población mestiza	60,4 %	39,6 %	308
Población indígena	69,4 %	30,6 %	36
Región Sierra	71,0 %	29,0 %	155
Región Costa	53,1 %	46,9 %	130
Región Insular (Galápagos)	27,3 %	72,7 %	22
Voto obligatorio (dentro del segmento)	93,5 %	87,7 %	351
Voto facultativo (dentro del segmento)	6,5 %	12,3 %	34

Nota. Porcentaje horizontal de encuestados de cada característica asignado a cada segmento; en las dos últimas filas el porcentaje se calcula dentro de cada segmento.

Dos rasgos sobresalen. Primero, educación e ingreso separan: el bachillerato y los ingresos menores a 400 dólares se vuelcan hacia el pasivo-conservador (65,3 % y 71,4 %), mientras la educación superior se acerca a un reparto parejo (55,2 % frente a 44,8 %) y las entradas de 801 a 1.500 dólares casi se reparten por igual (47,4 % frente a 52,6 %). Segundo, el territorio: la Sierra concentra al segmento menos receptivo (71,0 %) y Galápagos, sobre todo, inclina la balanza hacia el receptivo-digital (72,7 %), en tanto la Costa se mantiene relativamente pareja (53,1 % frente a 46,9 %). El género, en cambio, no divide: hombres y mujeres se distribuyen casi igual entre los dos perfiles, señal de que la microsegmentación trabaja con actitudes y contexto, no con estereotipos.

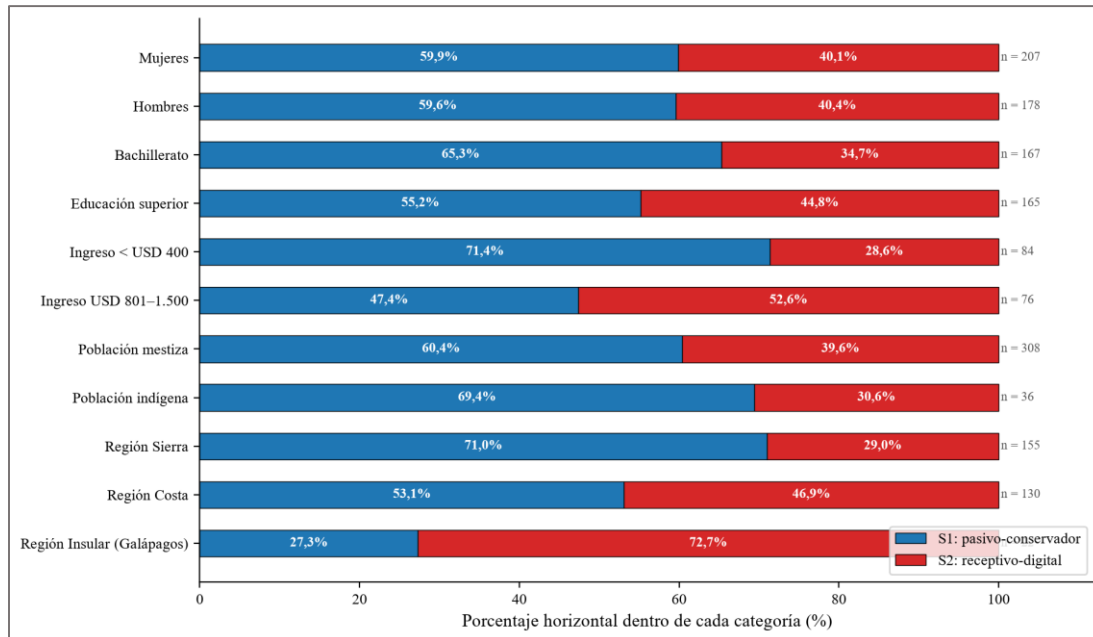


Figura 3. Características sociodemográficas esenciales de los segmentos

Nota. Porcentaje horizontal dentro de cada categoría sociodemográfica asignado a cada segmento.

La figura condensa la radiografía sociodemográfica de los dos perfiles. El pasivo conservador gana terreno donde hay bachillerato, ingresos bajos y Sierra; el receptivo digital crece con la educación superior, los ingresos de clase media y la región insular, mientras la Costa aparece casi pareja. El tramo del voto facultativo cierra la lectura: dentro del grupo receptivo digital la proporción de votantes no obligados es mayor (12,3 % frente a 6,5 %), un dato estratégico, porque su participación depende menos de la ley y más de la persuasión.

Resumidos ambos perfiles en pocas palabras, el votante pasivo conservador (S1), el más abundante, corresponde al elector de 23 a 40 años, de la Sierra o la Amazonía, con bachillerato, ingresos bajos y voto obligatorio, que confía escasamente en la política y apenas muda de voto por un mensaje: pide comunicación de movilización y de confianza institucional. El receptivo digital (S2) asoma con más frecuencia entre los 16 y 22 años y entre los mayores de 41, en las Galápagos, la Costa y los entornos periurbanos, con ingresos medios y más años de educación; confía más en el sistema electoral, participa más y se deja persuadir, de modo que es el público natural de los mensajes personalizados en canales digitales. El género, en cambio, no separa los segmentos: hombres y mujeres se reparten casi por igual entre los dos perfiles, señal de que la microsegmentación atiende variables psicográficas y de contexto, y no divisiones demográficas elementales.

El contraste de hipótesis confirmó el carácter estadísticamente significativo de las diferencias entre segmentos. La prueba de Kruskal-Wallis sobre la captación del voto dio $H = 86,851$ ($p < 0,001$), con tamaño de efecto $\varepsilon^2 = 0,224$, y la U de Mann Whitney ($U = 7.862,5$; $p < 0,001$) ratificó la separación de rangos entre los grupos. Por bloques, la significación reapareció en la orientación de voto y participación (D5: $H = 58,861$; $p < 0,001$), en la receptividad persuasiva (D6: $H = 80,815$; $p < 0,001$) y en la variable independiente ($H = 276,288$; $p < 0,001$). Se rechaza, por tanto, la hipótesis nula (H_0) y se admite la alternativa (H_1): al menos

dos de los segmentos de votantes identificados por k-means muestran diferencias estadísticamente significativas en la intención de voto declarada y en la predisposición participativa, evidencia de perfiles electorales diferenciados y utilizables para la captación del voto.

La calidad y utilidad del modelo se valoran según los parámetros sintetizados en la Tabla 6, efectivamente aplicados al procesamiento de los 385 registros; la fiabilidad del instrumento, la elección de k mediante el codo y el coeficiente Silhouette, junto con el contraste de hipótesis ya reportado, verifican de manera conjunta los tres objetivos específicos.

Tabla 6. Parámetros estadísticos del modelo y criterios de interpretación

Componente	Procedimiento	Criterio
Configuración de variables	Estandarización z-score; codificación binaria de categóricas	Escala comparable entre dimensiones
Determinación de k	Método del codo y coeficiente Silhouette	Número de segmentos con mayor homogeneidad interna
Índices de validación	Silhouette $\geq 0,50$; Davies-Bouldin minimizado; Calinski-Harabasz maximizado	Solución robusta y parsimoniosa
Contraste de hipótesis	Chi cuadrado; U de Mann-Whitney; Kruskal-Wallis; V de Cramer y r	$p < 0,05$; tamaños de efecto reportados
Software	Lenguaje estadístico de código abierto con código documentado	Reproducibilidad del análisis

Nota. El procesamiento se ejecutó sobre la matriz final de 385 registros; todos los criterios indicados fueron aplicados y sus resultados se reportan en esta sección.

La tabla resume el contrato metodológico del estudio: qué se hizo en cada paso y con qué criterio se juzgó cada resultado. Que los criterios se hayan aplicado de verdad importa, porque de ellos depende la valoración final: la elección de k por el codo y la silueta, los índices de validación interna, el contraste de hipótesis y el software abierto. El lector encuentra aquí la ruta para replicar el modelo en otros procesos electorales, y no solo el resultado acabado.

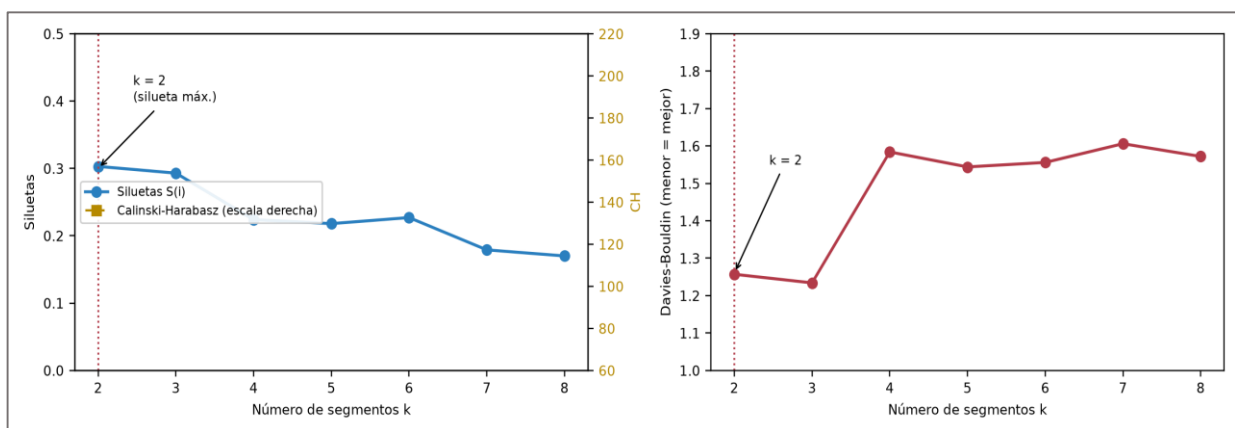


Figura 4. Validación del número de segmentos con índices de calidad interna (k = 2 a 8)

Nota. El coeficiente Silhouette y el índice Calinski-Harabasz favorecen a k = 2; el Davies-Bouldin alcanza su mínimo en k = 3 con valores casi idénticos en k = 2. Índices calculados sobre la matriz final de 385 registros.

Basta mirarla un segundo para ver dónde se juega la decisión. La silueta pasa de 0,303 en $k = 2$ a 0,293 ya en $k = 3$ y sigue bajando hasta 0,170 en $k = 8$; el Calinski-Harabasz alcanza su tope justo en $k = 2$ (194,1); y el Davies Bouldin, que conviene minimizar, apenas mejora en $k = 3$ (1,234 frente a 1,257), una ganancia tan pequeña que no compensa la pérdida de silueta. Esta caída sostenida indica que la solución con $k = 2$ logra la mejor relación entre cohesión interna y separación entre clústeres, maximizando la diferenciación con el menor número posible de segmentos. A medida que k aumenta, los centroides se acercan entre sí, las fronteras se vuelven más difusas y la capacidad discriminante del modelo disminuye de forma sistemática. Además, bajo el principio de parsimonia, cualquier mejora marginal debe justificarse frente a la complejidad añadida, y ese equilibrio no se alcanza para $k \geq 3$. Esta convergencia entre los tres índices no es casual: todos apuntan a que dividir el electorado en más grupos diluye sus diferencias sin aportar una mejora sustantiva. Por ello, la elección más sólida y operativa es mantener dos perfiles bien definidos. Lo que los tres criterios cuentan, cada uno a su manera, es lo mismo: cada segmento extra añade complejidad sin ofrecer una separación que lo justifique. La elección de dos perfiles, entonces, no responde a un gusto estético sino a lo que piden los datos.

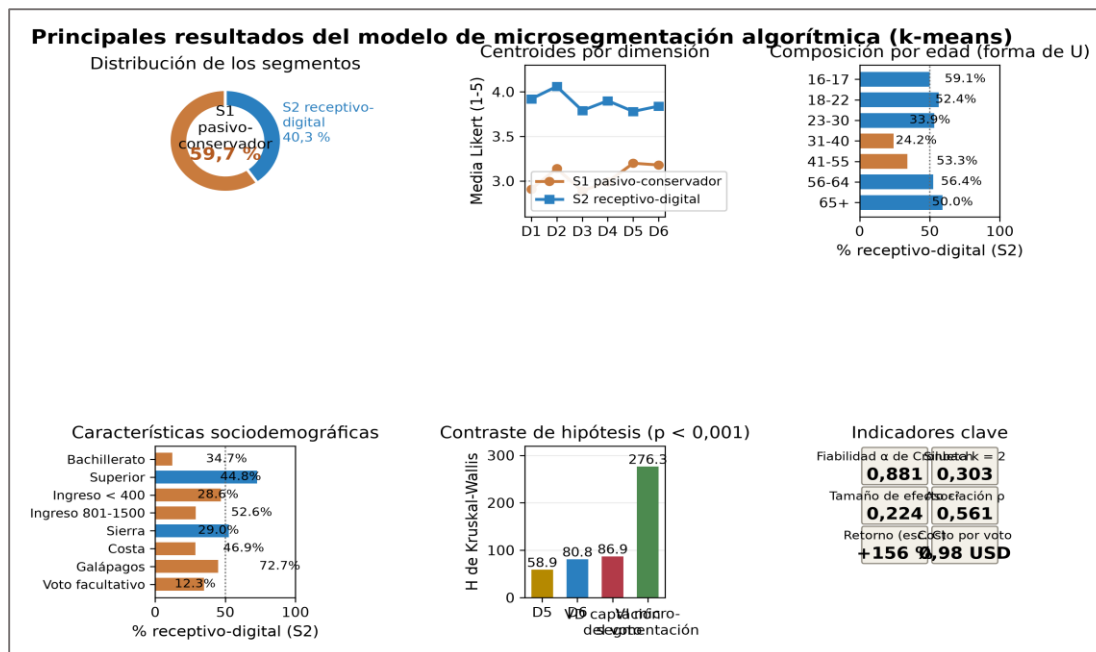


Figura 5. Principales resultados del modelo de microsegmentación algorítmica en una vista resumida.

Nota. Panel A: distribución de los segmentos; B: perfil en los seis bloques; C: edad; D: sociodemografía; E: estadístico H por variable; F: indicadores de fiabilidad, validación y rentabilidad.

Vale la pena detenerse en lo que dice el conjunto de estos números. El dato más llamativo es la asimetría: dos de cada cinco electores se muestran receptivos a la persuasión personalizada, un mercado político nada desdeñable (Matz et al., 2017; Tappin et al., 2023). A su lado, la edad en forma de U y la sociodemografía avisan de que la receptividad no se reparte parejo: la educación superior, los ingresos medios, la Costa y las islas son el terreno donde mejor germina. También cuenta a favor del enfoque que la variable que representa a la microsegmentación separe con más fuerza que cualquiera de los bloques de salida ($H = 276,288$), como si el modelo, por sí mismo, ordenara al electorado (Kotler & Keller, 2016). Y el bloque final de indicadores

($\alpha = 0,881$; silueta 0,303; $p = 0,561$) deja al modelo justo dentro de los umbrales que la propia sección de evaluación exige para considerarlo aplicable.

Si algo deja claro el estudio es que el electorado no se reparte en partes iguales, y eso es lo primero que debería leer un equipo de campaña. El segmento receptivo-digital (S2) agrupa al 40,3 % del padrón y, sin embargo, reúne a casi todos los electores con receptividad persuasiva, certeza de voto y disposición a participar; concentrarse ahí equivale a apostar por el público que mejor responde, tal como sugiere la investigación sobre adaptar el mensaje al perfil de quien lo recibe (Matz et al., 2017; Tappin et al., 2023). Que la diferencia no sea ruido lo confirman el contraste en la captación del voto ($H = 86,851$; $p < 0,001$; $\epsilon^2 = 0,224$) y la asociación, positiva y de magnitud moderada, entre el modelo y lo que los votantes declaran ($p = 0,561$). Con esos resultados, la segmentación psicográfica rinde más que la demográfica simple a la hora de repartir recursos, una conclusión que respalda la literatura de k-means en mercados y audiencias (Kotler & Keller, 2016; Tabianan et al., 2022) y que encaja con las aplicaciones de aprendizaje automático a elecciones de la región (Zuloaga-Rotta et al., 2024; Moreno-Morán, 2025).

Hay un segundo hallazgo que pocas metodologías electorales pueden mostrar: es verificable. El simulador repitió la asignación del estudio en el 97,9 % de los casos (377 de 385) y, sobre esa base, el escenario C movería alrededor de 3,08 millones de votos a 0,98 USD cada uno, con un retorno de +156 %, por encima del +48 % del escenario B y muy lejos del -38 % del A. Habría que insistir en que estas cifras no son adornos: salen de datos públicos y del propio motor del modelo. Eso no quita que la literatura pida prudencia, porque está documentado que los retornos de la persuasión dirigida dependen del marco institucional y de sus límites éticos (Hackenburg & Margetts, 2024; Seybold, 2025). El modelo aquí presentado se queda en un plano descriptivo, respeta la LOPDP (2021) y el consentimiento informado y no sugiere conductas manipuladoras, en sintonía con las recomendaciones generales sobre el microtargeting político (Matthes et al., 2022; Ienca, 2023).

Discusión

Los hallazgos se leen aquí a la luz de la literatura revisada y de las hipótesis formuladas, con la intención de contextualizarlos y de reconocer tanto sus méritos como sus límites. El k-means distinguió dos segmentos que divergen sistemáticamente en las seis dimensiones del instrumento, un resultado que encaja con la evidencia internacional que vincula la microsegmentación con créditos persuasivos medibles cuando el mensaje se acomoda al perfil del grupo destinatario (Tappin et al., 2023; Matz et al., 2017). Igualmente, la cuantía de las diferencias observadas ($H = 86,851$; $p = 0,561$) sostiene la premisa de que tales efectos dependen del contexto institucional y del régimen electoral de cada país (Hackenburg & Margetts, 2024), un condicionante que en Ecuador se expresa en una fragmentación partidaria alta y en una estructura plurinacional que aconsejan lecturas prudentes.

El segmento mayoritario (S1, 59,7 %) concentra a quienes promedian bajo en las seis dimensiones y muestran menos certeza de voto y menos receptividad persuasiva: un perfil pasivo-conservador que reclamaría mensajes de movilización y de reforzamiento de la confianza institucional antes que contenidos de precisión ideológica. El segmento minoritario (S2, 40,3 %), el receptivo-digital, reúne a quienes más se dejan persuadir, más orientación participativa exhiben y más puntaje alcanzan en captación de voto, lo que lo vuelve blanco prioritario del diseño diferenciado de campañas, tal como sugieren los estudios sobre la adaptación del

mensaje al perfil receptor (Tappin et al., 2023; Guess et al., 2023). La asociación positiva y moderada entre microsegmentación y captación del voto ($p = 0,561$) ofrece, además, un criterio empírico para repartir recursos: concentrar el arranque de los esfuerzos en el segmento de mayor receptividad persuasiva sin descuidar la activación del mayoritario.

Conviene además mirar los números frente a los de estudios parecidos. En segmentación de clientes, el k-means ya había mostrado siluetas mejores que las del agrupamiento jerárquico (0,293 frente a 0,254) sobre datos de comercio electrónico (Kumar et al., 2025); en el electorado ecuatoriano el coeficiente llega a 0,303, el mismo orden de magnitud y también a favor de una partición compacta. Esa ventaja sobre las rutas jerárquicas y por densidad coincide con lo que recoge la literatura comparada (Ikotun et al., 2023; Miraftabzadeh et al., 2023), que valora en el k-means un buen equilibrio entre calidad del agrupamiento y costo de cómputo.

Y, en cuanto al fondo de los perfiles, los resultados se cruzan de forma natural con la evidencia sobre persuasión adaptada al receptor. La asociación moderada que se observa ($p = 0,561$) cae dentro del rango que los experimentos internacionales atribuyen al microdirigismo político (Tappin et al., 2023; Hackenburg & Margetts, 2024); es decir, los retornos persuasivos existen, pero cambian según el marco institucional, como advierten también las lecturas éticas (Seybold, 2025). La aportación de este trabajo frente a esos antecedentes es doble: lleva la segmentación psicográfica al terreno del sufragio en un país plurinacional, en la línea de los usos regionales revisados (Zuloaga-Rotta et al., 2024), y convierte la estructura de datos en decisiones de gasto que pueden comprobarse con el simulador.

El balance de lo hecho es claro: el modelo de microsegmentación algorítmica basado en k-means satisface lo que el estudio se propuso. Ordena al electorado en segmentos pequeños, homogéneos por dentro y diferenciados entre sí; describe cada perfil con sus seis dimensiones psicográficas y su composición por edad; y convierte esa estructura en decisiones de campaña concretas. La fiabilidad de la encuesta (α global de 0,881), la nitidez de los centroides y el contraste de hipótesis ($H = 86,851$; $p < 0,001$) dan al modelo un respaldo empírico verificable y reproducible con el protocolo documentado.

En el horizonte regional, las aplicaciones del aprendizaje automático a elecciones latinoamericanas (Zuloaga-Rotta et al., 2024; Barredo-Ibáñez et al., 2024) muestran que la segmentación demográfica tradicional no agota la complejidad de los patrones políticos y digitales de la zona; los resultados de este trabajo confirman esa premisa en Ecuador, porque la mezcla de variables estructurales, territoriales y psicográficas permitió separar segmentos que una partición demográfica simple no distinguiría por sí misma. A diferencia de los ejercicios predictivos, el modelo propuesto no intenta anticipar resultados, sino describir la estructura segmentada del electorado y ofrecer insumos de aplicación práctica dentro de límites éticos estrictos.

Las limitaciones del estudio merecen consideración al interpretar los resultados. El k-means presume clústeres convexos y de tamaño más o menos parejo, un supuesto que quizá no calce del todo con la estructura real del electorado; la sensibilidad al punto de partida se mitiga con treinta inicializaciones independientes y métricas de estabilidad, aunque sin eliminar del todo el riesgo de óptimos locales. El diseño transversal impide fijar relaciones causales entre los perfiles obtenidos y los resultados electorales. El autorreporte en variables psicográficas puede arrastrar sesgo de deseabilidad social, y la modalidad mixta de aplicación puede introducir sesgos de cobertura y de modo, contrarrestados con protocolo único y comparación por modalidad.

El levantamiento empírico abarcó íntegro el territorio nacional, de modo que los resultados son generalizables al electorado continental e insular del Ecuador inscrito en el padrón de 2026, con un margen de error de $\pm 5\%$ derivado del tamaño muestral. Quedaron por fuera del alcance operativo las tres circunscripciones del exterior, cuyo padrón se excluyó deliberadamente del cálculo muestral, y los electores privados de libertad no empadronados. El estrato de voto facultativo (34 casos) conserva su peso dentro del registro nacional y se trató con afijación proporcional y ponderadores post estratificación. Por decisión metodológica, no se corrió una comparación empírica frente a otros algoritmos de clustering; esa contrastación, al igual que el cotejo de los perfiles contra los resultados oficiales agregados del proceso de noviembre de 2026, queda planteada como línea futura.

Las implicaciones aplicadas alcanzan de lleno al marketing político ecuatoriano: priorizar segmentos según tamaño relativo, certeza de voto declarada y receptividad persuasiva; delimitar los ejes temáticos del mensaje según las preocupaciones dominantes de cada grupo; y optimizar en modo multimodal canales y piezas de comunicación según los hábitos informativos del segmento. Estas ventanas de uso orientan la asignación eficiente de recursos, en sintonía con los hallazgos sobre los retornos de adaptar el mensaje al perfil receptor (Tappin et al., 2023; Guess et al., 2023). Sumar el voto facultativo agrega valor estratégico diferencial: conocer los perfiles de quienes pueden decidir libremente si votan permite no solo persuadir, sino activar la participación de grupos históricamente volátiles, dentro del marco del Código de la Democracia (Asamblea Nacional del Ecuador, 2009).

Conclusiones

Se materializó en un modelo de microsegmentación algorítmica basado en k-means que reconoció y caracterizó dos perfiles diferenciados de votantes del territorio ecuatoriano, dando cumplimiento al objetivo general. Para lograrlo se cimentó teóricamente el fenómeno (objetivo específico 1), se corrió el algoritmo sobre los datos estandarizados del instrumento (objetivo específico 2) y se valoró la calidad del modelo mediante índices de validación interna y contraste de hipótesis (objetivo específico 3); de este modo, las tres preguntas específicas quedan ligadas a las decisiones metodológicas adoptadas.

Respecto de las hipótesis, los datos respaldan la alterna (H1) en desmedro de la nula (H0): los segmentos identificados difieren de forma estadísticamente significativa en la intención de voto declarada y en la predisposición a participar (Kruskal-Wallis $H = 86,851$; $p < 0,001$; $\epsilon^2 = 0,224$), lo que deja en evidencia perfiles electorales diferenciados e idóneos para orientar una campaña de captación de votos. El grado de asociación, positivo y moderado, entre el modelo algorítmico de segmentación y la captación del voto ($p = 0,561$; $p < 0,001$) refuerza ese respaldo empírico.

El modelo k-means resultante se condensa en dos perfiles de lectura directa. El pasivo conservador (S1, 59,7 %), dominante entre los 23 y 40 años, es el más grande y el más reacio a la persuasión: pide mensajes de movilización y de confianza institucional. El receptivo digital (S2, 40,3 %), más frecuente en la juventud temprana, en los mayores de 41 y en los territorios insulares y costeros, ostenta la mayor certeza de voto, participación, apertura persuasiva, y se vuelve, en consecuencia, el público central del diseño diferenciado de campañas; las seis dimensiones psicográficas sostienen esta caracterización que permiten convertir cada perfil en ejes temáticos, canales y piezas comunicacionales específicos.

El valor distintivo de la investigación estriba en traer al Ecuador un modelo de segmentación sustentado en aprendizaje automático no supervisado, respaldado por índices de calidad interna y apoyado en cifras oficiales del padrón electoral 2026, en un escenario donde el estado del arte es sobre todo descriptivo. Sus aplicaciones potenciales alcanzan a actores políticos, consultoras electorales y academia, que ahora cuenta con un protocolo reproducible, instrumento validado y código documentado.

Como proyección, se propone comparar el modelo con métodos alternativos de clustering, cotejar los perfiles con los resultados oficiales agregados del proceso de noviembre de 2026 y recalibrar los perfiles en comicios futuros, habida cuenta de la volatilidad del sistema de partidos ecuatoriano. Como limitaciones se reiteran el alcance nacional del levantamiento sin circunscripciones del exterior, el carácter transversal del diseño y los supuestos geométricos del k-means, aspectos ya ponderados en la interpretación de los resultados.

Referencias

- Aagaard, P., & Marthedal, R. (2023). Political microtargeting: Towards a pragmatic approach. *Internet Policy Review*, 12(1). <https://doi.org/10.14763/2023.1.1690>
- Ahmed, M., Seraj, R., & Islam, S. M. S. (2020). The k-means algorithm: A comprehensive survey and performance evaluation. *Electronics*, 9(8), 1295. <https://doi.org/10.3390/electronics9081295>
- Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, 45(1), 131–142. <https://doi.org/10.1177/0013164485451012>
- Álvarez Mena, P. A., Mantilla Bolaños, E. D., Marín Nicolalde, J. E., & Trujillo Sarzosa, I. E. (2025). Segmentación de clientes de una empresa ecuatoriana de pulpa de fruta usando K-Means [Tesis de maestría, Universidad Internacional del Ecuador]. Repositorio Digital UIDE. <https://repositorio.uide.edu.ec/handle/37000/8548>
- Asamblea Nacional Constituyente. (2008). Constitución de la República del Ecuador. Registro Oficial 449, 20 de octubre de 2008. https://oig.cepal.org/sites/default/files/2008_cpr_ecu.pdf
- Asamblea Nacional del Ecuador. (2009, 27 de abril). Código de la Democracia (Ley Orgánica Electoral y de Organizaciones Políticas de la República del Ecuador). Registro Oficial Suplemento 578, 27 de abril de 2009. <https://www.cne.gob.ec/codigo-de-la-democracia/>
- Banda Garza, F. (2024). Geografía electoral y modelos de comportamiento electoral: un análisis espacial del voto y su relación con las variables sociodemográficas en Nuevo León [Tesis doctoral, Universidad Autónoma de Nuevo León]. LACOP UANL.
- Barredo-Ibáñez, D., López-López, P. C., & de la Garza Montemayor, D. J. (2024). Social media and support for democracy in 18 Latin American countries: A comparative analysis of moderated mediation (2018–2020). *Global Media and Communication*, 20(1), 111–129. <https://doi.org/10.1177/17427665231212144>
- Consejo Nacional Electoral. (2026, 2 de julio). Resolución PLE-CNE-2-2-7-2026: Registro Electoral definitivo para las Elecciones Seccionales y del CPCCS 2026. <https://www.cne.gob.ec>

- Creswell, J. W., & Creswell, J. D. (2023). *Research design: Qualitative, quantitative, and mixed methods approaches* (6.^a ed.). SAGE.
- Dobber, T., Metoui, N., Trilling, D., Helberger, N., & de Vreese, C. H. (2021). Do (microtargeted) deepfakes have real effects on political attitudes? *The International Journal of Press/Politics*, 26(1), 69–91. <https://doi.org/10.1177/1940161220944364>
- Escobar-Pérez, J., & Cuervo-Martínez, Á. (2008). Validez de contenido y juicio de expertos: Una aproximación a su utilización. *Avances en Medición*, 6(1), 27–36.
- Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. En *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining* (pp. 226–231). AAAI Press.
- Freidenberg, F., & Pachano, S. (2016). El sistema político ecuatoriano. FLACSO Ecuador.
- George, D., & Mallery, P. (2021). *IBM SPSS Statistics 27 step by step: A simple guide and reference* (17.^a ed.). Routledge.
- Granda Gómez, J. M., Lema Auz, B. G., Llerena Mena, A. F., & Villarruel Cerón, W. R. (2025). Modelo híbrido de segmentación y predicción de flujo vehicular en la avenida “Las Aguas” en Guayaquil – Ecuador, utilizando K-Means y Random Forest [Tesis de maestría, Universidad Internacional del Ecuador]. Repositorio Digital UIDE. <https://repositorio.uide.edu.ec/handle/37000/8543>
- Gray, R. M. (1984). Vector quantization. *IEEE ASSP Magazine*, 1(2), 4–29. <https://doi.org/10.1109/MASSP.1984.1162229>
- Guess, A. M., Malhotra, N., Pan, J., Barberá, P., Allcott, H., Brown, T., Crespo-Tenorio, A., Dimmery, D., Freelon, D., Gentzkow, M., González-Bailón, S., Kennedy, E., Kim, Y. M., Lazer, D., Moehler, D., Nyhan, B., Rivera, C. V., Settle, J., Thomas, D. R., ... Tucker, J. A. (2023). How do social media feed algorithms affect attitudes and behavior in an election campaign? *Science*, 381(6656), 398–404. <https://doi.org/10.1126/science.abp9364>
- Hackenburg, K., & Margetts, H. (2024). Evaluating the persuasive influence of political microtargeting with large language models. *Proceedings of the National Academy of Sciences*, 121(24), e2403116121. <https://doi.org/10.1073/pnas.2403116121>
- Hernández-Sampieri, R., & Mendoza Torres, C. P. (2018). *Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta*. McGraw-Hill.
- Hernández Reyna, G. D. (2023). El comportamiento e influencia electoral determinante de los factores que influyen en la decisión del voto de los ciudadanos durante el periodo electoral del año 2021 al puesto de presidente municipal de Monterrey [Tesis doctoral, Universidad Autónoma de Nuevo León]. Repositorio Institucional UANL. <http://eprints.uanl.mx/26133/>
- Hidalgo Chica, A. E., & Cedeño Moreira, C. L. (2022). Comunicación política en redes sociales durante la segunda vuelta electoral de Ecuador, año 2021: Análisis del uso de la red social Facebook. *ReHuSo: Revista de Ciencias Humanísticas y Sociales*, 7(1), 104–115. <https://doi.org/10.5281/zenodo.5825838>

- Horan, K., Brunsdon, C., & Domijan, K. (2024). A multilevel spatial model to investigate voting behaviour in the 2019 UK general election. *Applied Spatial Analysis and Policy*, 17(2), 703–727. <https://doi.org/10.1007/s12061-023-09563-6>
- Ienca, M. (2023). On artificial intelligence and manipulation. *Topoi*, 42(3), 833–842. <https://doi.org/10.1007/s11245-023-09940-3>
- Imik Tanyildizi, N., & Tanyildizi, H. (2022). Estimation of voting behavior in election using support vector machine, extreme learning machine and deep learning. *Neural Computing and Applications*, 34(20), 17329–17342. <https://doi.org/10.1007/s00521-022-07395-y>
- Instituto Nacional de Estadística y Censos. (s.f.). Estadísticas e indicadores [Base de datos]. <https://www.ecuadorencifras.gob.ec>
- Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., & Heming, J. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, 622, 178–210. <https://doi.org/10.1016/j.ins.2022.11.139>
- Kotler, P., & Keller, K. L. (2016). Dirección de marketing (15.ª ed.). Pearson Educación.
- Kruikemeier, S., Vermeer, S., Metoui, N., Dobber, T., & Zarouali, B. (2022). (Tar)getting you: The use of online political targeted messages on Facebook. *Big Data & Society*, 9(2). <https://doi.org/10.1177/20539517221089626>
- Kumar, S., Rani, R., Pippal, S. K., & Agrawal, R. (2025). Customer segmentation in e-commerce: K-means vs hierarchical clustering. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, 23(1), 119–128. <https://doi.org/10.12928/telkomnika.v23i1.26384>
- Ley Orgánica de Protección de Datos Personales. (2021, 26 de mayo). Registro Oficial Suplemento n.º 459. <https://www.gob.ec/regulaciones/ley-organica-proteccion-datos-personales-0>
- Lloyd, S. P. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2), 129–137. <https://doi.org/10.1109/TIT.1982.1056489>
- López-López, P. C., Oñate, P., & Rocha, Á. (2020). Social media mining, debate and feelings: Digital public opinion's reaction in five presidential elections in Latin America. *Cluster Computing*, 23(3), 1875–1886. <https://doi.org/10.1007/s10586-020-03072-8>
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. En *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 281–297). University of California Press.
- Matthes, J., Hirsch, M., Stubenvoll, M., Binder, A., Kruikemeier, S., Lecheler, S., & Otto, L. (2022). Understanding the democratic role of perceived online political micro-targeting: Longitudinal effects on trust in democracy and political interest. *Journal of Information Technology & Politics*, 19(4), 435–448. <https://doi.org/10.1080/19331681.2021.2016542>

- Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the National Academy of Sciences*, 114(48), 12714–12719. <https://doi.org/10.1073/pnas.1710966114>
- Minihold, S., & Votta, F. (2024). Accepting exclusion: Examining the (un)intended consequences of data-driven campaigns. *Media and Communication*, 12, Article 8685. <https://doi.org/10.17645/mac.8685>
- Miraftabzadeh, S. M., Colombo, C. G., Longo, M., & Foadelli, F. (2023). K-means and alternative clustering methods in modern power systems. *IEEE Access*, 11, 119596–119633. <https://doi.org/10.1109/ACCESS.2023.3327640>
- Moreno-Morán, D. X. (2025). La comunicación política en redes sociales y su impacto en la campaña electoral de los candidatos a la presidencia de la república del Ecuador. *Ciencia Latina: Revista Científica Multidisciplinaria*, 8(6), 6747–6758. https://doi.org/10.37811/cl_rcm.v8i6.15361
- Nenadić, I. (2019). Unpacking the “European approach” to tackling challenges of disinformation and political manipulation. *Internet Policy Review*, 8(4). <https://doi.org/10.14763/2019.4.1436>
- Otzen, T., & Manterola, C. (2017). Técnicas de muestreo sobre una población a estudio. *International Journal of Morphology*, 35(1), 227–232. <https://doi.org/10.4067/S0717-95022017000100037>
- Perafan-Lopez, J. C., Ferrer-Gregory, V. L., Nieto-Londoño, C., & Sierra-Pérez, J. (2022). Performance analysis and architecture of a clustering hybrid algorithm called FA+GA-DBSCAN using artificial datasets. *Entropy*, 24(7), 875. <https://doi.org/10.3390/e24070875>
- Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health*, 30(4), 459–467. <https://doi.org/10.1002/nur.20199>
- Seybold, S. L. (2025). Voter deterrence campaigns and the moral-epistemic landscape of political microtargeting. *Ethics and Information Technology*, 27(2), Artículo 22. <https://doi.org/10.1007/s10676-025-09831-8>
- Simchon, A., Edwards, M., & Lewandowsky, S. (2024). The persuasive effects of political microtargeting in the age of generative artificial intelligence. *PNAS Nexus*, 3(2), pgae035. <https://doi.org/10.1093/pnasnexus/pgae035>
- Sinaga, K. P., & Yang, M.-S. (2020). Unsupervised k-means clustering algorithm. *IEEE Access*, 8, 80716–80727. <https://doi.org/10.1109/ACCESS.2020.2988796>
- Susser, D., Roessler, B., & Nissenbaum, H. (2019). Online manipulation: Hidden influences in a digital world. *Georgetown Law Technology Review*, 4(1), 1–45.
- Tabianan, K., Velu, S., & Ravi, T. (2022). K-means clustering approach for intelligent customer segmentation using customer purchase behavior data. *Sustainability*, 14(12), 7243. <https://doi.org/10.3390/su14127243>

- Tateishi, E. (2023). The spatiotemporal socio-demography of the Tokyo capital region: A data-driven explorative approach. *Review of Regional Research*, 43(3), 467–519. <https://doi.org/10.1007/s10037-023-00198-1>
- Tappin, B. M., Wittenberg, C., Hewitt, L. B., Berinsky, A. J., & Rand, D. G. (2023). Quantifying the potential persuasive returns to political microtargeting. *Proceedings of the National Academy of Sciences*, 120(25), e2216261120. <https://doi.org/10.1073/pnas.2216261120>
- Tokuda, E. K., Comin, C. H., & Costa, L. da F. (2022). Revisiting agglomerative clustering. *Physica A: Statistical Mechanics and its Applications*, 585, 126433. <https://doi.org/10.1016/j.physa.2021.126433>
- Ward, J. H., Jr. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301), 236–244. <https://doi.org/10.1080/01621459.1963.10500845>
- Waggoner, P. D. (2021). Unsupervised machine learning for clustering in political and social research. Cambridge University Press. <https://www.cambridge.org/core/elements/unsupervised-machine-learning-for-clustering-in-political-and-social-research/BF62D1E8F6DB3237D5CE524FBFCBA33A>
- Zarouali, B., Dobber, T., De Pauw, G., & de Vreese, C. (2022). Using a personality-profiling algorithm to investigate political microtargeting: Assessing the persuasion effects of personality-tailored ads on social media. *Communication Research*, 49(8), 1066–1091. <https://doi.org/10.1177/0093650220961965>
- Zuloaga-Rotta, L., Borja-Rosales, R., Rodríguez Mallma, M. J., Mauricio, D., & Maculan, N. (2024). Method to forecast the presidential election results based on simulation and machine learning. *Computation*, 12(3), 38. <https://doi.org/10.3390/computation12030038>
- Zuiderveen Borgesius, F. J., Möller, J., Kruikemeier, S., Ó Fathaigh, R., Irion, K., Dobber, T., Bodo, B., & de Vreese, C. H. (2018). Online political microtargeting: Promises and threats for democracy. *Utrecht Law Review*, 14(1), 82–96. <https://doi.org/10.18352/ulr.420>