

Big data y elecciones

Cómo se usan los datos para segmentar votantes

Carlos Merlo Pastrana

Universidad Anáhuac México

Victory, Ciudad de México · dr@carlosmerlo.com

ORCID: <https://orcid.org/0009-0008-7804-9775>

DOI: <https://doi.org/10.5281/zenodo.23789103>

Versión preliminar sin revisión por pares. Se distribuye para discusión académica. Licencia CC BY 4.0.

Cómo citar: Merlo Pastrana, C. (2026). Big data y elecciones: Cómo se usan los datos para segmentar votantes

(Documento de trabajo n.º 3). Victory. <https://doi.org/10.5281/zenodo.23789103>

Resumen

Las campañas políticas usan cada vez más datos para decidir a quién hablar, con qué mensaje y por qué canal. Este documento explica cómo funciona la segmentación de votantes basada en datos, qué tan eficaz es según la evidencia experimental y qué riesgos plantea. La literatura muestra que las campañas construyen bases de datos de ciudadanos y modelos que estiman tres probabilidades: que una persona vote, que apoye a un candidato y que pueda ser persuadida. La evidencia sobre persuasión es más modesta que la promesa comercial: 59 experimentos durante la campaña presidencial de 2016 en Estados Unidos encontraron efectos muy pequeños de la publicidad política, y la microsegmentación aporta ventajas limitadas a uno o dos atributos. Los riesgos documentados incluyen la falta de transparencia, el uso indebido de datos personales, como en el caso Cambridge Analytica, y en México la exposición del padrón electoral en 2016. Se cierra con implicaciones para campañas y para la regulación.

Palabras clave: big data, ciencia de datos electoral, microsegmentación, campañas políticas, persuasión, privacidad, protección de datos, México.

Abstract

Political campaigns increasingly use data to decide whom to contact, with which message and through which channel. This working paper explains how data-driven voter targeting works, how effective it is according to experimental evidence and what risks it poses. Campaigns build citizen databases and models that estimate turnout, support and persuadability. Experimental evidence on persuasion is more modest than commercial claims: 59 experiments during the 2016 US presidential campaign found very small advertising effects, and

microtargeting adds limited gains beyond one or two attributes. Documented risks include lack of transparency, misuse of personal data, as in the Cambridge Analytica case, and in Mexico the 2016 exposure of the voter roll. It closes with implications for campaigns and regulation.

Keywords: big data, electoral data science, microtargeting, political campaigns, persuasion, privacy, data protection, Mexico.

1. Introducción

Durante décadas, las campañas decidieron con encuestas generales, experiencia territorial e intuición. Desde mediados de la década de 2000, eso cambió: las campañas empezaron a tratar a cada votante como un registro en una base de datos, con probabilidades estimadas de comportamiento (Nickerson y Rogers, 2014). Issenberg (2012) documentó cómo los experimentos de campo y las ciencias del comportamiento entraron a las campañas estadounidenses; Kreiss (2016) mostró, con datos de 629 colaboradores de campañas presidenciales entre 2004 y 2012, cómo esa infraestructura tecnológica se construyó y se difundió.

El modelo se exportó a otras democracias mediante consultores y plataformas digitales (Bennett y Lyon, 2019). En América Latina, el término *big data* electoral se usa a menudo como sinónimo de eficacia garantizada. Este documento ordena lo que se sabe: cómo funciona la segmentación, cuánto funciona realmente y qué riesgos implica, con atención al caso mexicano.

2. Cómo funciona la segmentación

Una campaña basada en datos convierte registros en probabilidades y probabilidades en prioridades de contacto.

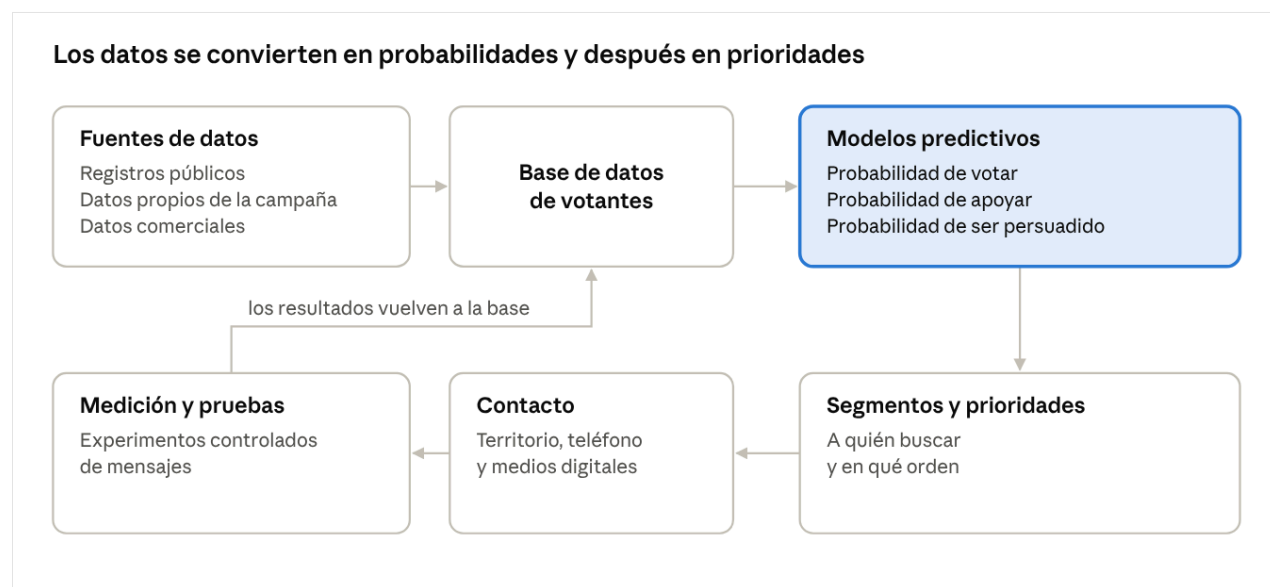


Figura 1. Flujo de datos en una campaña: de las fuentes al contacto y de vuelta a la base. Elaboración propia con base en Nickerson y Rogers (2014).

El punto de partida son tres tipos de fuentes. Los registros públicos son la base: Hersh (2015) muestra que las campañas conocen a los votantes sobre todo por un conjunto reducido de datos en manos del Estado, y que cada campaña percibe un electorado distinto según qué datos tenga disponibles. A eso se suman los datos propios de la campaña y las bases comerciales.

Con esa base, las campañas construyen modelos que estiman tres probabilidades por persona: que vote, que apoye al candidato o la causa, y que pueda ser persuadida (Nickerson y Rogers, 2014). Según estos autores, los datos más útiles son los que los propios ciudadanos generan al interactuar con la campaña, más que los comerciales. Las puntuaciones definen segmentos y el orden de contacto; los resultados del contacto y de las pruebas de mensajes vuelven a alimentar la base.

3. Cuánto funciona: la evidencia sobre persuasión y microsegmentación

La promesa comercial del big data electoral es mayor que la evidencia. La prueba más amplia disponible son 59 experimentos aleatorios realizados en tiempo real durante la campaña presidencial de 2016 en Estados Unidos, con 49 anuncios y unas 34,000 personas (Coppock et al., 2020). El efecto promedio sobre la intención de voto fue de 0.7 puntos porcentuales, sin significancia estadística, y los efectos fueron pequeños en todos los subgrupos, incluidos los estados competidos.

La microsegmentación mejora ese resultado, pero poco. Tappin et al. (2023), con más de 23,000 participantes, encontraron que segmentar por un solo atributo, como la afiliación partidista, hizo los mensajes alrededor de 70% más eficaces que mostrar a todos el mejor anuncio único. Agregar más atributos no produjo ganancias adicionales, y el beneficio varió según el tema.

La segmentación psicológica tiene evidencia más fuerte, pero en otro terreno. Matz et al. (2017), en experimentos que alcanzaron a más de 3.5 millones de personas, encontraron que los anuncios ajustados a la personalidad lograron hasta 40% más clics y hasta 50% más compras. Se trató de publicidad comercial, no política, por lo que su extrapolación a la decisión de voto debe tomarse con cautela.

La síntesis es consistente con Bennett y Lyon (2019): el efecto más claro de las campañas basadas en datos probablemente está en movilizar a simpatizantes y donantes, no en convertir a indecisos. El dato sirve más para decidir a quién buscar que para cambiar lo que piensa.

4. Riesgos: Cambridge Analytica, privacidad y el caso mexicano

Opacidad. Zuiderveen Borgesius et al. (2018) señalan que la microsegmentación permite a un partido mostrar caras distintas a públicos distintos sin que nadie vea el conjunto, y que la recolección de datos plantea problemas serios de privacidad. En el Reino Unido, la autoridad de protección de datos encontró que los tres partidos principales usaban software que estimaba la etnia y la edad de las personas a partir de su nombre, y que subían datos de contacto de votantes a Facebook sin transparencia suficiente (Information Commissioner's Office, 2018).

Cambridge Analytica. La misma autoridad concluyó en 2020 que la empresa combinó datos comerciales con los obtenidos por medio de una aplicación de Facebook. También concluyó que su afirmación de tener 5,000 datos sobre 230 millones de adultos estadounidenses probablemente era exagerada, y que sus modelos usaban herramientas comunes de código abierto, con una precisión real probablemente mucho menor que en sus datos de entrenamiento (Denham, 2020). El caso muestra un doble riesgo: el abuso de datos personales y la venta de una capacidad de influencia que no se ha demostrado.

El caso mexicano. En abril de 2016 se descubrió en un servidor de Amazon, sin contraseña, una copia de la lista nominal con 93.4 millones de registros: nombres, domicilios, fechas de nacimiento y claves de elector (ESET, 2016). Movimiento Ciudadano reconoció que había subido esa copia (Expansión, 2016). El INE impuso al partido una multa de 34.1 millones de pesos, que el Tribunal Electoral confirmó por unanimidad (Vanguardia, 2018).

El marco legal también cambió. La nueva Ley Federal de Protección de Datos Personales en Posesión de los Particulares se publicó en el *Diario Oficial de la Federación* el 20 de marzo de 2025. Abrogó la ley de 2010 y trasladó las funciones del INAI a la Secretaría Anticorrupción y Buen Gobierno. Cualquier estrategia de datos electorales en México debe revisarse contra ese marco (Ley Federal de Protección de Datos Personales en Posesión de los Particulares, 2025).

5. Implicaciones para campañas y para la regulación

Para las campañas:

1. **Invertir primero en datos propios.** Lo que los ciudadanos hacen y dicen ante la campaña (registros, voluntariado, respuestas, asistencia) predice mejor que las bases comerciales (Nickerson y Rogers, 2014).
2. **Usar los modelos para movilizar y priorizar, no para prometer persuasión.** La evidencia de efectos pequeños (Coppock et al., 2020) aconseja asignar el presupuesto según la probabilidad de participar y de apoyar.
3. **Segmentar poco y probar mucho.** La ganancia de la microsegmentación se agota con uno o dos atributos (Tappin et al., 2023). Es más rentable probar mensajes con experimentos controlados que multiplicar segmentos.
4. **Tratar la privacidad como riesgo reputacional.** El caso de la lista nominal muestra que una filtración puede costar más que cualquier ganancia de segmentación.

Para la regulación:

1. Transparencia en los anuncios segmentados: quién paga, a quién se dirige y con qué criterios.
2. Reglas claras sobre el uso de la lista nominal y su resguardo por parte de partidos y consultores.
3. Capacidad técnica de la autoridad para auditar modelos y bases de datos de campaña, que en México debe definirse tras el nuevo reparto de funciones de 2025.

6. Conclusiones

El big data cambió la forma de organizar las campañas: hoy el votante es un registro con probabilidades estimadas de participar, apoyar y ser persuadido. Pero la evidencia experimental desmiente la idea de que los datos permitan manipular a voluntad. Los efectos de persuasión son pequeños y la microsegmentación aporta poco más allá de uno o dos atributos.

El valor real del dato electoral está en la eficiencia: saber a quién buscar, en qué orden y con qué recursos. Sus riesgos, en cambio, son concretos y están documentados: opacidad, abuso de datos personales y filtraciones como la de la lista nominal en 2016.

Para México, la agenda pendiente incluye medir con experimentos locales la eficacia de la segmentación y adaptar la práctica de los partidos y consultores al nuevo marco de protección de datos de 2025.

Referencias

- Bennett, C. J. y Lyon, D. (2019). Data-driven elections: Implications and challenges for democratic societies. *Internet Policy Review*, 8(4). <https://doi.org/10.14763/2019.4.1433>
- Coppock, A., Hill, S. J. y Vavreck, L. (2020). The small effects of political advertising are small regardless of context, message, sender, or receiver: Evidence from 59 real-time randomized experiments. *Science Advances*, 6(36), eabc4046. <https://doi.org/10.1126/sciadv.abc4046>
- Denham, E. (2020, 2 de octubre). *Letter to Julian Knight MP, Chair of the DCMS Select Committee* (ICO/O/ED/L/RTL/0181). Information Commissioner's Office. https://cy.ico.org.uk/media2/migrated/2618383/20201002_ico-o-ed-l-rtl-0181_to-julian-knight-mp.pdf
- ESET. (2016, 25 de abril). *Información de 93 millones de mexicanos expuesta*. WeLiveSecurity. <https://www.welivesecurity.com/la-es/2016/04/25/informacion-93-millones-mexicanos-expuesta/>
- Expansión. (2016, 27 de abril). *Movimiento Ciudadano admite que subieron el padrón electoral a Amazon*. <https://politica.expansion.mx/politica/2016/04/27/movimiento-ciudadano-admite-que-subieron-el-padron-electoral-a-amazon>
- Hersh, E. D. (2015). *Hacking the electorate: How campaigns perceive voters*. Cambridge University Press. <https://doi.org/10.1017/CBO9781316212783>
- Information Commissioner's Office. (2018). *Democracy disrupted? Personal information and political influence*. <https://ico.org.uk/media2/migrated/2259369/democracy-disrupted-110718.pdf>
- Issenberg, S. (2012). *The victory lab: The secret science of winning campaigns*. Crown.
- Kreiss, D. (2016). *Prototype politics: Technology-intensive campaigning and the data of democracy*. Oxford University Press.
- Ley Federal de Protección de Datos Personales en Posesión de los Particulares. *Diario Oficial de la Federación*, 20 de marzo de 2025 (México).
- Matz, S. C., Kosinski, M., Nave, G. y 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>
- Merlo, C. (2021). *Data science electoral*. <https://www.amazon.com.mx/dp/B0FTXWT4HW>

- Nickerson, D. W. y Rogers, T. (2014). Political campaigns and big data. *Journal of Economic Perspectives*, 28(2), 51–74. <https://doi.org/10.1257/jep.28.2.51>
- Tappin, B. M., Wittenberg, C., Hewitt, L. B., Berinsky, A. J. y 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>
- Vanguardia. (2018, 20 de septiembre). *TEPJF confirma multa a Movimiento Ciudadano con 34 mdp por negligencia con padrón*. <https://vanguardia.com.mx/articulo/tepfj-confirma-multa-movimiento-ciudadano-con-34-mdp-por-negligencia-con-padron>
- Zuiderveen Borgesius, F. J., Möller, J., Kruikemeier, S., Ó Fathaigh, R., Irion, K., Dobber, T., Bodo, B. y de Vreese, C. (2018). Online political microtargeting: Promises and threats for democracy. *Utrecht Law Review*, 14(1), 82–96. <https://doi.org/10.18352/ulr.420>

Obra relacionada

Este documento amplía temas tratados en *Data Science Electoral* (Merlo, 2021), disponible en Amazon México: <https://www.amazon.com.mx/dp/B0FTXWT4HW>