

# El 'big data' no gana unas elecciones

[Gonzalo Toca](#)



***En los últimos años los datos masivos han pasado de una rareza a convertirse en una herramienta clave, pero insuficiente para ganar elecciones.***

La puesta de largo de la aplicación electoral de los datos masivos se produjo, seguramente, en [la campaña presidencial estadounidense de 2012](#), cuando Barack Obama utilizó las nuevas capacidades de almacenamiento y análisis que le proporcionó Narwhal, un programa que integró y cruzó la información que se obtenía mediante plataformas digitales y aquella que se ordeñaba gota a gota gracias a las llamadas telefónicas, las identidades y perfiles que ofrecían voluntariamente los simpatizantes que participaban en las campañas, las encuestas, etcétera.

Aunque muchos medios, como de costumbre, sacaron las trompetas para anunciar que el triunfo de Obama se debía a los datos masivos y que esto significaba el inicio de una nueva era, ambas conclusiones [se demostraron falsas](#). Para empezar, Barack Obama ganó gracias a sus méritos, a la inercia de quien ha ocupado la presidencia durante los cuatro años anteriores y a que sus rivales estaban [profundamente divididos](#) en torno a la figura de Mitt Romney, un republicano tibio y mormón en exceso para el gusto del cristianísimo y beligerante Tea Party. Los candidatos, sus mensajes y el contexto social ganan o hacen ganar las elecciones; los datos masivos, hasta la fecha, no.

Quienes pensaron que Obama había triunfado gracias al *big data* estaban lógicamente convencidos de que estas nuevas tecnologías iban a predecir con muy poco margen de error los resultados de los comicios británicos de mayo de 2015. Los analistas, con toneladas de información novedosa (a los clásicos sondeos de YouGov se sumaron los esfuerzos de Tata Consultancy Services entre otras) no tuvieron problema en afirmar que ganarían tímidamente los conservadores. Lo confirmaban los posos del té de las redes sociales.

Pues bien, [todos los sondeos](#) se equivocaron y también las aplicaciones móviles que rastreaban Facebook y Twitter. Los *torios* de David Cameron arrasaron y nadie lo vio venir. Por supuesto, las empresas de *big data* echaron la culpa a [la calidad de los datos](#), a que por desgracia éstos no eran tan dinámicos como los cambios de opinión y los reflejaban tarde, a las groseras extrapolaciones de los viejos sondeos o a [la discrepancia](#) entre las emociones expresadas en Internet y la verdadera decisión de los votantes frente a la urna. Acababan de descubrir que muchas acciones humanas eran imprevisibles, inconfesables y caprichosas... Y que las muestras de las encuestas, a veces, no eran representativas. Bravo.

De todos modos y a pesar del fiasco de las elecciones británicas, nadie puede decir que los datos masivos sean inútiles o que sus notables imprecisiones no sean un camino interesante para comprender mejor a los votantes. Volvamos a Barack Obama: él [amplió su ventaja frente a los republicanos](#) y tomó mejores decisiones que ellos durante la campaña gracias a la revolución del *big data* que se había producido en los años anteriores y que había consistido en la multiplicación de la capacidad de procesamiento de equipos y servidores virtuales, en el hundimiento del coste del almacenamiento de los datos, en la aparición de los expertos que eran capaces de interpretar y construir visualizaciones de datos, y en el cambio de mentalidad de los políticos y consultores.

Estos nuevos instrumentos le ayudaron, [como recuerdan](#) los profesores David Nickerson y Todd Rogers, a que su propaganda alcanzase a los posibles simpatizantes en un territorio donde mayoritariamente no le votaban y a que, tanto en los lugares donde tenía mayoría como

en los que no, sus mensajes no movilizaran en exceso a sus enemigos. Por todo ello, en un contexto profundamente polarizado como el de las elecciones presidenciales estadounidenses de 2012, los datos masivos alcanzaron un protagonismo muy especial.

## Privacidad a la americana

No es casualidad que la puesta de largo del *big data* se produjera en Estados Unidos. Al fin y al cabo, es allí donde la regulación para obtener información privada resulta especialmente laxa, al menos a ojos de otros países occidentales y, ciertamente, para los miembros de la Unión Europea.

Debemos recordar que en la primera potencia mundial las empresas [no están obligadas](#) a avisar o pedir consentimiento para almacenar muchos tipos de datos personales, y que les basta normalmente con ofrecer a los afectados la posibilidad de darlos de baja en sus ficheros (las normas son más restrictivas en algunos ámbitos como el sanitario). A diferencia de eso, en países como España los particulares deben decir claramente que sí quieren compartir esa información y tienen derecho a consultarla, rectificarla, cancelarla y a oponerse a que alguien la almacene.

Estas cuestiones son importantes, porque implican que el *big data* político surgió en el lugar donde era más fácil y barato acceder a los datos personales sin el consentimiento de sus titulares. Desde entonces, la revolución se ha extendido rápidamente gracias al cambio tecnológico y a que las multinacionales estadounidenses que no hayan abierto, por ejemplo, una sede en Francia o en España [pueden recopilar los datos personales](#) de los franceses y de los españoles sin cumplir los límites que imponen las leyes nacionales. Los líderes políticos europeos están utilizando ya en sus campañas la misma información que prohibieron recopilar cuando estaban en el Gobierno... Y es legal porque se la compran a multinacionales sin domicilio fiscal en su país.

No hay que confundir esas empresas –llamadas *data brokers*– que extraen los datos y hacen un primer barrido de análisis antes de comercializarlos (como TLO, Neustar o Datalogix) con las enormes compañías que los almacenan y peinan hasta convertirlos en información para sus clientes. Aquí es donde se encuentran no solo los gigantes informáticos de toda la vida, como IBM o Dell, sino también consultoras como Deloitte, Accenture o PwC o nuevos e influyentes jugadores como Palantir Technologies, Pivotal y Splunk.

¿Pero cuáles son los datos que sirven los *data brokers* a los políticos en bandeja de plata? Por lo general, estamos hablando de números de teléfono, direcciones, régimen de propiedad o alquiler de la vivienda habitual, años de formación reglada o estado de sus hipotecas. Estos

datos no aportan nada si no se cruzan después con los que los simpatizantes de los partidos ofrecen voluntariamente y con los que pueden obtenerse de las estadísticas oficiales, que suelen recoger cifras macro –y anónimas– sobre la participación política en distintas regiones, los ingresos medios de los hogares o el número de hijos.

### **Datos útiles e indicadores**

No todos los datos que se compran son igual de efectivos. De hecho, su utilidad dependerá sobre todo de que la información privada provenga de un número de individuos que represente al colectivo de votantes que se pretende estudiar. Para ello, los individuos deberán ser muchos y esto quiere decir que, por ejemplo, la información sobre consumidores de productos relativamente minoritarios (como los suscriptores de prensa, por ejemplo) no interesan a los candidatos que intentan ganar unas elecciones.

Los datos relevantes, tanto los públicos como los privados, se emplearán después para construir tres tipos de indicadores: los que afectan a la conducta, a los apoyos y a la respuesta de los electores. Los que afectan a la conducta utilizan comportamientos previos e información demográfica para [identificar el tipo de participación](#) que va a tener un colectivo en la campaña. Básicamente, si sus miembros estarían interesados en votar, donar, asistir a los mítines o en ser voluntarios.

Los que se refieren a los apoyos expresan sobre todo las conclusiones de las encuestas que se realicen para valorar la puntuación que les dan a los candidatos y a los temas de su interés los posibles votantes. El objetivo en este caso es saber qué les interesa y quién les genera más confianza.

Por último, tenemos los indicadores que intentan adivinar la respuesta de los electores a los distintos estímulos y promesas de los candidatos. Aquí lo que se busca es procesar millones de datos dispares y tratar de evaluar tres cuestiones distintas: cuántos responderán positivamente a determinados mensajes, cuántos se volcarán de forma muy activa en las donaciones y actividades de voluntariado y a cuántos no merece la pena dirigirse porque o no se movilizarán o lo harán en nuestra contra.

A pesar de la asombrosa complejidad de los datos masivos y de la expectación que suscitó el éxito de la campaña de Barack Obama en 2012, debemos evitar las exageraciones. Tenemos que recordar, para empezar, que la victoria de Obama se habría producido [de todas formas](#) (aunque hubiera sido más ajustada en algunos estados) y que sus técnicos utilizaron muchos menos datos de los que se cree popularmente: fueron solo diez *terabytes*, es decir, el equivalente a la capacidad de almacenamiento de aproximadamente diez ordenadores

---

portátiles para procesar información relevante en un país de 300 millones de personas.

También debe recordarse que el acierto de una predicción depende menos de la potencia de la máquina y de la cantidad masiva de los datos que de su calidad y del talento del científico a la hora de construir variables, buscar correlaciones e identificar tendencias. Por último, no podemos olvidar que estamos sacrificando nuestra privacidad mientras los gobernantes que deberían protegerla se benefician de su violación para ganar las elecciones.

**Fecha de creación**

17 diciembre, 2015