

Macrodatos y mentiras masivas

Posted on 6 de mayo de 2019 by Sergio Parra

Nuestro cerebro es la caja negra de un avión accidentado en la zona más profunda de un océano freudiano. Incluso para nosotros mismos. O, al menos, lo era hasta ahora. Porque los macrodatos (el *big data*, en su traducción anglosajona más sofisticada) constituyen una visión sin precedentes de la psique humana que ya habría querido para sí un psicoanalista vienés. El poder de los datos masivos puede asistirnos en cada vez más tareas pero, sobre todo, puede revelar qué pasa por nuestra cabeza, incluso si lo hace de forma inconsciente para nosotros mismos. Y en ese último punto es en el que se centra Seth Stephens-Davidowitz, excientífico de datos de Google y profesor de The Wharton School, en el libro objeto de esta reseña.

Hasta ahora habían aparecido otros títulos que nos desgranaban las bondades de los macrodatos, como *Big Data*. *La revolución de los datos masivos*, de Viktor Mayer-Schönberger y Kenneth Cukier. Pero si éste se centraba en las implicaciones sociales, culturales y tecnológicas de la minería de datos, *Todo el mundo miente* incide en las partes más esquivas de nuestra personalidad: qué pensamos en la soledad de nuestro hogar, qué opinamos de verdad sobre los demás, qué preocupaciones bullen a cada momento y no nos atrevemos a compartir con nadie. Como el mismo autor admite, *Todo el mundo miente* sería una versión mejorada y supervitaminada del clásico libro *Freakonomics*, del economista

Steven Levitt, un volumen que llegó a encaminar los pasos profesionales de Stephens-Davidowitz tras licenciarse en Filosofía. Si el libro de Levitt trataba de examinar montañas de datos para describir cómo funciona realmente el mundo, derribando un buen puñado de mitos e ideas que todos dábamos por sentados, Stephens-Davidowitz hace lo propio asesorado por un volumen de datos que excede todo lo imaginable hace apenas una década.

Más datos que nunca

En el siglo XIX, el médico británico John Snow llevó a cabo uno de los estudios de salud pública más exitosos de todos los tiempos gracias al análisis de datos. De resultas de un brote de cólera en Londres, en vez de abordar el problema con las herramientas disponibles, quiso ir un poco más allá: concibió un mapa de todos los casos registrados en la ciudad, descubriendo que la enfermedad se concentraba alrededor de una bomba de agua en concreto. En aquel entonces, la idea más popular era que las enfermedades se transmitían a través del aire, pero Snow, gracias al poder de los datos, descubrió que todo el mundo estaba equivocado: el origen del contagio era el agua infectada de gérmenes. Por primera vez, de una forma elegantemente eficaz, se ponía de manifiesto cuán lastrados habíamos estado en los últimos siglos, e incluso milenios, por opiniones personales, corazonadas, prejuicios y, sobre todo, las largas sombras de las falacias de autoridad (lo dijo el sabio X) o *ad populum* (de todos es sabido que…).

A partir de entonces, si un investigador quería probar una hipótesis debía esforzarse por encontrar datos que la confirmaran. Datos contantes y sonantes lo más alejados posible de sesgos y subjetivismos. Esto es relativamente frecuente en el campo de las ciencias duras, como la Física o la Química, pero en Psicología o Sociología, por el contrario, reunir datos objetivos nunca ha sido tarea fácil y básicamente se han empleado métodos imprecisos como rellenar casillas de cuestionarios, realizar sondeos y encuestas telefónicas, llevar a cabo estudios longitudinales y poco más. Todo ese cuerpo de evidencia, pues, siempre ha tenido un alcance limitado, tanto en cantidad como en calidad de las fuentes.

Ahora, gracias a Internet, las fuentes se han multiplicado exponencialmente. Todos los usuarios están cediendo datos personales que pueden almacenarse para siempre: palabras, clics, visitas de páginas, enlaces, errores tipográficos, búsquedas en Google, estados de Facebook, fotografías en Instagram. Originalmente, Internet fue creado para que las personas pudieran comunicarse entre sí, pero no para que los investigadores pudieran averiguar cosas de la gente. Sin embargo, toda esa información que distribuimos de buena gana por la red no es sólo ruido de conversación: son migas de pan, pistas que nos permiten establecer correlaciones estadísticas muy significativas. Nadie pudo imaginar que Internet pudiera llegar a ser una fuente adecuada de información para realizar investigaciones académicas, pero lo

es. Para Stepens-Davidowitz, el motor de búsqueda Google, de hecho, es el más poderoso contenedor de datos sociológicos del mundo: «La acción cotidiana de escribir una palabra o frase en una caja compacta y rectangular blanca deja un pequeño rastro de verdad que, multiplicado por millones, acaba revelando realidades profundas».

Pero no se trata sólo de mayor cantidad de fuentes, sino de mayor calidad de las mismas: cuando buscamos algo en Google, lo hacemos con el convencimiento de que nadie está mirándonos, así que nos conducimos con absoluta libertad, algo que, oh, sorpresa, no suele parecerse en nada a lo que solemos responder cuando nos someten a un cuestionario telefónico, por ejemplo. Como si fuera un confesionario de ceros y unos, en Google trasladamos inquietudes que no nos atrevemos a compartir con nadie más. Sobre todo si guardan relación con temas sensibles, como el sexo, nuestro parecer sobre otras etnias o qué nos desagrada de nuestra pareja o incluso de nosotros mismos. Por ejemplo, ¿los votantes de derechas son más racistas que los de izquierdas? Según los datos de búsqueda de Google, el racismo se reparte de forma bastante equitativa por todo el abanico ideológico. De hecho, parece ser más importante dónde has nacido que tu orientación política a la hora de determinar cuán racista eres. Tampoco influye en el racismo la inseguridad y la vulnerabilidad económicas. Ni siquiera aumenta el racismo cuando aumenta el desempleo. Estas revelaciones resultan de todo punto desconcertantes, porque hasta ahora sólo podíamos saber si alguien era racista si lo admitía públicamente (o si se sometía a una serie de pruebas psicológicas, como la de asociación implícita, de alcance ciertamente limitado).

Éste es el deslumbrante poder de los datos: demostrarnos que, en muchos ámbitos, el mundo funciona justo al revés de lo que creíamos. Particularmente, la capacidad de radiografiar la naturaleza humana de Google no reside en la cantidad de datos que alberga, sino en la honestidad inaudita de los usuarios cuando utilizan su motor de búsqueda.

Con todo, después de que Stephens-Davidowitz haya desplegado toda suerte de correlaciones que desafían nuestro sentido común o los estudios que dábamos por válidos, en temas relacionados con el aborto, la homosexualidad, la importancia de la universidad a la que acudimos o de qué debe hablarse en una cita romántica si se aspira a tener una segunda, el autor se cuida de advertirnos muy juiciosamente que los macrodatos no son la panacea: en primer lugar, hay que formular las preguntas adecuadas para que los macrodatos las respondan; en segundo lugar, no debemos desestimar por completo los resultados de investigaciones de otra naturaleza, incluso de las encuestas; y, tercero y más importante, los macrodatos sólo son verdaderamente útiles si reúnen una serie de características muy concretas. Algunas de ellas, por cierto, están aún por definir.

Fiabilidad

Estamos ante una ciencia muy reciente: carecían de ella todos los pensadores e investigadores del mundo hasta hace sólo veinte años. Es decir, que todavía queda mucho trabajo por hacer a fin de optimizar los datos de fuentes como Google. Acabamos de rascar la superficie. Las primeras pistas son emocionantes, pero hay que proceder aún con la cautela de quien abre senda en un campo minado.

Dado que el autor se dedica o se ha dedicado profesionalmente a la ciencia de los datos, el hecho de que advierta de las limitaciones de la misma es toda una rareza en un libro de divulgación, donde la épica, el sesgo de confirmación y el *cherry-picking* suelen prevalecer sobre la honestidad intelectual y la búsqueda de la verdad. «Mucha gente ha hecho grandes afirmaciones sobre el poder de los macrodatos, pero han aportado pocas pruebas», nos avisa, echando con ello piedras sobre su propio tejado. También se cuida de señalar que los macrodatos son interesantes si se usan en masa, no de forma individualizada, y, por supuesto, desvaneciendo la identidad de cada una de las personas que los producen: «debemos guardarnos de permitir que el Gobierno intervenga a nivel individual en respuesta a los datos de búsqueda», tanto por motivos éticos como legales.

Por otro lado, si bien podemos detectar correlaciones significativas, por la propia naturaleza de los datos (son demasiados, básicamente), somos incapaces de averiguar su causalidad. Por ejemplo, se ha podido determinar una correlación absurda pero consistente en la que, ante el anuncio de un tornado en Estados Unidos, los consumidores suelen comprar más Pop-Tarts (unos dulces que se calientan en la tostadora) y linternas. Se ignora la razón de ello. De hecho, puede que ni siquiera los consumidores sean capaces de explicar por qué lo hacen. Sin embargo, la cadena de supermercados Wal-Mart se hace con un mayor número de existencias de estos dos artículos cuando se anuncian tornados para hacer frente a los picos en la demanda. Es decir: sabemos lo que pasa, pero no la razón por la que pasa. Ganamos fiabilidad, pero no lo sabemos todo.

Muchas cosas buenas

Todo el mundo miente no es un libro sesudo ni resultará indigesto para paladares poco habituados a esta clase de temas. Más bien constituye un primer acercamiento a una nueva ciencia que debería interesar a cualquier persona, con independencia de cuál sea su desempeño o sus intereses personales, porque los macrodatos aspiran a influir en todas las áreas del conocimiento, al mismo tejido profesional e interpersonal. Precisamente por eso, el autor se ha preocupado de salpimentar todos los capítulos del libro con anécdotas, hallazgos curiosos y otras hierbas. Más que un libro, pues, en ocasiones parece más bien un conjunto de reportajes de semanario dominical hilvanados en un libro, o la versión *digest* de un estudio sobre *big data* con su dosis de pleonasmos y tautologías. Pero esto no debe

interpretarse necesariamente como un demérito: la obra aspira así a seducir a toda clase de lectores, y eso es, también, muy necesario en un campo del conocimiento que parece restringido a Silicon Valley.

Además, consciente de esta avalancha de datos que cambian el paradigma de la forma en la que adquirimos conocimiento, el autor también se cuida de deslizar una serie de ventajas psicológicas y sociales que considera inherentes a los macrodatos y que mitigan un poco el temor a perder la privacidad o a que nos utilicen como conejillos de indias: que todos somos mucho más parecidos de lo que creemos (precisamente en lo que menos pensamos que lo somos), que hay gente que sufre aunque no lo sepamos y que merece ser asistida, que podemos hallar mejores soluciones a algunos problemas que parecen no tener remedio y que, en suma, los macrodatos permiten a los científicos realizar ensayos aleatorizados de forma masiva, barata y rápida.

En definitiva, *Todo el mundo miente* es el libro ideal para iniciarse en el mundo del *big data*. También es una forma estupenda de que, a cada página, a uno se le desencaje la mandíbula o se le dilaten los ojos por la sorpresa o la sensación de maravilla. Porque la ciencia de los datos va a cambiarlo todo, forjando a nuevos investigadores en todas las áreas, como concluye Stephens-Davidowitz:

> Tengo la firme sospecha de que el próximo Kinsey será un científico de datos. El próximo Foucault será un científico de datos. El próximo Freud será un científico de datos. El próximo Marx será un científico de datos. El próximo Salk, sin duda, será un científico de datos.

Sergio Parra es editor y coordinador de Xataca Ciencia www.xatakaciencia.com, editor de los medios digitales *Diario del Viajero* https://www.diariodelviajero.com/ y *Papel en Blanco* https://papelenblanco.com/, y colaborador de *QUO, Mètode, Revista Arbitrada Columnata, Yorokobu* y *Jot Down*. Es autor de *300 lugares de verdad que parecen de mentira* (Barcelona, Martínez Roca, 2013).

Autor: Seth Stephens-Davidowitz

Edicion: Madrid, Capitán Swing, 2019

Paginas: 296 pp. 19 €

Precio: COMPRAR ESTE LIBRO

Titulo: Todo el mundo miente. Lo que Internet y el big data pueden decirnos sobre nosotros mismos