

**Transversal**
José García MontalvoCatedrático de
Economía (UPF)

‘Big data’ con fin social



Invierno en Minnesota. Unos 15 grados bajo cero. Un coche se acerca a un gran centro comercial. Su dueño sale y se dirige a Target, unos grandes almacenes. Pregunta por el encargado y descarga sobre el mostrador un montón de cupones. Entre los mismos se pueden ver descuentos para pañales, potitos, peluches, leche en polvo y cunas. El encargado no entiende. El cliente comienza a gritar: “¿Cómo se atreven a enviarme a mi hija, que todavía está en el instituto, todos estos descuentos inapropiados para su edad? ¿Están intentado que se quede embarazada?”. El encargado, muy afectado, se disculpa muchas veces, pidiéndole el número de teléfono al cliente para contactarle y ofrecerle un regalo por los trastornos ocasionados.

Al cabo de dos días el encargado llama al cliente y comienza la conversación volviendo a disculparse. El cliente le interrumpe y le pide que no se disculpe. “Bajo el techo de mi casa estaban pasando cosas que desconocía”. La hija estaba embarazada y Target lo había sabido antes que su propio padre. El *New York Times Magazine* atribuyó la confección de este sistema predictivo a Andrew Pole, un estadístico y economista que llevaba trabajando desde

el 2002 en Target. Pole fue capaz de identificar 25 productos que, analizados conjuntamente, le permitían asignar a cada comprador una puntuación predictiva de embarazo. Tener un buen sistema para predecir el embarazo de las clientas es una especie de Santo Grial para los grandes almacenes. Si una clienta puede ser identificada en el segundo trimestre de embarazo hay una elevada probabilidad de captarla para muchos años. Se empieza con los pañales y la leche para bebés y se acaba cogiendo también un paquete de detergente o un DVD.

Aunque existen dudas sobre la veracidad de esta historia, éstas son irrelevantes para el argumento. De hecho, en este caso incluso Target tiene interés en negarla aunque fuera verdad. Lo cierto es que este sistema predictivo sería factible sin ningún problema dado los datos que recopilan los grandes almacenes sobre los hábitos de consumo de sus clientes. Como este ejemplo, podríamos encontrar cientos como las recomendaciones muy acertadas de Amazon y Netflix, la capacidad de Google de individualizar la publicidad con gran éxito, etcétera. El *big data* están contribuyendo de forma muy destacada a mejorar la eficiencia de multitud de actividades privadas. Las empresas privadas están teniendo gran éxito en resolver problemas de mer-



Datos sociales
Hay entes públicos que no están interesados en conocer si un programa ha funcionado porque la respuesta podría ser negativa

cado usando la tecnología y el *big data*.

Pero, ¿y su utilización para resolver problemas sociales? Aquí las cosas van mucho más retrasadas a pesar de las enormes ventajas que proporcionarían para la investigación, la gestión de la movilidad, la gestión del parque de viviendas, etcétera, aunque algunos países están más avanzados que otros. Por ejemplo, en EE.UU. es posible saber a nivel de calles la probabilidad que los hijos acaben viviendo mejor que sus padres. Raj Chetty y sus colaboradores han fusionado los datos de la Agencia Tributaria norteamericana con datos de los censos de población. Así han podido asociar padres e hijos y, analizando bastante años de datos, comparar la posición en el ranking de la renta de los padres con la posición de sus hijos. Y esto a un nivel geográfico increíblemente detallado. El atlas se puede consultar en <https://opportunityinsights.org/data/>. Con estos datos sabemos que más del 90% de los nacidos en 1940 han tenido un nivel de vida superior al de sus padres, mientras que entre los nacidos en 1984 sólo el 50% vivirá mejor que sus padres. Y esto lo podemos saber por estados, ciudades e incluso distritos. Los datos muestran con claridad que el sueño americano ha dejado de ser realizable para una gran mayoría de la población de Estados Unidos. Este es un ejemplo de la enorme capacidad de los datos, utilizados adecuadamente, para responder a preguntas sociales importantes.

¿Se imaginan que algún investigador intentara hacer esto en España? Pues si lo han hecho tienen una imaginación portentosa. En España todo son problemas y más problemas. Y retrasos y más retrasos. Los que profesionalmente nos dedicamos a la investigación social hemos de pasar larguísimo procesos de solicitudes, resoluciones, y *re-resoluciones* para acabar con el silencio en la gran mayoría de los casos. Las excusas son de todo tipo: solicitar autorizaciones a organismos superiores, supuestos problemas con la ley de protección de datos, la consulta a los servicios jurídicos durante varios años, etcétera. En un ayuntamiento cercano nos tuvieron mareando la perdiz cuatro años por una solicitud de datos *anonimizados* sobre solicitantes y beneficiarios de loterías de viviendas de protección oficial para acabar argumentando que, como ya había pasado tiempo desde que los datos se generaron, tenían que destruirlos. ¡Encima pitorreo! Mientras las empresas privadas resuelven sus problemas de mercado utilizando los datos intensivamente, los problemas sociales y la evaluación de las políticas públicas siguen sin avanzar a pesar de la ingente cantidad de datos administrativos existentes. Es evidente que en muchas ocasiones son los propios organismos públicos los que no están interesados en saber si un programa ha funcionado puesto que la respuesta podría ser negativa y políticamente inconveniente.

El colmo del despropósito es la polémica generada por el estudio del INE sobre movilidad basado en el análisis de la geoposición de teléfonos móviles durante 8 días. Esta metodología es muchísimo más útil y barata para los estudios de movilidad que la anterior, basada en los censos. A pesar de que la información recibida será anónima y agregada por zonas, lo que impedirá cualquier posibilidad de reidentificación, se ha montado un gran revuelo. Es difícil avanzar en el uso de los datos para el bien social en una sociedad de personas que no tiene ningún problema en ceder toda su información personal para instalarse un app de linterna en el móvil, pero se escandaliza cuando un organismo oficial quiere usar los datos para mejorar la movilidad de los ciudadanos. |

Administración ausente

Mientras las empresas privadas se benefician de los datos de sus clientes y usuarios, los entes públicos desaprovechan el *big data* para crear políticas sociales adecuadas

