



*Facultad de Ciencias Económicas
Universidad de Buenos Aires*



Big data, opacidad y nula reflexión

Alberto Müller¹
IIE-CESPA-FCE-UBA

Nota Breve N°47
12/07/2018

¹ Se agradecen los útiles comentarios de Javier Perdomo y Julio Ruiz.

Av. Córdoba 2122
2do. Piso, Departamentos Pedagógicos
(C 1120 AAQ) Ciudad de Buenos Aires
Tel.: 54-11-5285-6583 – E-mail: dircespa@econ.uba.ar
<http://www.econ.uba.ar/cespa>
www.blogdelcespa.blogspot.com

Una vez, en un curso sobre Economía Urbana, un urbanista presentó un conjunto de ejercicios basados en el cálculo sistemático de correlación de variables estadísticas, explicando la formación de determinados patrones en la configuración urbana. Allí señaló – la cita es de memoria – que “no se necesita teoría” para asociar diferentes variables, porque con los datos es suficiente.

Empirismos tan crudos e ingenuos han sido rebatidos ya reiteradamente. Porque en realidad este urbanista - a la manera del burgués gentilhomme de Moliere que hablaba en prosa sin saberlo – estaba haciendo teoría sin saberlo. Su “teoría” - bastante simple – consiste en suponer que los patrones obtenidos mediante correlaciones que ha observado permanecerán en el futuro; en caso contrario, no serían patrones relevantes.

No es que esta presunción sea necesariamente equivocada; tratándose de áreas urbanas relativamente consolidadas y de crecimiento relativamente lento (la de la anécdota es el Área Metropolitana de Buenos Aires), no es desatinado suponer que un conjunto importante de características permanezcan en el tiempo: las poblaciones urbanas son bastante sedentarias, las actividades productivas no se re-localizan permanentemente, etc. O sea, ese análisis puede ser la base de una buena predicción; pero una explicación que fundamenta esta predicción no surge de los hechos observados, sino de comportamientos que suponemos que tienen una razón de ser. O sea, en una teoría.

Porque siempre hay teoría, en el sentido de construcción en abstracto, para establecer por ejemplo relaciones causales. Abstracción y juicio causal no son resultados de la mera experiencia, sino una forma de nuestro pensamiento.

Pero una y otra vez reaparece la fantasía del empirismo ingenuo, que plantea prescindir de la reflexión teórica, atenernos a los hechos y extraer verdades de allí. La anécdota inicial es del año 1978; el urbanista es Horacio Torres; el instrumento empleado entonces es el cluster analysis (análisis de conglomerados). Hoy día, este papel le toca al big data y a los análisis que éste viabiliza.

Pero hay claras diferencias entre el cluster analysis y lo que hoy damos en llamar big data. El cluster analysis consiste en un conjunto de técnicas que permiten investigar patrones en un conjunto amplio pero delimitado de datos. Big data es un término que se refiere a la posibilidad de gestionar volúmenes inusualmente grandes de información, muchas veces crecientes a ritmos elevados, mediante equipamiento y algoritmos apropiados. Por extensión, el término se asocia a ciertos usos específicos que permite este procesamiento, como por ejemplo la evaluación del desempeño de educadores, la definición del perfil crediticio de los individuos, etc. Acá emplearemos el término en su sentido más acotado.

Ahora bien, la posibilidad de procesar grandes y crecientes cantidades de información a bajo costo es algo reciente. Por ejemplo, los textos de las décadas de 1950 y 1960 que trataban la temática del entonces muy popular Modelo de Leontieff (Modelo Insumo-Producto), dedicaban buena parte de sus páginas a explicar procedimientos para simplificar la operación de inversión de una matriz cuadrada, dado que los métodos canónicos eran extremadamente trabajosos. Este problema desapareció 20 años más tarde. No podemos decir que hoy día la capacidad de cálculo es ilimitada, porque nada lo es; pero para un rango muy amplio de cuestiones, es como si lo fuera.

Por otro lado, el empleo de soportes informáticos para las actividades diarias ha crecido exponencialmente; y esto ha incrementado la disponibilidad de datos procesables. Si antes se requería “traducir” la información a un formato que fuera legible para el equipo de computación, hoy día ella en buena medida ya está en este formato, desde el inicio. La informática integra nuestras prácticas diarias, en varios ámbitos, y ésta es la base para un crecimiento continuo de los volúmenes de información procesable..

O sea: hemos ampliado la capacidad de cálculo, y la información se encuentra lista – o casi – para el procesamiento. Esto comporta un cambio muy grande, que al parecer, ninguno de los “visionarios” en cuestiones tecnológicas (desde Leonardo da Vinci hasta Julio Verne o

Tomás Edison) anticipó. No es algo para lo que estemos mental y culturalmente preparados; de allí que no debe sorprendernos que, una vez más, muchos hayan caído en el empirismo ingenuo.

Es de suponer que esta creencia de que los datos pueden suplantar una teoría no perdurará. Esto vale no solo para la ciencia, sino también para la tecnología (sin entrar ahora en disquisiciones acerca de la relación que existe entre ambas). Nadie duda de que la estadística es una herramienta indispensable en la medicina o la ingeniería; pero ningún médico o ingeniero prescinde de un marco teórico de referencia para sus decisiones, sea éste originado en la ciencia o en la tecnología. Ambos saben que los hechos ocurren por alguna razón, que debe haber una explicación satisfactoria.

Borges ilustró con la maestría del poeta de lo intelectual que supo ser la falacia empirista, al relatar el caso de Funes el memorioso, un individuo que repentinamente queda dotado de la capacidad de recordar absolutamente todo. Borges nos dice que Funes “no era muy capaz de pensar. Pensar es olvidar diferencias, es generalizar, abstraer. En el abarrotado mundo de Funes no había detalles, casi inmediatos.” El pensamiento, la abstracción, es, en el argumento de Borges, un sucedáneo para la desmemoria. Pero seguramente Borges no suscribía un pensamiento tan ingenuo; era demasiado inteligente, y además muy afecto a la ironía.

Jocosamente, diríamos que Borges habría anticipado en Funes lo que otros no pudieron anticipar: el big data. Porque efectivamente el big data (unido quizá a internet) nos produce esa percepción: la de que ahora podemos “acordarnos de todo”. No nosotros, sino los equipos informáticos. Y la forma de gestionar ese volumen de información son los algoritmos del big data, una suerte de cerebro operador sobre esa información. Y allí esta asociación entre datos y procesamiento en niveles asombrosos aparece sustituyendo nuestro razonamiento, incapaz de operar con semejantes algoritmos, manejando semejante nivel de información.

Pero, ¿actúan estos algoritmos emulando nuestro cerebro? Hay una tendencia ya clásica a aceptar esta tesis: alguna vez a las computadoras se las denominaba “cerebros electrónicos”.

No lo creo así. Hay un capacidad de asociación creativa del cerebro que el otro cerebro (“electrónico” o como se lo llame) no logrará nunca. Se dirá que los algoritmos asocian, al igual que el cerebro. Esto recuerda la clásica historia de los monos y los teclados: un conjunto de monos ante un conjunto de teclados puede escribir el Quijote (o el Hamlet, o cualquier otro clásico, ya que la historia es clásica). La computadora podría asociar igual que nosotros, solo hay que darle tiempo. Y lo cierto es que las computadoras cada vez requieren menos tiempo.

Ahora bien, sabemos que los monos necesitan infinito tiempo para escribir el clásico y lograr el Premio Nóbel de Literatura. Y ésta no es sino una forma elegante (o retorcida, dirán algunos) de decir que no es posible en los hechos; esto es lo que nos pasa cada vez que invocamos al infinito para tratar nuestras prácticas. En infinito tiempo, la computadora será capaz de reproducir todas las asociaciones posibles que puede hacer el conjunto de los cerebros humanos; o sea, no podrá hacerlo en la práctica.

Pero hay algo más: juzgamos que el Hamlet, el Quijote, la Divina Commedia o el Martín Fierro son obras cumbre, geniales, originales, penetrantes... Y esto nunca una computadora podrá decidirlo por sí misma, aun si fuera capaz de escribirlos. Sobre todo, porque el juicio sobre esas obras es diverso – no faltarán quienes digan que el Martín Fierro es muy superior al Quijote, mientras que otros dirán que es sacrílego pretender que el Martín Fierro sea comparable al Quijote – y cambia con los tiempos y las culturas. Porque este juicio emerge de prácticas individuales y colectivas para las que hasta ahora no encontramos forma de racionalizar o explicar, y por ende en convertir en un algoritmo. Un cuadro de Miró en el Siglo XII europeo habría resultado quizá inaceptable; pero tal vez

habría sido apreciado sobremanera por los mayas o los chinos de aquel entonces, quién sabe. Pretender lo contrario es suponer que es posible teorizar sobre cómo se forman estas preferencias; pero esta formación de preferencias no reconoce patrones o recurrencias, es una sucesión de hechos únicos.

Igualmente, ya en el plano del desarrollo científico, la sustitución de la teoría del flogisto por la del oxígeno no ocurrió por obra de correlaciones estadísticas. Como señalara Kuhn (en su clásico “La estructura de las revoluciones científicas”) en cierto sentido Lavoisier “inventó” el oxígeno, cuando las experiencias le fueron mostrando las fallas de la teoría del flogisto.

En definitiva, es una tontería pensar que las computadoras piensan como nosotros; es posible que piensen por nosotros (si insistimos en el uso del término “pensar”) en determinadas instancias. A veces conviene liberarnos de la obligación de ese pensamiento, porque las computadoras lo hacen más eficientemente. Si voy al supermercado y quiero saber cuánto me habría costado el carrito de bienes que estoy comprando hace un año, hacerlo manualmente me llevaría un tiempo muy grande; pero no falta mucho para que esta pregunta sea respondida en forma automática (o a pedido) por el equipo de facturación que regentea el cajero del supermercado (o uno mismo, en el supermercado).

Es posible que ayuden a asociar, también. Pero nunca tendrán la capacidad de sustituirnos en la formación de un deseo, de un plan, o de una teoría. Plantear que la computadora puede sustituir nuestro pensamiento es como pensar que la lanza del hombre primitivo podría decidir qué animal cazar.

Dicho esto, está claro que hay peligros en el big data, peligros que emergen precisamente cuando decidimos abdicar del pensamiento propio, en nombre de la superioridad del empirismo ingenuo.

Un caso que ha surgido reiteradamente en Estados Unidos es la aplicación de algoritmos para evaluar el desempeño docente; así lo relata Cathy O’Neal , en su libro “Weapons of Math destruction”. Allí relata el caso de una docente que recibe una calificación negativa, y por lo tanto es despedida. Lo interesante es que cuando solicita explicaciones al director del establecimiento, la respuesta es que el resultado de un algoritmo complejo, cuyo funcionamiento él desconoce. Analizando luego lo ocurrido, concluye que su mala calificación es producto de que los docentes del año anterior calificaron bien a sus alumnos, precisamente para no tener inconvenientes por el hermético algoritmo que evalúa...

Si se tratara de un algoritmo sencillo aplicado manualmente – la planilla de puntaje confeccionada por una comisión evaluadora, por ejemplo – esto no sería aceptable. Se supone que los procedimientos de evaluación deben ser conocidos y transparentes, porque de esa forma el evaluado sabe cómo desempeñar sus funciones. Lo contrario implica incurrir en arbitrariedades (aunque esto ocurre; lo he constatado actuando en comisiones de evaluación de investigaciones).

Pero cuando la evaluación procede de un complejo algoritmo retro-alimentado con datos masivos en forma permanente, el procedimiento adquiere una respetabilidad distinta. Un algoritmo complejo es elaborado necesariamente por gente inteligente y perspicaz, y si además se funda en grandes volúmenes de datos “reales y actuales”, parece estar dotado de inatacable legitimación, y no nos resta sino atenernos a lo que produce. Es más, pareciera que la opacidad le brinda objetividad: si nadie conoce cómo funciona, seguramente es neutro, y no está diseñado para favorecer a alguna persona o grupo en especial.

Se produce así una enajenación: un procedimiento producido por personas por encomienda legítima de la sociedad se convierte en un instrumento de dominación sobre la propia sociedad. El algoritmo gobierna autónomamente, y “cobra vida”, lo que en este caso significa que es imprevisible. Es más, cobra carácter divino: una posible definición de “dios”

es la de un ser imprevisible, pero que no por eso deja de tener razón (aunque no hace falta ser un dios; los tiranos son también así).

Pero el rústico empirismo que legitima este procedimiento no impide errores. El ejemplo típico, también citado por O'Neal, es el de la asignación de una persona a un cierto grupo de riesgo crediticio. Ello se hace no en función de sus particulares características, sino en función de su similitud con otros individuos. Aquéllos que en el pasado tuvieron las características A (ser afros), B (tener secundaria incompleta) y C (vivir en un suburbio mal entrazado) entraron en default en el 60% de los casos. Entonces, todo individuo que cumpla con A, B y C tendrá un recargo de x% en un eventual préstamo que se le conceda. El banco puede justificar el procedimiento, como forma de salvaguardar sus activos. Pero lo cierto es que esto no implica que pueda preverse el comportamiento del próximo deudor. Es un procedimiento convencional, que la gerencia del banco puede adoptar y justificarse así ante sus accionistas o ante las autoridades regulatorias. Pero no es más que eso. Afirmar que la probabilidad de que un individuo no pague cuotas de su préstamo es de 60% no tiene sentido alguno, porque no hay probabilidades sobre hechos únicos. El default es como el embarazo, ocurre o no ocurre; y no sabemos si ocurrirá o no con el próximo deudor. Y además – pero esto ya no es atribuible al big data sino al sistema bancario – el mayor recargo incrementa la posibilidad del default.

Pero cuando este procedimiento se generaliza, señala O'Neal, se produce una suerte de círculo vicioso. Quienes cumplen con las características X, Y, Z pueden no ser aceptados en universidades, o ser rechazados en posibles empleos. Todo esto favorece la exclusión o la segregación social: para algunos, no hay crédito, no hay acceso a educación, no hay empleos de calidad; por lo tanto, las mencionadas características se mantendrán sin posibilidad de corrección.

Y esto no es sino el resultado de haber renunciado al conocimiento como entendimiento, interpretación o análisis. El big data nos promete mucha, mucha información, y la permanente realimentación, con lo que se nos asegura que los eventuales errores se corrigen con más información. Este procedimiento parece prescindir del razonamiento teórico, cuando en realidad lo que hace es aplicar una interpretación subrepticia y muy corta de miras; correlaciones inmediatas leídas como causalidades, que no llegan a articularse en alguna visión más amplia, y por lo tanto carentes de sabiduría, útiles para el gestor de una organización sujeto a determinados objetivos, pero no para la gestión de una sociedad.

Contrariamente a lo que sostiene O'Neal, sin embargo, esto no es “culpa” del big data, en sí mismo, de la misma forma que un cuchillo no es culpable de un asesinato. Lo que sí hace el big data es potenciar enormemente estos mecanismos perversos, que ya existen, de la misma forma en que un arma de fuego es más eficaz que un cuchillo, a la hora de un robo o asesinato.

El capitalismo es una organización social cambiante, pero que, con relativamente pocas excepciones, no se destaca por ser inclusivo, por cuanto responde en forma primordial a la motivación del lucro individual. Y esto ocurre con o sin big data. Se pueden morigerar estas características por la acción contrarrestante de la sociedad civil, a través principalmente, pero no únicamente, del Estado.

Big data nos plantea un desafío precisamente a propósito de evitar que las tendencias excluyentes del capitalismo campeen sin límites a partir de esta herramienta. Se trata en realidad de un aprendizaje, por parte de la sociedad civil, para sobreponernos a la perspectiva enajenante que parece prevalecer hoy día. Pero será un aprendizaje como muchos otros que han debido encararse, ante cambios tecnológicos profundos, abundantes en las últimas dos centurias en buena parte del mundo.