

MÁS ALLÁ DE LA ETIQUETA BIG DATA: EL DATO EN EL NÚCLEO DE LA GENERACIÓN DE VALOR Y LA INNOVACIÓN

1. LA DIGITALIZACIÓN EN EL ORIGEN

Las etiquetas y las modas temáticas facilitan la divulgación acelerada de conceptos y tendencias prácticamente con la misma potencia con la que, al tiempo, incurren en un efecto no deseado, pero aparentemente inevitable: el riesgo de la trivialización y vaciamiento de contenido de aquello que pretenden difundir, quién sabe si por un exceso de uso en un breve lapso temporal o por una escasa profundización en su comprensión y alcance real. Esa moneda de dos caras que conforman el haz y el envés de la comunicación moderna afecta, como no podía ser de otra manera, a uno de los conceptos más citados de los últimos tiempos. No cabe duda: el Big Data está de moda. Y como acostumbra a ocurrir con los conceptos en boga, su simplificada etiqueta mistifica gratuitamente alguna de sus potencialidades técnicas y efectos -que desarrollaremos en las páginas que siguen- y omiten pliegues, paradojas, limitaciones y algún que otro dilema.

Previamente a deliberar sobre el *tamaño* de los datos y la pertinencia del uso indiscriminado del adjetivo 'Big' para definir el surgimiento de un nuevo paradigma, deberíamos atender a su etimología, al origen conceptual del 'dato': procedente del latín 'dare', ofrecer, los datos constituyen elementos *crudos* mostrados por cualquier fenómeno susceptible de ser observado y registrado. Y que, por ser más precisos, cabe tomar, captar, bien sea a través de la observación, la experimentación, el registro, más modernamente la programación, y -en definitiva- cualquier fórmula de medición.

JON GOIKOETXEA GOIRI

SOCIÓLOGO. PROFESOR DEL PROGRAMA
BIG DATA Y BUSINESS INTELLIGENCE DE LA
UNIVERSIDAD DE DEUSTO

En esta era digitalizada que habitamos la auténtica revolución -el auténtico salto paradigmático y hasta epistemológico- consiste en la infinitud de fuentes de información y datos, en su anonadante exhaustividad y trazabilidad. En la progresiva capacidad de nuestras sociedades, de nuestros modelos económicos en plena vía de transformación -desde las empresas y sus procesos internos, los mercados y la actividad económica, pasando por nuestros datos financieros, hábitos de consumo o nuestras identidades digitales, y llegando hasta la sensorización misma de nuestras ciudades- para generar constante información traducible en toda suerte de métricas que alimentan modelos analíticos iterativos y de elevada complejidad que se incorporan al núcleo mismo de la productividad. De ahí que, como señala Rob Kitchin¹, más que hablar de una 'revolución Big Data' en sí misma deberíamos referirnos como sustrato de todo ello -como su condición de posibilidad- a la 'revolución en la capacidad para la generación y obtención de datos'².

En el año 2004 Jeffrey Dean y Sanjay Ghemawat -ambos ingenieros de Google- presentaron un nuevo modelo de programación centrado en la simplificación del procesamiento de grandes volúmenes de datos³, acuciados por el éxito del buscador para el que trabajaban y las necesidades de gestión derivadas de un volumen en crecimiento exponencial formado por documentos, páginas webs o referencias de todo tipo creados desde cualquier lugar del mundo a cada segundo. El tradicional algoritmo page rank, creado a finales de la década de los

90 por los fundadores de Google, Larry Page y Sergey Brin, se reveló insuficiente en apenas unos años ante el crecimiento inmanejable de la información generada, en una suerte de repositorio infinito.

Bajo aquel modelo denominado 'Map Reduce' surgieron industrias esenciales para el surgimiento de la economía digital como el SEO -'Search Engine Optimization'- la optimización del posicionamiento en los motores de búsqueda de cualquier recurso web- y el SEM -'Search Engine Marketing', el marketing y promoción de webs y la optimización de su visibilidad en los motores de búsqueda-. Más de 12 años después, estas potencialidades continúan constituyendo el principal eje estratégico de Google -reconvertida en Alphabet-: el eje de generación de valor que le ha permitido llegar a ser la empresa de mayor capitalización bursátil del mundo.

En síntesis, el sustrato de este nuevo modelo de programación provenía de la necesidad de procesar grandes volúmenes de datos en una línea de crecimiento tendencialmente infinito a través del andamiaje de un esquema de computación en paralelo que permitiera distribuir, reasignar, la carga de procesamiento entre diferentes nodos, de tal modo que se pudiera reducir el tiempo total resultante de procesamiento de una manera más eficiente. La configuración de los procesos en paralelo reducirían, fragmentarían -y dispararían en alguna medida-, lo que de otro modo no sería sino un único *gran procesamiento* concentrado y con una exigencia de disponibilidad de recursos más acusada e inasequible.

El rendimiento de esta nueva aproximación al procesamiento de datos y su eficiencia impulsó su uso en otros entornos. Esencialmente, a través de versiones de software de código abierto que facilitaron su rápida adopción y difusión a escala global y materializaron de manera progresiva -en

1. Kitchin, Rob, *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*, Sage Publications, London, 2014.

2. "The Capta Revolution", tal y como la denomina Rob Kitchin, op. cit. página 5 y siguientes.

3. Dean, J. Ghemawat, S. "MapReduce: simplified data processing on large clusters", *Communications of the ACM* - 50th anniversary issue: 1958 - 2008, Volume 51 Issue 1, 2008.

un proceso que aún continúa- el objetivo de la universalización del procesamiento de ingentes volúmenes de datos en paralelo. Uno de los *frameworks* más notorios para lograrlo fue desde su origen -y continúa siéndolo en buena medida- Apache Hadoop⁴, creado por un ingeniero de Yahoo! -con origen académico de la Universidad de Stanford- llamado Doug Cutting.

Con todo ello, el volumen de datos disponible y la velocidad de su procesamiento dejaron de ser una condición limitativa insalvable. El tamaño dejó de ser un problema y quedó expedito el campo para el sueño omnisciente de todo analista: disponer no ya de una muestra de los datos sobre los que aplicar modelos de extrapolación, sino de su totalidad. De todos ellos, de todos aquellos que hayan sido conformados y resulten captables a través de procesos de programación y computación, y tenerlos a su disposición de manera trazable.

En definitiva, esta condición de posibilidad -con sus consecuencias desde técnicas y analíticas hasta epistemológicas- supuso el advenimiento del 'Big Data'.

Las empresas que conformaron el mascarón de proa de la llamada economía digital (Google, Yahoo, Twitter, Facebook, LinkedIn), impulsaron y propiciaron el desarrollo de tecnologías Big Data -que posteriormente han sido y siguen siendo utilizadas en otros sectores-, no en menor medida a como contribuyeron a la visualización de sus aplicaciones y potencialidades.

Y, sin embargo, el uso del análisis de datos para la mejora de la información disponible en la toma de decisiones, la competitividad y la eficiencia de las organizaciones no es en absoluto novedoso en sí mismo. No lo es, por ejemplo, para el sector financiero, el asegurador o las

empresas del sector de las (tele)comunicaciones, que llevan decenios aplicando técnicas de minería de datos -sí, el célebre *Data Mining*- para extraer valor de sus capacidades para obtener grandes volúmenes de datos de las interacciones con sus amplísimas bases de clientes -quién no tiene una cuenta corriente, una tarjeta de crédito, una prima de seguro contratada, un teléfono móvil-. Sus ámbitos de aplicación han cubierto habitualmente desde la detección de fraude a la definición de perfiles de clientes con mayores niveles de propensión al impago, o para clasificarlos en términos de *scoring* y de su calificación crediticia. O para segmentarlos y tipologizarlos atendiendo a sus hábitos de consumo.

La novedad tampoco procede estrictamente de la evolución o desarrollo de algoritmos o de sofisticaciones analíticas o estadísticas, sino que responde a la extraordinaria disponibilidad de datos procedentes de muy diferentes y heterogéneas fuentes y formatos de almacenamiento que, por encima de todo, son generados a gran velocidad. Este sí que es un aspecto distintivo y definitorio de esta era del dato y que bien pueden sintetizarse en la combinación de las denominadas tres "V's"

del Big Data: volumen; velocidad; variedad⁵.

El 90% del volumen total de datos generados en todo el mundo⁶ han sido creados apenas en los dos últimos años. En 2011, en pleno Foro Económico de Davos, el dato surgió como un nuevo activo a gestionar y del que obtener valor añadido por parte de empresas y organizaciones. Ya para 2012 los paradigmas de gestión empresarial comenzaron a incluir el dato -el *data*

El volumen dejó de ser un problema y quedó expedito el campo para el sueño omnisciente de todo analista: disponer no ya de una muestra de los datos sobre los que aplicar modelos de extrapolación, sino de su totalidad. De todos ellos, de todos aquellos que hayan sido conformados y resulten captables a través de procesos de programación y computación, y tenerlos a su disposición de manera trazable

4. <http://hadoop.apache.org>

5. Zikopoulos, P. & Eaton, C., *Understanding big data: Analytics for enterprise class hadoop and streaming data*. McGraw-Hill Osborne Media, 2011.

6. SINTEF, "Big Data, for better or worse: 90% of world's data generated over last two years", ScienceDaily, 2016, from www.sciencedaily.com/releases/2013/05/130522085217.htm

management- en las proximidades del núcleo de valor, tal y como McAfee y Brynjolfsson definieron con visión precursora⁷.

Apenas unos años antes, en plenos años 90, el mundo del *Business Intelligence* aportaba sobre los fenómenos observados una visión retrospectiva, explicativa *ex post* acerca de lo ocurrido, se trataba de la evolución de las ventas y su evolución previsible o de una proyección demográfica.

El nuevo paradigma Big Data -pues de eso se trata, en cuanto que crea nuevas configuraciones de representación de la realidad- incorpora tanto el aquí y el ahora, el tiempo real, como el futuro a sus modelizaciones. Invierte los ejes temporales tradicionales de la investigación, volcándolos del pasado al presente -literalmente en tiempo real- para orientarse hacia la predicción del futuro. Y puede contemplar hacerlo por la creciente capacidad para el procesamiento de datos -extremado, optimizado, en la era *cloud* en la que nuestros equipos y dispositivos comparten sus archivos y cargas de trabajo en redes y servidores deslocalizados, en las célebres nubes, *clouds*-. Puede asumirlo porque paralelizar y distribuir las cargas de almacenamiento y procesamiento de datos incrementa exponencialmente la velocidad de los procesos hasta convertirlos en potencialmente activables en tiempo real. Puede posicionarlo en otra dimensión analítica porque no depende de la creación o generación *ad hoc* de indicadores, métricas y experimentaciones de laboratorio, sino de la captación *natural* de datos en entornos vitales y hábitats comunicativos digitalizados como los que habitamos. La digitalización -la programación, la computación- incluye la generación de datos en su mismo mapa genético constitutivo. La digitalización, en definitiva, es dato. Y ese rastro de códigos, marcaciones y programación a través de la que se configura es el sustrato de su ADN, que se extiende a cualquier fuente de datos, desde la más estructurada a la más heterodoxa, como puedan ser los análisis semánticos o de influencia y emociones vía redes sociales como Twitter.

Contemplan cualquier acción realizada, desde su smartphone hasta las operaciones de compra de su tarjeta de crédito; incluyan todos

aquellos procesos de su trabajo cotidiano mediados digitalmente; recuerden el seguimiento de ese blog que alimentan semanalmente, su itinerario por las redes sociales o el gps que emplean para orientarse cuando van de viaje o suben al monte. La lista sería interminable. Hasta tal punto que tal vez finalizaríamos antes tratando de identificar los escasísimos ámbitos exentos de creación de datos. Desde la aparición de las webs y aplicaciones de búsqueda de pareja, tal vez no quede fuera del mundo del dato ni el amor, ni las emociones, si me permiten la boutade.

De esta mirada de fuentes de ingente volumen, de su variedad en formatos y estructuras posibles y de su velocidad acelerada de procesamiento surgen nuevos modelos de análisis e inteligencia del dato. Inteligentes en cuanto que el afinamiento estadístico es significativamente superior sobre grandes volúmenes que mejoran sus ajustes y la efectividad de sus resultados. E inteligentes igualmente por la aplicabilidad de sus modelos en tiempo real, incrementando sus resultados a través de una iteración -recursividad, repetición- continua.

En definitiva, todo lo programable es trazable; y todo lo trazable es analizable.

A todos estos factores de carácter trascendente coadyuvan aspectos económicos, de costes de producción y de mercado que no producen otro efecto que su aceleración: la computación se ha abaratado significativamente y **fabricar ordenadores resulta relativamente económico**, impulsando las teorías de Jeremy Rifkin⁸ sobre la tendencia al coste marginal cero de las sociedades digitalizadas. Este elemento descrito por Rifkin es exactamente el vector de la tecnificación y digitalización de nuestras sociedades, el precipitante de la universalización de dispositivos digitales cada vez más interconectados e integrados, incluyendo incluso objetos físicos a los que se incorpora la codificación, es decir, el núcleo mismo de su *analizabilidad*: hemos llegado ya al internet de las cosas. La raya divisoria de los mundos on line y off line ha difuminado definitivamente su grosor para pasar a entrelazarse o, quién sabe, acabar conformando un único plano, un hábitat con una nueva configuración.

7. McAfee, A. et al.: "Big data". The management revolution", Harvard Business Review, 90.10, 2012, páginas 61-67.

8. Rifkin, J., *La sociedad de coste marginal cero*. Editorial Paidós, Barcelona, 2014.

2. LA RELEVANCIA DEL DATO, NO DE SU TAMAÑO: LOS ÁMBITOS DE APLICACIÓN DEL BIG DATA

Si nos viéramos exigidos a sintetizar en un único elemento catalizador, en un aspecto motivador, en una filosofía, el proceso de digitalización y el consiguiente advenimiento del fenómeno Big Data que hemos venido describiendo hasta el momento, bien podríamos recurrir al célebre axioma de Peter F. Drucker: “lo que no puede medirse, no se puede mejorar”. Huelga comentar que nunca en la historia fue tan accesible como lo es en este momento realizar cualquier tipo de medición de cada vez más amplias capas de actividad humana. Pero más relevante que la capacidad de medición en sí misma es el sentido con el que se dote. Su orientación, su capacidad para responder a las preguntas originarias que definen los objetivos de la herramienta misma de medición. Se trate de dilemas de negocio ante un descenso de ventas de una cadena de distribución o de la sensorización del tránsito de ciudadanos por una ciudad para optimizar la distribución de los recursos públicos empleados. Los datos en sí mismos -informes, literalmente, sin forma si no proceden de una modelización correcta- no contienen las respuestas sin un diagnóstico asociado, sin unas *preguntas* idóneas que los soporten.

Al fin y al cabo, ningún dato -y por supuesto, ninguna tecnología- sustituirá nunca al criterio que lo inspira, alimenta y le confiere sentido. Ni la capacidad de procesamiento y almacenamiento, ni la tecnología, ni los algoritmos. Su potencialidad consistirá en amplificar el criterio, hacerlo más fiable, multiplicar su creatividad y capacidad innovadora. En convertir los datos en información significativa, en conocimiento para sedimentar las tomas de decisión. Y tanto más cuanto más relevantes resulten esas decisiones, cuanto más sensibles. El diseño

de toda estrategia de mejora futura, desde las administraciones públicas a las empresas, debe asentarse sobre este principio que aboga por la sustitución de la intuición y el sentido común por el análisis fundamentado sobre tres conceptos lógicos centrales: la trazabilidad de los fenómenos, la atribución de un hecho a sus causas y la comprensión de los procesos de comportamiento y toma de decisiones en las que nos vemos involucrados cada uno de nosotros a cada segundo.

Recordemos que hace apenas unos pocos años -y eso los avanzados, los denominados *early adopters*- adquiriríamos dispositivos GPS que tenían un precio entre 200 y 300 €.

Hoy en día aquellos aparatos han sido prácticamente reemplazados por aplicaciones como Google Maps o Waze -*app* (aplicación) que presenta el estado del tráfico y permiten conocer la ruta óptima en tiempo real-. Gratis, con el único coste del tráfico de datos empleado a través de nuestro smartphone. La *transacción* entre Google y nosotros constituye una metáfora adecuada de esta era del advenimiento del Big Data: la cesión de nuestros

datos de localización a cambio de un servicio de gran utilidad y diseñado para atender personalmente millones de peticiones de ubicación recurrentes y simultáneas en todo el mundo. El resultado es de unas dimensiones extraordinarias. Para Google, porque dispone -como le ocurre en otros entornos, como Analytics- de un volumen inabordable de datos -al menos concebido desde las unidades de medida tradicionales previas al Big Data- cedidos voluntariamente y sin coste alguno por sus propios usuarios. Para el usuario, porque nunca volverá a perderse en ningún lugar del mundo, por muy desconocido a priori que le resulte: bastará con introducir su destino, trazar la ruta y esperar las indicaciones.

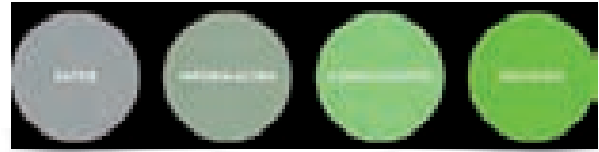
Si nos viéramos exigidos a sintetizar en un único elemento catalizador, en un aspecto motivador, en una filosofía, el proceso de digitalización y el consiguiente advenimiento del fenómeno Big Data bien podríamos recurrir al célebre axioma de Peter F. Drucker: “lo que no puede medirse, no se puede mejorar”

Y sin embargo, como en la vida misma, lo relevante no es el volumen de *data* disponible, sino su ubicación en el centro lógico de la toma de decisiones, su posicionamiento como fuente de alimentación de conocimiento para la definición de estrategias efectivas de empresas y organizaciones. Para innovar creativamente; para hallar fórmulas para incrementar la competitividad -en términos de costes, de reformulación de productos y servicios-. Lo auténticamente relevante de la gestión del dato y de su inclusión nuclear en los procesos de generación de valor no es su tamaño -si son 'Big' (grandes) o no-, sino si son gestionados y tratados de una manera suficientemente 'smart' (inteligente). De manera aplicable, orientados a estimular la efectividad, trazables y evaluables -¿de qué sirve analizar, planificar, si carecemos de herramientas para medir si nuestro planteamiento es eficaz o no?-. Si usted trabaja en una pequeña cafetería es altamente improbable que la cantidad de datos de la que disponga sea lo suficientemente volumétrica como para ser calificada 'Big'. Pero ni falta que le hace, porque pese a ello usted dispondrá, sin duda, de una cantidad estimable de métricas que contribuirán a mejorar su negocio y a comprender con mayor efectividad su capacidad para satisfacer a sus clientes y las pautas de la interacción que mantienen con su establecimiento. Una guía en un entorno económico cada vez más complejo, sincopado y variable. Cuyas claves de comprensión distan cada vez más de ser lineal e intuitivamente previsible.

Sea cual sea su actividad profesional y su ámbito, la explotación y la comprensión de los datos no le deben resultar ajenas: reducirán significativamente su grado de exposición a la incertidumbre y la improvisación irremediable.

Desde cualquier perspectiva posible -sea la financiera, la interna de procesos, la orientada al cliente, tan esencial; sea la innovación y el aprendizaje interno de organizaciones, instituciones y empresas-, los enfoques de explotación de datos pueden clasificarse en dos grandes grupos.

Bien sea orientados externamente, con una perspectiva que dirige su mirada hacia el cliente: focalizados en la perspectiva de mercado y de captación y fidelización de clientes sea a través de:



- la obtención de nuevas vías y fórmulas de ingresos: la definición y caracterización de nuevos mercados, nuevos clientes, nuevos márgenes de rentabilidad y definición de valor sobre clientes ya existentes, nuevas oportunidades de productos y servicios.
- la fidelización de los clientes: en un mundo caracterizado por la competitividad globalizada, la retención de los clientes actuales resulta de enorme interés estratégico para las empresas, más económica y eficiente que la captación de nuevos/as. No sólo eso: a mayor ciclo de vida de permanencia de cliente mayores probabilidades de obtención de rentabilidad con mecanismos como la venta cruzada de nuevos productos o la adquisición de nuevos productos o servicios.

O bien a través de una perspectiva orientada hacia el interior de empresas y organizaciones: procesos de aprendizaje e innovación focalizados en la mejora de los procesos de una compañía, organización o institución y que inciden en la optimización de costes mediante la detección de puntos de mejora y los factores que generan las desviaciones, o en la redefinición y *eficientización* de los procesos productivos -la base misma de la Industria 4.0-.

Organizaciones, empresas e instituciones que definen su actividad en términos de "*data-driven business*", y en las que el contraste de hipótesis delimita el campo de actuación de la toma de decisiones y de la gestión estratégica, táctica y operativa a través de procesos de Transformación -tal vez sería preciso definirlos en algunos sectores como procesos de *reconversión*- Digital. Los datos, en definitiva, viran las estrategias de compañías, organizaciones e instituciones -¿deberíamos afirmar, tal vez, con mayor precisión, que deberían hacerlo, por aquello tan básico de relacionarse con la realidad con la que interactúan?- comenzando por su localización en el centro de cualquier enfoque que adopten.

3. MODELOS ANALÍTICOS, ALGORITMOS Y REPRESENTACIONES DE LA REALIDAD: NUEVAS BRÚJULAS EN LA COMPLEJIDAD

Durante años, las herramientas de *Business Intelligence* estuvieron centradas en el *reporting*, en *deconstruir* el pasado. Hoy en día, los modelos analíticos lo están en la predicción y en la prescripción. Es decir, en adelantarnos a lo que pudiera ocurrir y ayudar a las organizaciones, empresas e instituciones a tomar decisiones orientadas con el mejor rumbo. Estos modelos adquieren el andamiaje analítico de lo que conocemos como algoritmos: una serie de conceptualizaciones abstractas para transformar y elucidar datos, modelos de expresión y representación del objeto de análisis delimitado. Brújulas entre la complejidad.

Sin embargo, estas reglas no constituyen una ciencia exacta, si es que esta fuera epistemológicamente posible. Tal y como apuntaba Kenneth Arrow “el buen pronóstico no es el que te dice que lloverá, sino el que informa acerca de sus probabilidades”⁹. Es decir, nada es seguro hasta que ocurre. La probabilidad cero no existe. Aprender a gestionar situaciones en este entorno, en este hábitat constitutivamente incierto, asignando probabilidades a las diferentes alternativas que puede tomar un determinado suceso, deviene crítico. Como en la filosofía futbolística de Marcelo Bielsa, la reducción de la incertidumbre se revela esencial, el frontispicio de la estrategia “data-driven”, su eje conceptual.

Estos algoritmos predictivos y prescriptivos modelan, representan, una realidad *ponderada* por la probabilidad de ocurrencia de un hecho concreto, en el sentido que todos concedemos a la probabilidad de acontecimiento

de un suceso específico. Es el principio que ha resultado predominante a lo largo del siglo XX, con Ronald A. Fisher a la cabeza, a través de la interpretación frecuentista¹⁰ de probabilidad. Un segundo enfoque, aún más antiguo, ha obtenido -pese a ello- mucha popularidad en los últimos tiempos. Se trata de un enfoque de base subjetiva, en el que cada persona tiene una percepción o expectativa de mayor o menor probabilidad. Este enfoque resultó preeminente en el siglo XIX, y de esta subjetividad en la interpretación de la probabilidad surge el teorema de Bayes¹¹. Dado que en muchas ocasiones para efectuar una predicción se posee un conocimiento limitado, la probabilidad es la expresión matemática

de ese nivel de conocimiento y, paralelamente, de su desconocimiento o grado de incertidumbre. Es decir, que “no se puede predecir con un 50% de probabilidades que saldrá cara” al lanzar una moneda, sino que “basándose en el conocimiento que se tiene, hay un 50% de certeza de que saldrá cara”.

El auge de los métodos bayesianos no sólo incorpora recursos analíticos y modelos, sino una nueva manera de afrontar predicciones y situaciones expresada en que a cada nueva capa

de información recibida cabrá una nueva probabilidad estimada de ocurrencia del fenómeno-hecho objeto de estudio o análisis. A medida que se vaya obteniendo nueva información, se mejoran las probabilidades iniciales, el ajuste y bondad estadística del modelo trazado. En síntesis: a más información, con mejor grado de probabilidad se puede estimar. Con menor grado de incertidumbre.

**Nada es seguro hasta que ocurre.
La probabilidad cero no existe.
Aprender a gestionar situaciones
en este entorno, en este hábitat
constitutivamente incierto,
asignando probabilidades a las
diferentes alternativas que puede
tomar un determinado suceso,
deviene crítico**

9. Arrow, K., “Planning and Uncertainty” International Journal of Development Planning Literature, 2.2. 1987.

10. Fisher, R. *Statistical method and scientific inference*, páginas 64-66. 1957.

11. Laplace, Simon, P., *A philosophical essay on probabilities, translated from the 6th French edition by Frederick Wilson Truscott and Frederick Lincoln Emory*. 1951.

A través de estos algoritmos predictivos y prescriptivos expresados bajo incertidumbre -siempre, en un grado u otro-, se podrán procesar los datos de una organización, empresa o institución para la obtención de valor de los mismos. Un tratamiento que se puede orientar bajo dos culturas, siguiendo la terminología introducida por Leo Breiman¹²: la cultura del modelado estadístico (regresión, reglas de asociación, Análisis de Componentes Principales) y la del modelado algorítmico e inteligencia artificial (redes neuronales, KNN). Más allá de disputas profesionales entre enfoques y paradigmas, el alcance de los modelos de explotación avanzada de datos residió en el desarrollo de una cultura algorítmica dotada de un potencial predictivo mucho más poderoso y preciso. Gracias a la cual pueden configurarse modelos generativos que iluminan la estructura inadvertida de los datos significativamente más allá que los producidos a través de los esquemas tradicionales -por ejemplo, a través de redes neuronales-.

En definitiva, comprender a bases de *clientes* masivas -trasládenlo al colectivo o agrupación humana que consideren-, a los consumidores y el “*por qué*” de sus decisiones. Detectar elementos que hacen de cuello de botella en una cadena de producción, abastecer de manera óptima la cadena de suministro u optimizar rutas logísticas para evitar ineficiencias en la gestión del tiempo, son algunos ejemplos de lo que estos algoritmos permiten representar y ajustar. Algoritmos cada vez más sofisticados, sobre los que pivota la transformación digital de muchas industrias. Millones de líneas de código, frente a las sesenta millones de líneas que definen Facebook y las cinco millones de líneas que componen la

programación del colisionador de hadrones. Nos hallamos, en suma, ubicados en el umbral mismo de los límites biológicos de la complejidad. En su mismo núcleo.

4. LOS ASPECTOS ÉTICOS DEL DATO Y EL BIG DATA EN EUSKADI

Las posibilidades y potencialidades del Big Data generan en la sociedad efectos contradictorios que oscilan entre las mistificaciones vacuas y el recelo ante un Gran Hermano¹³ omnisciente que pudiera alcanzar a conocer hasta el detalle más íntimo y fútil de nuestra intimidad, con la misma supuesta facilidad aparente con la que Amazon

-apenas con una decena de consultas en su web- es capaz de trazar su oferta personalizada basándose en nuestro historial de búsquedas. El acceso a nuestros datos, desde nuestros hábitos de consumo, situación financiera o identidad digital, constituye la condición de posibilidad de estas fases primigenias, pero desarrolladas, de la economía digital. Pero al respecto, la regulación de los niveles de accesibilidad a los datos por parte de empresas, organizaciones e instituciones

-desde el punto de vista tecnológico, sin lugar a dudas, pero también desde su dimensión jurídica- devendrá clave en su desarrollo en los próximos años, circulando en paralelo el desarrollo de estos ámbitos de actividad económica y los derechos de los usuarios, con la exigida anonimización de la explotación y uso de nuestros datos, la evitación de su transferencia a *terceros* y su regulación garantista como ejes. Y, por nuestra parte, como ciudadanos, teniendo siempre presente la importancia de la responsabilidad que adquirimos a través de nuestras autorizaciones para la cesión de datos

En pleno contexto de globalización económica e incertidumbre -profundizado por una crisis económica prolongada-, empresas -desde las de mayor tamaño y estructura hasta nuevos tejidos de actividad y startups-, instituciones y organizaciones vascas desarrollan programas y líneas de investigación en torno al framework Big Data con el objetivo de elevar los umbrales de competitividad de nuestra economía

12. Breiman, L., “Statistical modeling: The two cultures (with comments and a rejoinder by the author)”, *Statistical Science*, 16.3, páginas 199-231. 2001.

13. Huxley, A. 1984. Salvat Editores. 1970.

-frecuentemente inadvertidas, como cuando instalamos una aplicación en nuestros smartphones- en plena era digital en la que la creación, edición y distribución de contenidos está a la altura de un elemental click.

Mientras, en pleno contexto de globalización económica e incertidumbre -profundizado por una crisis económica prolongada-, empresas -desde las de mayor tamaño y estructura hasta nuevos tejidos de actividad y *startups*-, instituciones y organizaciones vascas desarrollan programas de innovación y líneas de investigación¹⁴ en torno al *framework* Big Data con el objetivo de elevar los umbrales de competitividad de nuestra economía. Con el fin de innovar para transferir a empresas y organizaciones las herramientas y los conocimientos necesarios para insertar en sus núcleos de actividad la cultura y conocimientos

14. Tales, entre otras, como Big Data by Tecnalía (<http://bigdata-by-tecnalia.com>), el Máster en Ingeniería Computacional y Sistemas Inteligentes de la UPV/EHU (<http://www.ehu.eus/es/web/kisa/aurkezpena>) o el Programa en Big Data y Business Intelligence de la Universidad de Deusto (<http://www.deusto.es/cs/Satellite/deusto/es/masteres/estudios-masteres/programa-en-big-data-y-business-intelligence/programa>)

derivados del dato. Para continuar creciendo en entornos económicos complejos y cambiantes que no nos esperan.

Para evitar lo que le ocurrió a aquel hombre que deambulaba de noche, agachado, casi en cuclillas, alumbrado apenas por una débil farola en lo que semejaba ser una búsqueda desesperada: “he perdido mis llaves”, respondió ante la pregunta de un policía que salió a su paso. “¿Las ha perdido por aquí?”, apuntó el agente. “No, pero es aquí donde dispongo de luz para buscarlas”, respondió el atribulado buscador¹⁵. Puede que en eso consista todo: en que no debemos, ni podemos, seguir buscando las llaves así. Ni darlas por perdidas tan sólo porque nuestras farolas actuales no tengan más alcance. Nuestro carácter, al fin y al cabo, nunca ha sido ese. Y alguna parte de nuestro futuro dependerá, además, de ello.

15. Pasquale, F, *The Black Box Society. The Secret Algorithms That Control Money and Information*, página 53, London, 2015.



IKER SERRANO ROBLES. PAMPLONA-IRUÑA 1982

Es licenciado en Bellas Artes por la Universidad del País Vasco donde actualmente está realizando su tesis doctoral en el Departamento de Pintura. Artista, ilustrador, profesor de dibujo y pintura en la Kultur Etxea de Basauri y coordinador de exposiciones de artistas emergentes en la Torre de Ariz de Basauri. Ha expuesto en numerosos espacios a nivel nacional y local: “Lur zati bat” Galería la Taller, Bilbao; FIG, Bilbao; Ediciones VuelaPluma, Madrid; “Capital Animal”; La casa encendida, Madrid (2015); “El guardián del Bosque”, Galería Utopia Parkway, Madrid; Sala Rekalde, Bilbao (2014); Viaje isla 2.1. Galería Windsor Kulturgintza, Bilbao (2012) etc. Al mismo tiempo, ha obtenido numerosas becas y ayudas para su formación: beca predoctoral del Gobierno Vasco, beca de producción de la Diputación Foral de Bizkaia, beca residente Fundación BilbaoArte etc.

www.ikerserrano.com

