



# A che servono questi famosi big data?

Sono le "tracce" che lasciamo quando andiamo al supermercato, prenotiamo il cinema, mettiamo un commento sui social. Vengono raccolte in enormi archivi digitali e poi analizzate. Per offrire vacanze su misura o prevedere il risultato delle elezioni

di **Giampaolo Cerri**



**Z**ettabyte: se vuoi capire che cosa sono i Big Data e come ci cambieranno la vita, tieni a mente questa bizzarra parola. È una unità di misura della quantità di informazioni che produciamo quando facciamo shopping, navighiamo in Rete, pubblichiamo una foto su Facebook. A quanto ammonta? A 10 elevato alla 21esima potenza. Per fartene un'idea, pensa ai dati che può contenere il disco esterno dove hai salvato il video del primo giorno di scuola di tuo figlio o di un'ecografia... moltiplicato per un miliardo. Di Big Data si parla da quando le nostre vite si sono digitalizzate, e ogni azione genera informazioni custodite da qualche computer. L'epistemologo Luciano Floridi, che insegna a Oxford, ha stimato i dati creati sin qui e quelli che si genereranno nei prossimi decenni: «Dai graffiti preistorici nelle caverne fino al 2013 l'uomo ha prodotto 4,4 Zettabyte di dati. Nel 2012 ne sono stati generati più che nei precedenti 5.000 anni. Dal 2013 al 2020 produrremo 44 Zettabyte di informazioni digitali». C'è poi un'altra differenza rispetto al passato. «Una volta, per raccogliere e conservare una parte infinitamente più piccola di dati occorrevano calcolatori enormi» osserva Francesco Sacco, docente dell'università Bocconi di Milano, esperto di economia digitale. «Oggi invece tutta la musica prodotta sin qui potrebbe stare in un hard disk che occupa 4 scaffa-



## PERCHÉ LEGGERE QUESTO ARTICOLO

Per fare chiarezza sui Big Data, di cui sentiamo parlare sempre più spesso.

→ L'analisi dei social ha funzionato meglio dei campioni demoscopici per pronosticare la vittoria di Trump alle elezioni Usa.

→ L'Ue vuole rilanciare l'industria aerospaziale europea. Perciò rende disponibili alle aziende i dati raccolti da 3 programmi (Egnos, Copernicus e Galileo) che usano le informazioni provenienti dalla osservazione della Terra dallo spazio.



li di una normale libreria. Intanto, la potenza di calcolo dei computer è cresciuta enormemente e siamo in grado di estrarre e dare cioè un senso, a miliardi di informazioni».

**Studiano le malattie.** «Negli Stati Uniti analizzando milioni di tweet e post di Facebook si anticipa di almeno 15 giorni l'arrivo dell'influenza, con la possibilità, quindi, di ridurre i danni» spiega Riccardo Bellazzi, ordinario di Bioingegneria all'università di Pavia. Come? Basta usare un buon algoritmo (uno dei cosiddetti algoritmi predittivi, capaci cioè di "leggere" cosa accadrà) per decifrare schiere infinite di parole e individuare la sintomatologia tipica dell'attacco influenzale. Con il reparto di Endocrinologia della clinica Maugeri nell'ambito del progetto europeo Mosaic, Bellazzi ha immesso migliaia di informazioni di pazien-





ti diabetici, da quelli delle visite a quelle dei farmaci assunti, e costruito un programma che permette ai medici di osservarli statisticamente nel corso tempo. «Studiando l'evoluzione di malati con caratteristiche simili» continua Bellazzi «è possibile rilevare quali altre patologie, secondarie ma ugualmente gravi, possono insorgere e quando. In questo modo i diabetologi hanno a disposizione strumenti per decidere, per esempio, di anticipare screening o di essere più aggressivi con le terapie sui nuovi pazienti».

**Cambiano il modo di comprare.** Il turismo è un settore emblematico dell'utilizzo dei Big Data. A settembre, la società Almax ha annunciato di aver raccolto per 45 giorni qualcosa come 21 milioni di messaggi Twitter e Instagram, in italiano e in inglese, aventi come argomento le vacan-

## Ci sono rischi per la privacy?

La gestione dei Big Data presenta 2 punti critici. «Il primo riguarda la "profilazione": i dati che generiamo creano una specie di nostra seconda identità su cui altri potrebbero prendere decisioni automatiche» dice Giovanni Ziccardi, docente di Informatica giuridica alla Statale di Milano. «Una compagnia assicurativa può negarci una copertura, perché dal nostro profilo risulta che avremo un'alta probabilità di ammalarci. O un'azienda può non assumerci perché, dai dati, emerge che abbiamo un certo orientamento politico. Perciò sono cruciali regole che stabiliscano chi raccoglie i dati, come lo fa e con quale finalità. Ed è decisivo che ciò avvenga con il nostro consenso». C'è un secondo nodo da sciogliere: «La sicurezza dei metodi di conservazione delle informazioni, come l'uso di sistemi di cifratura» dice Ziccardi. «Yahoo ha appena annunciato un colossale furto di dati dei propri utenti. Pensiamo se fossero stati milioni di cartelle cliniche...».

ze in Italia. Club Med, invece, ha annunciato l'arrivo per il 2017 di una piattaforma digitale attraverso la quale raccogliere e vagliare i dati dei potenziali turisti per offrire pacchetti sempre più mirati. L'Italia sarà il Paese pilota. Viene così rovesciata la prospettiva: dal cliente che sceglie il viaggio al viaggio che gli viene offerto in base ai suoi gusti, ossia la vacanza personalizzata.

### Permettono di mappare l'opinione pubblica.

Analizzare i dati, però, non è soltanto una questione di algoritmi. «L'apporto umano è decisivo» chiarisce Luigi Curini, professore di Scienza politica all'università Statale di Milano che, con i colleghi Andrea Ceron e Stefano Iacus, ha fondato nel 2012 Voices of the blogs, società che analizza le informazioni generate dai social network. I tre professori hanno inventato un algoritmo chiamato Isa, sigla che sta per Integrate sentiment analysis, ossia l'analisi integrata dell'opinione diffusa. Isa viene continuamente istruito da alcuni operatori specializzati: «I nostri analisti setacciano i social e distinguono gli interventi pertinenti a una certa ricerca da quelli che non lo sono. Ne bastano alcune migliaia, poi l'algoritmo impara a scartare ciò che non c'entra con l'argomento di cui ci si sta occupando». Durante le presidenziali americane, il team milanese ha analizzato quotidianamente fino a 10 milioni di messaggi dei social, arrivando a indicare il sorpasso di Donald Trump su Hillary Clinton diversi giorni prima dei sondaggi tradizionali.

**Aiutano a trovare lavoro.** «Noi lavoriamo con lo stesso metodo per le imprese» racconta Curini. «Per un'azienda italiana che produce assorbenti femminili biodegradabili siamo riusciti a individuare nella California il mercato ideale, fino a dare indicazioni sul packaging giusto per le consumatrici. Quell'azienda ora esporta in quel mercato». Gli fa eco il professor Sacco: «Ricordo il caso di una piccola impresa italiana che, analizzando i dati, è riuscita a capire come il suo olio d'oliva di grande qualità potesse avere in Giappone un mercato floridissimo. Come? Oltre 89.000 interventi sui social in pochi mesi parlavano di extravergine. E l'export è cominciato». È facile dedurre, quindi, che quella del data analyst sia la professione del futuro. Da una ricerca Adecco, big delle risorse umane, su 347 aziende di varie dimensioni, da Sud a Nord, emerge che il 97% dei direttori del personale ha chiara la percezione dell'importanza di questi professionisti. Ma solo nel 20,27% di quelle stesse imprese esistono progetti sui Big Data. Un gap notevole, che dovrà essere colmato in fretta.

