



Cosa sono i Big Data? Molto più che una questione tecnologica

di **Marco Fattore**

Dipartimento di Statistica e Metodi
quantitativi, Università di Milano-Bicocca

Che cosa sono i Big Data? Il tema è certo anche tecnologico, ma rispondere a questa domanda significa soprattutto affacciarsi alle sfide culturali del prossimo futuro e aprirsi a una serie di riflessioni che toccano aspetti educativi, sociali, economici e politici.

Tecnicamente, l'espressione "Big Data" indica insiemi di dati così ampi da renderne impossibile l'archiviazione e il trattamento attraverso i comuni strumenti hardware e software, e tipicamente generati da processi digitali non statisticamente controllati (cioè non costruiti secondo i criteri scientifici adottati per i censimenti e per le indagini statistiche "mirate"). Si pensi alla quantità di dati generati dagli smartphone, dai dispositivi "intelligenti", dai social network, dalle transazioni del commercio elettronico, dalle ricerche sul web, dalle immagini caricate in rete... Dati numerici, testuali, sonori, visuali, a volte strutturati (cioè organizzati secondo schemi formalizzati, come tabelle con righe e colonne definite), molto spesso non strutturati (come i testi e le immagini). Gli specialisti usano caratterizzarli con tre "V": Volume (quantità di dati), Varietà (dati di natura eterogenea e provenienti da fonti differenti) e Velocità (flussi continui di dati, spesso acquisiti in tempo reale).

La vera natura dei Big Data, tuttavia, si chiarisce meglio guardando alla loro origine: essi derivano direttamente dall'esperienza quotidiana delle persone e dal funzionamento automatico dei sistemi digitali. Vivendo, agendo, lavorando, giocando e utilizzando "cose" noi, senza accorgercene, "emettiamo dati" e così fanno gli innumerevoli dispositivi tecnologici tra loro interconnessi che costituiscono l'architettura, spesso invisibile, in cui ci muoviamo e alla quale ci appoggiamo quotidianamente. I Big Data sono quindi il continuo e incalzante fluire di segnali digitali provenienti dalla vita personale, sociale ed economica. Ci sono palloni da calcio che raccolgono i dati sui tiri che facciamo, dispositivi che tracciano come guidiamo o che tipo di consumo energetico abbiamo; guardando un filmato su Youtube, lasciamo tracce del nostro passaggio e dei nostri interessi; a breve il tacco delle nostre scarpe ci dirà come sta la nostra schiena e l'Internet of Things farà esplodere la "risoluzione" dei dati, rendendoli sempre più in grado di rivelare tratti differenti e particolari del mondo.

Questo processo, che è già in atto, cosa comporta, quali scenari apre e che impatto avrà sulla persona, sulla società e sull'economia? A queste domande tentano di rispondere, almeno in parte, gli altri articoli del presente numero di *Atlantide*. Qui mi limito a suggerire alcune prospettive di riflessione, muovendomi liberamente fra diversi punti di vista, più per generare scenari che per fornire risposte.

Il primo aspetto che sottolineo è che una disponibilità così ampia di dati, direttamente o indirettamente legati al "fenomeno umano", apre spazi di osservazione e di conoscenza nuovi ed è in grado di modificare strutturalmente i processi decisionali, a livello personale e collettivo, inducendo comportamenti adattativi che modificano l'economia e la società, rendendole più dinamiche e complesse.

A chi non è successo, entrando in un negozio per fare acquisti, di usare lo smartphone per confrontare i prezzi e verificare i giudizi di altri utenti, così da "ribilanciare" le informazioni fornite dal venditore (al quale, a questo punto, non conviene più essere troppo parziale)?

Con tutti i limiti del caso, TripAdvisor induce comportamenti più trasparenti in chi eroga servizi legati al turismo e la condivisione di informazioni genera processi di valutazione che possono valorizzare la qualità, indurre processi di emulazione e di adattamento. La diffusione dell'informazione genera, inoltre, dinamiche di auto-organizzazione e fenomeni collettivi, che possono poi avere effetti sistemici, come dimostra il crollo del valore delle licenze dei taxi a New York, dettato dalla presenza di Uber (che di fatto non fa altro che mettere in circolo informazioni) o l'apertura di nuovi mercati, come quello generato da Airbnb, il portale per l'affitto di abitazioni.

L'economia e la società si stanno "complessificando" e in un ambiente complesso la capacità di controllare e prevedere cede il posto alla capacità di adattarsi ai mutamenti, in parte assecondandoli o valorizzandoli e in parte governandoli. Questa attitudine, a sua volta, richiede una capacità di vedere e percepire i cambiamenti con tempestività e granularità; un uso intelligente dei Big Data è in grado di sostenere questi processi, non tanto per la possibilità di misurare con precisione i fenomeni (cosa spesso impossibile e concettualmente inappropriata, quando i fenomeni riguardano l'umano), quanto per la capacità di cogliere segnali strutturali e agire in costante dialogo con la realtà.

Per esempio, mentre i dati ufficiali hanno segnalato l'impatto della crisi economica sull'occupazione con mesi di ritardo, le informazioni amministrative generate dalle comunicazioni che i datori di lavoro devono conferire al Ministero in occasione di assunzioni, cessazioni o trasformazioni di rapporti di lavoro, hanno raccontato in tempo quasi reale il "congelamento" del mercato del lavoro e la sua progressiva ristrutturazione e precarizzazione, rivelando molto più di un semplice dato aggregato. Quanto la politica avrebbe potuto avvantaggiarsi di questa possibilità di "vedere"?

In questo senso, si aprono spazi interessanti per il cosiddetto "policy-making adattativo", nel quale la capacità di osservare i fenomeni tempestivamente e con più dettaglio e la possibilità di valutare gli impatti delle azioni messe in campo si traducono in politiche dinamiche e più aderenti alla realtà. I dati generati dalle automobili possono essere usati per studiare la mobilità, i comportamenti di guida, l'inquinamento veicolare e così impostare politiche, regolamenti e incentivazioni, testarne gli effetti ed eventualmente estenderne la validità, monitorare il sistema...

Le nostre case e le nostre città si riempiranno a breve di dispositivi domotici in grado di rilevare i consumi, gli sprechi, la temperatura, l'umidità, la rumorosità... tutte informazioni che consentiranno di impostare politiche energetiche, valutare il grado di "educazione" energetica del territorio, individuare aree territoriali più o meno disagiate per condizioni abitative inadeguate... Per non parlare della medicina e del monitoraggio della salute. Gli esempi potrebbero essere infiniti.

Le stesse opportunità si aprono per i processi sussidiari, in cui l'adattamento dei corpi sociali alle esigenze che si pongono sul territorio si potenzia grazie alla possibilità di "in-formarsi", cioè di conoscere e prendere la forma del bisogno emergente, in modo tempestivo. Oltretutto, i dati permetterebbero di oggettivare il valore generato in modo sussidiario, consentendo al livello istituzionale di mantenere le leve di governo sistemico, in un corretto rapporto e in un giusto equilibrio con la presenza operativa e qualificante dei corpi intermedi.

Qual è l'impatto della carità sull'economia e sulla società? Nessuno oggi riesce a rispondere a questa domanda in modo oggettivo e di fatto la questione non è certamente al centro del pensiero economico. Ma se solo con qualche dispositivo domotico potessimo stimare in tempo reale il numero di persone che ogni giorno vanno alle mense dei poveri e se potessimo oggettivare, e quindi renderci conto, dell'imponenza di quel che avviene ogni giorno sul territorio e della capacità dei soggetti di rispondere ai bisogni che cambiano, quanto muterebbe la percezione del valore della sussidiarietà nella società, nell'accademia e nella politica?

L'impatto dei Big Data sull'economia privata e sul mondo del lavoro sarà altrettanto significativo. Le aziende hanno bisogno di informazioni per governarsi, per conoscere il proprio contesto, il comportamento dei propri clienti... Nascono nuovi servizi, basati sulla trasformazione di dati in informazioni, nuove "app"; cresce il bisogno di "ingegneri dei processi informazionali", strano profilo caratterizzato da un mix di competenze statistiche, informatiche e ingegneristiche, che va sotto il nome di "data scientist". Si aprono mercati nuovi, si creano professioni nuove e si rinnovano quelle classiche.

Dunque cresce l'economia "informazionale" (cioè basata sulla generazione, trasmissione, fruizione, e interpretazione di informazioni) e la capacità di generare conoscenza diviene sempre più cruciale. Ma i Big Data sono solo frammenti (bit,

appuntamento); chi li trasforma e li "condensa" in "informazioni significative"? E chi interconnette le "informazioni significative" in conoscenza? Recentemente un'azienda attiva nell'ambito della domotica mi ha contattato per farsi aiutare a "immaginare" che cosa fare con i dati a disposizione; non mi ha chiesto che algoritmo usare o che tecnica applicare; mi ha chiesto di "immaginare" e di generare "ipotesi". E non è la prima volta che questo mi accade. Chi mette assieme i frammenti in modo ragionevole, per rispondere a domande aperte e a bisogni non codificati?

Qui si apre il secondo tema, quello educativo. Il mondo dei Big Data ha bisogno di soggetti capaci di senso, capaci di nessi e con una tensione sintetica, che non confonda lo zapping tra i dati con il percorso conoscitivo, la quantità di micro-informazioni con il valore di un giudizio di sintesi, la velocità di fruizione con i tempi di un pensiero solido e ragionevole. È questa la personalità che sarà capace di trasformare un mondo di frammenti in "cose" dotate di valore, a qualunque livello essa si trovi ad agire.

Purtroppo, non è questa la struttura di personalità che il mondo digitale tende a generare, soprattutto nei giovani "millennials". La cultura digitale e le modalità di accesso e di fruizione dei relativi contenuti tendono a de-strutturare i processi di conoscenza e a polverizzarli in atti staccati, in "happy-hour a base di noccioline informative", dove il numero delle scelte possibili (in fondo, spesso equivalenti) chiude e fa muovere in spazi pre-immaginati e pre-costituiti, che chiedono più abilità che creatività. I nativi digitali sono fruitori eccezionali delle forme di comunicazione odierne, e sono padroni dell'uso delle opzioni messe a disposizione, ma generalmente sono del tutto inconsapevoli delle strutture concettuali, e quindi dei limiti, di questi strumenti; pensando di dominare la tecnologia nel senso dell'uso, ne sono di fatto dominati per quanto riguarda l'orizzonte di senso (paradossalmente la mia generazione, che ha cominciato ad avere in casa i primi home computer, è più consapevole della natura di quella "ferraglia" e ha un rapporto con la tecnologia molto meno di dipendenza simbiotica). E così occorre mettersi alla ricerca di chi è capace di "immaginare" e fare "nessi", perché questa è la capacità che sta venendo meno. Inoltre, l'enfasi sui sistemi intelligenti (Machine Learning, Deep Learning, Intelligenza Artificiale...), portando agli estremi la sequenza "dati-computazione-ottimizzazione", suggerisce che l'intelligenza sia primariamente connessa alla "capacità analitica di calcolo" rispetto a un obiettivo fisso, misurabile e chiuso, piuttosto che alla capacità di sintesi e di domanda in un percorso di ricerca di significati, in campo aperto. Essere immersi in una cultura che esaspera gli aspetti cognitivi di immediatezza e facilità di accesso, rimuovendo ogni ostacolo che si frappone tra il "client" e il "server", disabituata alla fatica della conoscenza di ciò che della realtà non può essere ridotto a bit e deprime l'esperienza di attrazione affettiva, ma anche di alterità irriducibile, che ogni rapporto "onesto" con la "cosa" implica. E così facendo riduce la potenza conoscitiva sintetica della persona, proprio quando questa sarebbe più necessaria.

Qui è inevitabile un passaggio sulla scuola. Se i percorsi formativi devono far crescere le competenze tecniche legate, direttamente o indirettamente al "mondo 4.0" (non si può stare al mondo senza possedere i linguaggi del proprio tempo), è indispensabile che si rafforzino le discipline che insegnano ad argomentare indirettamente, a "mettere insieme i pezzi", a ricercare l'unità concettuale dai frammenti che si hanno a disposizione (per non dare l'impressione di fare un discorso "etereo", dico che questo è forse il modo migliore per descrivere il mio lavoro, quando parto da "un mucchio di dati" e devo ricavarne qualcosa per cui il committente è disposto a investire le sue risorse). Piaccia o non piaccia ai tecnoburocrati della scuola, queste attitudini si sviluppano attraverso le materie umanistiche. In effetti, la difficoltà dei miei studenti a mettersi davanti a un insieme complesso di dati, e ricavarne una visione coerente del fenomeno sottostante, non è dovuta a una carenza di competenze tecniche, ma a una disabitudine a interconnettere segni, come quando si deve interpretare o ricostruire un testo.

Il tutto è reso ancora più urgente dal fatto che i grandi generatori/utilizzatori di dati temono la capacità di senso e di pensiero autonomo che non rende prevedibili i soggetti. Perché i Big Data sono anche uno strumento di controllo e di misurazione e, in una cultura dell'"ottimizzazione", si aprono scenari ambigui, soprattutto quando l'informazione è, come in molti casi, asimmetrica. Se infatti la disponibilità crescente di dati può ridurre l'asimmetria informativa e generare processi virtuosi, è anche vero che in molti casi i dati stessi, e la capacità di trattarli e di estrarne informazione, sono appannaggio esclusivo dei produttori di beni, di servizi o di "opinioni".

In realtà, questa è una situazione classica, non legata ai Big Data; il problema è che oggi le conseguenze di questa asimmetria possono essere molto più pervasive, a livello individuale, e sistemiche, a livello sociale, perché la tecnologia digitale permette di agire su livelli più profondi e su scale più ampie. Oggi esistono sistemi di Machine Learning in grado di individuare i soggetti più sensibili a specifiche comunicazioni commerciali via web e quindi capaci di "pressare" gli individui, inducendoli a comportamenti anche autodistruttivi, come nel caso del gioco d'azzardo on-line.

Con dati e algoritmi adeguati, un'assicurazione, una banca, una clinica o qualche altro agente "ad alto impatto sociale" può perseguire a tal punto obiettivi di ottimizzazione economica, da generare, direttamente o indirettamente, processi discriminatori verso particolari classi di soggetti. Non mi sto riferendo a comportamenti di per sé illeciti (peraltro possibili), ma al possibile effetto sistemico di comportamenti leciti, ma puramente strumentali. Possiamo ancora affidarci alla "mano invisibile" o i possibili effetti sono troppo rischiosi per non porci un problema politico? Con i dati e gli algoritmi, infatti, oggi si fa cambiamento economico e sociale e quindi si fa anche politica.

Vengo così all'ultimo passaggio. L'uso dei dati pone anche problemi legali, di regole, di tutela della privacy, di diritto all'oblio... Ma più profondamente, se i processi sociali, economici e politici si svilupperanno sempre più in questo "spazio informazionale" fatto di dati, dispositivi interconnessi, algoritmi, misure, comunicazione, allora chi governa questo spazio? Di chi è questo spazio, che è totalmente trasversale ai confini, alle istituzioni e alle giurisdizioni statuali?

Prendiamo l'esempio di Google. Quest'azienda è riuscita a costruire un sistema per cui la maggior parte delle azioni che effettuiamo nel web "visibile" (cioè escludendo il cosiddetto "deep web") non solo sono svolte usando strumenti da essa messi a disposizione, ma sono percettivamente svolte nel suo alveo, al punto tale che ormai l'espressione "googolare" è divenuta sinonimo di "andare a cercare nel web" e che l'insieme dei servizi "intelligenti" resi disponibili da Google (dai calendari sincronizzati, all'archiviazione e condivisione di documenti, da gmail alle mappe geografiche e all'espressione "Ok Google, guidami a...") costituiscono l'ambiente in cui ci muoviamo (ed esistiamo? siamo ormai forse sempre più la "stirpe di Google"?) e senza i quali sarebbe ormai difficile agire. E intanto Google acquisisce dati, elabora dati, usa dati, prevede, ottimizza, ci fa trovare quello che cerchiamo (ma in base a quali criteri?) ci dice che cosa comperare, ci dà le notizie, anticipa i nostri bisogni... un perfetto sistema di welfare informazionale che guadagna indefinitamente.

Ma se Google interrompesse i suoi servizi? Avete in mente l'ultima scena di *Fuga da Los Angeles*, quando Jena Plissken "spegne" il mondo con un click? O se decidesse di rendere visibili certi contenuti o certe informazioni in alcuni Stati e non in altri? O di veicolare certe campagne di opinione? Che impatto può avere sulle dinamiche sociali ed economiche? E che relazione ci potrà essere tra il governo di uno Stato, nel contesto di processi democratici e di rappresentanza, e un potere privato di fatto sovranazionale di questo tipo, che risponde a logiche indipendenti dai processi di autodeterminazione nazionali?

No, il tema dei Big Data non è solo tecnologico.