Come creare valore con l'economia dei dati

*di Paola Cillo, Gaia Rubera**

CONTESTO DI RIFERIMENTO

Si dice che viviamo nella cosidetta «data-driven economy», vale a dire un'economia in cui i dati sono la risorsa fondamentale per generare vantaggio competitivo. Per capire come le imprese possano creare valore in questo scenario economico è importante capirne i due fattori fondamentali – i big data e il machine learning – sfatando alcuni dei luoghi comuni che limitano la loro comprensione da parte delle imprese e quindi la possibilità di estrarne valore.

Nella prima metà di questo decennio, il termine «big data» è diventato popolare nel mondo manageriale. Questo venne coniato per la prima volta nel 1997 da due ricercatori della NASA, Michael Cox e David Ellsworth. I due, sopraffatti dalla difficoltà di gestire una mole sempre crescente di dati generati dai loro studi, scrissero: «Insieme di dati generalmente molto grandi, tali da mettere a dura prova le capacità della memoria, il disco fisso e anche del disco locale [di un computer]. Chiamiamo questo il problema dei big data»[1].

In ambito manageriale, nel 2011 il McKinsey Global Institute definì i big data come: «Dataset il cui volume è talmente grande da eccedere le capacità dei software di catturare, immagazzinare, gestire e analizzare»[2].

CHE COSA SONO I BIG DATA E PERCHÉ SONO COSÌ IMPORTANTI

Alla luce di queste definizioni, si è quindi spesso portati a identificare la caratteristica fondamentale dei big data nella loro quantità – «big», appunto – cui si accompagna un problema prettamente tecnologico, vale a dire la difficoltà dei nostri computer nel gestire tale mole di dati. Tuttavia, questo problema tecnologico non è certamente nuovo, anzi potremmo dire che è vecchio tanto quanto l'invenzione del computer, come certamente converrà chiunque abbia visto il film The Imitation Game. Per questo motivo, ridurre i big data a un semplice problema tecnologico e di dimensione è quantomeno fuorviante. Cosa rende allora allora i big data così speciali?

Per capirlo, proviamo a guardare da vicino i dati che milioni di persone producono ogni giorno su Spotify. Molti di noi credono, a ragione, che Spotify abbia una conoscenza molto approfondita dei nostri gusti musicali. In realtà Spotify sa molto di più: sa, per esempio, quando corriamo, quando guidiamo, quando stiamo facendo la doccia. Come fa a saperlo? Nei server di Spotify è tutto un fiorire di playlist «Run, Michael, Run!» o «Driving Songs» o «Songs to Sign in the Shower». La maggior parte delle persone crea playlist col nome dell'attività in cui è impegnata nel momento stesso in cui l'ascolta. Sono gli utenti stessi a dire: «Sto correndo», «Sto guidando», «Sto facendo la doccia». Questo è il livello di conoscenza, minuziosamente dettagliata, sulle nostre attività quotidiane cui Spotify ha accesso (e su cui ha costruito gran parte del suo business model).

* ***Paola Cillo** è Professore Associato di Management & Tecnologia presso l'Università Bocconi e Direttore della Divisione Ricerca "Claudio Dematté" presso la SDA Bocconi. **Gaia Rubera** è Amplifon Chair in Customer Science, Direttore del Dipartimento di Marketing e Professore Ordinario di Marketing presso l'Università Bocconi di Milano.*

DOI: 10.1485/1120-5032-202102-3

Spotify rappresenta la norma della data-driven economy. Siamo circondati da oggetti che registrano qualsiasi attività facciamo e la trasformano in un dato, immagazzinato da qualche parte e pronto per essere utilizzato. Il vero vantaggio dei big data, quindi, non risiede tanto nella quantità, quanto nella capacità di fornire informazioni estremamente dettagliate su un individuo. È la prima volta nella storia dell'umanità, questo sì, che riusciamo non solo a tracciare in tempo reale, ma anche a immagazzinare attività, parole e stati d'animo di ogni individuo. È proprio questa capacità di fornirci informazioni minuziose sulla vita di miliardi di persone che rappresenta l'unicum dei big data. Ma cosa ha portato all'emergere dei big data? Perché adesso e non vent'anni fa?

Così come siamo soliti pensare che i big data siano collegati a un problema tecnologico, allo stesso modo tendiamo a credere che un cambiamento tecnologico abbia originato tali dati. Sentiamo spesso parlare di «digital economy» come di un fenomeno nuovo. A quanti eventi avete partecipato in cui si parlava di questo tema che veniva presentato come del trend del futuro? E quanti articoli avete letto dello stesso tenore? In realtà la digital economy è un fenomeno iniziato circa vent'anni fa, mentre i big data sono un fenomeno molto più recente.

Questo disallineamento tra i tempi dovrebbe rendere evidente come la digitalizzazione sia la causa necessaria, ma non sufficiente, per la proliferazione dei big data. Certo, è necessario che in qualche modo la playlist che ascolto venga trasformata in un byte, digitalizzato appunto, perché possa poi essere aggregata con altri miliardi di byte e dare vita ai big data. Eppure, finché la musica che ascolto rimane un dato all'interno delle mie mura domestiche nessuno potrà accedervi. Figuriamoci costruirci un'intera economia intorno. Perché i dati che creiamo ogni giorno diventino big data è necessario essere disposti a condividere tali dati. Qual è, dunque, il cambiamento fondamentale che ha reso possibile trasformare dei semplici byte in big data? A nostro parere, la risposta va trovata nei social network, un fenomeno sociologico che ha cambiato per sempre il nostro senso della privacy. È come se i social network ci avessero di colpo trasformati in essere sociali, non solo disposti, ma anche vogliosi di condividere con chiunque quello che facciamo. Non è un caso che, quasi in contemporanea con la crescita dei social network, sia stato coniato un nuovo termine per descrivere la nostra società: «Sharing economy». Oggi siamo disposti a condividere con perfetti estranei oggetti fino a poco tempo fa assolutamente privati, quasi intimi, come auto, case, moto, biciclette. Allo stesso modo, non abbiamo nessuna remora a condividere informazioni private che ci riguardano. È questo cambiamento sociologico il vero responsabile della proliferazione dei big data. La necessità di condividere la nostra vita con estranei ha fatto sì che quasi nessuno trovi strano che ogni passo, ogni nota ascoltata, ogni pensiero espresso, escano dalle pareti della propria casa per finire in super computer ai confini dell'Artico.

Ignorare tale cambiamento – culturale e non tecnologico – vuol dire avere una comprensione solo parziale del valore dei big data. Questi hanno rivoluzionato la nostra economia perché hanno invertito il processo di creazione del valore. In passato, era l'impresa che andava dal consumatore per capirne i bisogni tramite questionari o focus group. Potevamo ottenere scarne informazioni su un numero limitato di consumatori. Oggi sono i consumatori, nuovi animali sociali ansiosi di condividere, che forniscono direttamente o indirettamente alle imprese miriadi di informazioni su qualsiasi cosa facciano o pensino, ogni minuto della loro vita. E non i pochi consumatori che a fatica riuscivamo a intervistare, ma milioni di persone. Come possono le imprese usare questa nuova risorsa? La risposta si chiama «machine learning».

CREARE VALORE COI BIG DATA: MACHINE LEARNING

Intorno al 2017, le ricerche su Google per il termine «machine learning» diventano più numerose di quelle per il termine «big data», testimoniando come l'attenzione si sia spostata dalla risorsa grezza – i big data – all'uso di tale risorsa attraverso, appunto, il machine learning. Così come esiste una certa confusione sui big data, allo stesso modo il machine learning è sovente ammantato da un velo misterioso che porta molti manager a pensare che esso sia una sorta di sfera di cristallo in cui leggere il futuro. È davvero così? Per rispondere a questa domanda, dobbiamo capire meglio come funzionano questi modelli di machine learning.

Il processo è diviso in due fasi: addestramento e predizione. Nella fase di addestramento al computer vengono forniti due tipi di input: i dati e le risposte che vogliamo che la macchina impari, quelle che in termine tecnico vengono chiamate «label». Famoso è il caso di Target che nel 2012 sviluppò un modello per identificare donna incinte. Vediamo come. Nella fase di addestramento, ci focalizziamo su un gruppo di clienti, alcune delle quali sappiamo essere incinte (perché a distanza di qualche mese osserviamo che comprano pannolini o cibo per neonati) e altre che sappiamo non esserlo. Iniziamo a passare alla macchina i dati di acquisto di una cliente e chiediamo di dirci se è incinta oppure no. Forniamo alla macchina anche la label, in questo caso una variabile binaria: «sì» (se la cliente è incinta) oppure «no» (se non lo è).

La macchina crea un primo modello e fornisce una risposta (all'inizio totalmente casuale). Confronta poi questa risposta con quella corretta, cioè la label che abbiamo fornito. Se la risposta è giusta passa alla cliente successiva. Usando il modello appena sviluppato, anche questa volta fornisce una risposta e la confronta con la label. Se la risposta è sbagliata, modifica il modello per

poter fornire, la prossima volta, una risposta migliore. Per poter imparare, la macchina dovrà eseguire questo loop migliaia, in alcuni casi milioni, di volte. A ogni iterazione, la macchina aggiusta il proprio modello, fino a quando riuscirà a fornire la risposta esatta in un numero adeguato di casi (di solito, intorno al 70/80 per cento).

Al termine di questa fase, il modello di Target aveva imparato che le donne incinte hanno certe caratteristiche distintive, vale a dire acquistano 25 prodotti in misura maggiore rispetto alle altre clienti, quali cotton fioc, saponi naturali e vitamine. A questo punto, la fase di addestramento termina e il computer è pronto a usare il modello che ha appena imparato per analizzare i dati di nuovi clienti, che non sappiamo se siano incinte o meno. Si passa, quindi, alla cosiddetta fase di predizione (o inferenza), dove daremo al computer dati di nuovi clienti e chiederemo se sono incinte o meno. Come fa la macchina a rispondere («predire» in termini tecnici)? Usando il modello imparato durante l'addestramento.

Come si vede, il meccanismo tramite cui le macchine imparano è del tutto simile al modo in cui gli umani imparano. Per fortuna, gli umani imparano molto velocemente, mentre al momento le macchine hanno bisogno di essere addestrate su centinaia di migliaia di dati, che siano acquisti dei clienti, post su Facebook o immagini su Instagram. Questa esigenza spiega perché, seppur molti dei modelli di machine learning siano stati concettualizzati anni e a volte decenni fa, in ambito manageriale se ne sia iniziato a parlare solo recentemente. I big data sono la benzina che permette di mettere in moto questi modelli e usarli.

Dovrebbe essere chiaro, a questo punto, perché i big data sono considerati la risorsa fondamentale nella cosiddetta data-driven economy. Chi possiede tali big data, ed è in grado di analizzarli, ha un fondamentale vantaggio competitivo. Già, ma chi possiede i big data?

CHI POSSIEDE I BIG DATA?

Per rispondere a questa domanda presentiamo due casi: Target e Nathan Ruser.

Abbiamo già visto il caso Target, che nel 2012 sviluppò un modello di machine learning per identificare donne incinte. Questo caso è famoso perché è stato uno delle primissime volte in cui, in ambito manageriale, abbiamo visto i big data in azione. L'attività di Target è infatti un'operazione squisitamente di marketing, resa possibile dai big data. Analizzando i dati delle proprie carte fedeltà, Target si era accorta che le donne che hanno appena partorito sono un segmento particolarmente fedele di consumatrici (probabilmente perché non hanno molto tempo a disposizione per provare nuovi negozi). Da qui la decisione di fidelizzarle, mentre ancora incinte, e quindi l'invio di coupon promozionali.

Target ha potuto permettersi una tale operazione perché aveva in casa i big data necessari, che genera attraverso le carte fedeltà di milioni di clienti in tutti gli Stati Uniti. Allo stesso modo, imprese come Google, Amazon, Facebook generano big data immagazzinando i dati delle attività sui propri siti di miliardi di persone nel mondo. Sembrerebbe quindi che avere una base cliente estesa sia condizione necessaria per generare big data. Se queste sono le regole del gioco, allora solo le grandi multinazionali potrebbero competere nella data-driven economy, in quanto gli unici soggetti in grado di generare big data tramite i propri prodotti.

Dal 2012 passiamo al 2017, quando Nathan Ruser riuscì a scoprire delle basi segrete dell'esercito americano in Siria, semplicemente usando i dati di Strava – un social network in cui corridori e ciclisti di tutto il mondo riportano i propri allineamenti. Con suo, e nostro, grande stupore, Ruser trovò profili, foto e dati di allenamento di centinaia di soldati americani di servizio in zone guerra. Ruser notò che decine di soldati correvano in zone deserte, almeno stando a Google Maps. Com'era possibile che decine di militari americani corressero sistematicamente a centinaia di chilometri di distanza dalla base più vicina? La risposta è intuitiva: doveva esserci una base non segnalata sulle mappe, una base segreta appunto. Per inciso, questa storia evidenzia un elemento che è frutto di quel cambiamento sociologico sopra descritto: vale a dire la scarsissima, quasi nulla, attenzione per la privacy della nostra società, che spinge addirittura soldati impegnati in pericolosissime zone di guerra a condividere con sconosciuti le proprie attività.

Ma torniamo a Nathan Ruser. Probabilmente qualcuno si starà chiedendo: «Ruser? Chi era costui?». Un

hacker di qualche potenza straniera? Un terrorista? Niente di tutto ciò. Nel 2017, Nathan era un tranquillo studente alla Australian National University di Canberra, un ventenne che aveva trovato un passatempo durante le proprie vacanze estive.

LA CENTRALITÀ DEI SOCIAL NETWORK

I casi di Target e Ruser pongono in evidenza il ruolo fondamentale dei social network nella data-driven economy. Essi hanno in qualche modo spezzato il monopolio delle grandi imprese nell'accesso e nell'uso dei big data, rendendoli disponibili a una platea più ampia, potenzialmente a chiunque abbia un minimo di competenze di programmazione, come appunto un adolescente in vacanza. Questa democratizzazione dell'accesso ai big data è uno degli elementi di maggior valore dei social network per la società, ma purtroppo anche quello più ignorato. Demonizzati, anche a ragione, come la causa di molti mali contemporanei – dal cyberbullismo alla deriva populista di molte democrazie occidentali – in realtà i social network hanno permesso agli scienziati di tutto il mondo di accedere a questa risorsa fondamentale, dando nuovo impulso alla ricerca, con evidenti vantaggi per la società. Abbiamo oggi modelli basati sui dati Twitter o Facebook che ci permettono di identificare soggetti maggiormente a rischio di sviluppare disturbi mentali, malattie cardiovascolari o suicidarsi, con tassi di successo maggiori dei modelli tradizionali sviluppati da esperti.

Se è vero, come crediamo, che i big data siano la benzina dell'economia e dello sviluppo scientifico, è assolutamente necessario garantire che quanti più soggetti possano avere accesso a tali dati. Lasciare che essi siano appannaggio esclusivo delle grandi multinazionali finirebbe per concedere a queste ultime un'ulteriore, decisiva fonte di vantaggio competitivo: il monopolio della ricerca nell'ambito del machine learning e dell'intelligenza artificiale.

[1] *M. Cox, D. Ellsworth, «Application-Controlled Demand paging for Out-of-Core Visualization», Proceedings, Visualization, 1997, p.1*

[2] *«Big Data: The Next Frontier For Innovation, Competition, and Productivity», McKinsey Global Institute, 2011 («"Big data" refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze», p. 1, Executive summary).*

IN SINTESI

- Il vero vantaggio dei big data non risiede tanto nella quantità di dati registrati da un software o da un applicativo, quanto dalla capacità di questi di fornire informazioni estremamente dettagliate su un individuo. È su questo elemento distintivo che si fonda la data-driven economy in cui siamo immersi.
- I social network sono stati quel fenomeno sociologico che ha cambiato per sempre il nostro senso della privacy e che ha permesso di trasformare dei semplici byte in big data, spingendo il singolo individuo a cedere e condividere i dati che produce.
- I big data sono la benzina dell'economia e dello sviluppo scientifico, e per questo è necessario garantirne l'accesso a quanti più soggetti possibili. Lasciare che essi siano appannaggio esclusivo delle grandi multinazionali finirebbe per concedere a queste ultime un'ulteriore, decisiva fonte di vantaggio competitivo: il monopolio della ricerca nell'ambito del machine learning e dell'intelligenza artificiale.