Tantissimi Dati, tantissimi dubbi. La dittatura imperfetta della tecnologia

Alessandro Giuliani Angela Schito | 6 Dicembre 2017

Questi strani tempi ci costringono a soffermarci su argomenti piuttosto bizzarri che qualche anno fa non solo non sarebbero stati meritevoli di alcun commento, ma neanche avrebbero preteso un nome particolare. Un caso paradigmatico è quello dei cosiddetti 'Big Data' (Tantissimi Dati in Italiano).

Questi strani tempi ci costringono a soffermarci su argomenti piuttosto bizzarri che qualche anno fa non solo non sarebbero stati meritevoli di alcun commento, ma neanche avrebbero preteso un nome particolare. Un caso paradigmatico è quello dei cosiddetti '*Big Data*' (*Tantissimi Dati* in Italiano). Nel suo articolo apparso sulla rivista 'Wired' del 2008, Chris Anderson annunciava con malcelata soddisfazione 'la fine del metodo scientifico' provocata dall'avvento dei '*Big Data'* che consentirebbero a sistemi automatici di esplorazione (*Data Mining*) di mettere in luce tutta l'informazione ivi racchiusa. Questa operazione di estrazione di conoscenza *può fare del tutto a meno dell'intervento umano* (e quindi dell'analisi critica del senso da dare ai risultati). Potremmo insomma affidarci totalmente alla potenza dei computer che esplorano masse immense di dati e che ci forniranno la risposta a tutti i nostri quesiti e la sistemazione definitiva del mondo.

Per esaminare nel dettaglio questa annunciata rivoluzione, occorre partire dall'Analisi dei Dati (sezione della statistica descrittiva multidimensionale che si occupa di estrarre informazione rilevante da grandi raccolte di dati): vi chiediamo un piccolo sforzo che però ci aiuterà a fondare su basi solide e non ideologiche cosa sta avvenendo nel mondo del lavoro (e non solo, purtroppo!). Per prima cosa chiariamo una volta per tutte che, nonostante l'affascinante aura di modernità del termine "Big Data", lo schema teorico della statistica multidimensionale era già stabile alla metà del secolo scorso, il progresso è esclusivamente tecnico, non certo concettuale: prima dell'avvento dei PC, per analizzare una matrice di dati ci volevano mesi di calcoli laboriosi (non difficili), oggi lo sviluppo di computer potenti e a basso costo consente di ottenere gli stessi risultati in pochi minuti.

La matrice dei dati (comunque grande) è il materiale di partenza di qualsiasi esplorazione statistica: essa è formata da righe che rappresentano le unità statistiche (persone, animali, cellule, banche) e da colonne (variabili), che ne sono i descrittori, sperimentali (volume, densità, numero di globuli rossi) o demografici (genere, età, luogo di nascita). L'informazione rilevante viene estratta da dette matrici in termini di *associazioni notevoli* (correlazioni significative) tra le righe e/o tra le colonne. Un esempio è rappresentato dall'analisi dei *cluster* che individua associazioni notevoli tra unità statistiche se queste, invece di essere disperse casualmente nello spazio, si aggregano in sotto-insiemi (cluster) di individui simili tra di loro e molto differenti da quelli appartenenti ad altri gruppi.

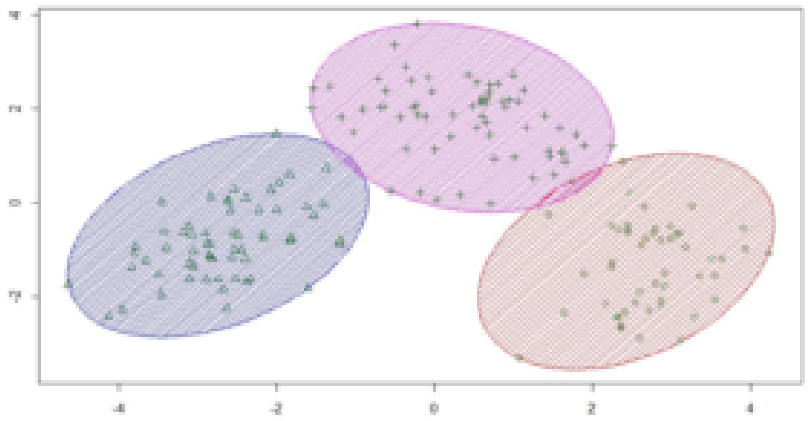


Figura 1 – Analisi dei cluster

L'analisi dei cluster riportata in figura separa le unità statistiche in tre gruppi (indicati da simboli e colori diversi) sulla base del calcolo delle loro distanze reciproche: il risultato numerico è il prodotto di un calcolo automatico *ma l'osservatore intelligente è essenziale per dare un senso al risultato, che altrimenti sarebbe nulla più che un bel disegno!* Le associazioni fra variabili (Fig.2) corrispondono all' individuazione di coppie di descrittori (X,Y) che variano in maniera coerente. Se 'valori grandi' della X si associano a 'valori grandi' della Y abbiamo una correlazione positiva (pannello di sinistra di Fig.2), altrimenti se l'associazione è tra valori grandi e valori piccoli si ha una correlazione negativa (pannello di destra); entrambe sono associazioni notevoli. L'assenza di correlazione è riportata nel pannello centrale della Fig.2.

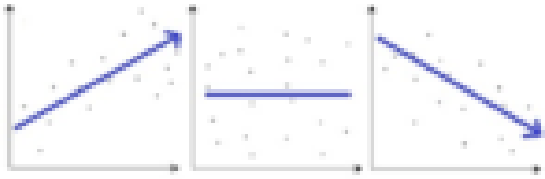
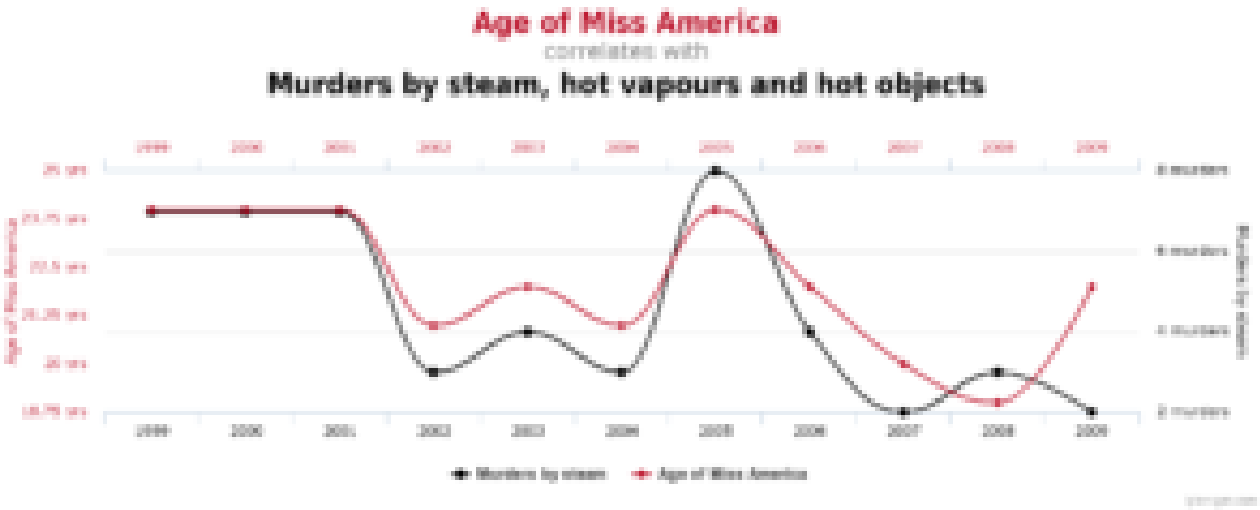


Figura 2 Relazioni tra variabili

Queste nozioni ci servono per arrivare ad uno dei problemi classici dei "*Big Data*", cioè le correlazioni spurie. L'immagine di seguito è estratta da un sito incredibilmente interessante e svelerebbe una oscura congiunzione astrale che lega l'affermarsi di una reginetta di bellezza agli incidenti domestici.



Ebbene sì, secondo il grafico ci sarebbe una stretta relazione tra gli andamenti temporali dell'età di Miss America (in rosso) e il numero di persone morte per colpa di incidenti con il vapore o con oggetti bollenti (e.g. pentole che bollono in cucina, ferri da stiro). Questo è un esempio di correlazione spuria. Il sito si basa su tantissimi dati (centinaia di migliaia di variabili) disponibili gratuitamente sul web dove un semplice algoritmo calcola il coefficiente di correlazione tra tutte le coppie di variabili e riporta sullo schermo le correlazioni elevate. Per puro effetto del caso si ottengono delle relazioni fortissime fra variabili che non hanno in realtà nulla a che fare fra loro.

La metodologia statistica *consente di stimare con molta precisione il numero (non la natura!) delle correlazioni spurie* che ci attendiamo a fronte della numerosità delle coppie di variabili messe a confronto. Basta una semplice occhiata a questi risultati per rendersi conto della vacuità della profezia di Anderson e della gravità delle dichiarazioni dei "filosofi dei Big Data" che si scagliano contro la scienza intesa come attività eminentemente umana e guidata dalle ipotesi. Questa realtà terrificante non ha a che vedere solo con la sensatezza o meno delle 'macchine-sputa-correlazioni': il punto è che la tecnologia, in quanto prodotto della modernità, dotata dell'aura di oggettività (nessun intervento umano), si sta candidando a diventare il 'dittatore perfetto'.

Uno degli ambiti più interessati da questo fenomeno è quello del lavoro dove un numero crescente di aziende destina risorse all'analisi di Big Data per ricavare informazioni utili a prendere decisioni nelle strategie di business o negli aspetti organizzativi e di gestione del personale. Viene chiamata *workforce* o *people analytics* e si basa sull'idea che l'analisi dei dati permetta non solo di conoscere al meglio il funzionamento dei processi produttivi, ma anche di prevedere le potenzialità di un candidato o la possibile resa di un lavoratore all'interno di un determinato contesto lavorativo. Tale deriva idolatrica, dunque, vedrebbe riconoscere potenzialità predittive ai dati non solo rispetto agli obiettivi delle aziende, ma anche rispetto alla valutazione dei lavoratori, *scongiurando ogni intervento umano e cancellando la componente relazionale del rapporto lavorativo.*

Sulla base del presupposto che la valutazione umana può essere influenzata da pregiudizi e preconcetti che ledono il principio di imparzialità con cui le "risorse umane" (se ancora sono ritenute tali) devono essere gestite. E' il caso di Mevaluate, una piattaforma, approdata anche in Italia (e poi bloccata dal Garante per la Privacy), nata per la qualificazione reputazionale che, attraverso un algoritmo, misura l'affidabilità delle persone in campo economico e professionale, con l'obiettivo di "*contribuire alla creazione di un mondo migliore, più sano, proficuo… dove intelligenze, capability e spiritualità vengono riconosciute e valorizzate, tutelando contemporaneamente la collettività da delinquenti e incompetenti, aumentando prevenzione e sicurezza*".

Oppure, il caso di Percolata, un'azienda informatica della tanto celebrata Silicon Valley, specializzata nella realizzazione di algoritmi in grado di profilare i dipendenti di aziende che operano nel settore del commercio al dettaglio: in questo modo le macchine riuscirebbero a stabilire il rendimento dei lavoratori secondo una scala di valutazione (alto o basso) e organizzerebbero orari e turni, scegliendo le combinazioni ottimali della forza lavoro. Le aziende che hanno adottato questo sistema avrebbero dichiarato un aumento del 30% delle vendite: l'algoritmo che sostituisce il lavoro del direttore del personale renderebbe più efficiente il servizio di vendita perché non sarebbe influenzato da errori di valutazione cui il giudizio umano è solito incorrere (favorendo qualcuno al posto di un altro). E a farne le spese

sono i lavoratori ("Come faccio a spiegare al software che devo andare a prendere mio figlio all'asilo perché non sta bene?") e il senso del lavoro, completamente svuotato dalla logica della disintermediazione e della falsa equità.

L'esaltazione del progresso digitale sta spingendo verso la cancellazione del fattore umano, l'unico in grado di restituirci un senso nella complessità, di coglierne le sfumature, di farci sentire parte di un progetto condiviso, non pedine di un meccanismo oscuro. Come quello messo in piedi in Cina dove, in via obbligatoria dal 2020, ogni persona sarà valutata in base al 'credito sociale', cioè passando attraverso un sistema nazionale di valutazione (del tutto automatico) destinato a misurare la loro affidabilità agli occhi del governo. Una sorta di sistemazione definitiva del mondo (come quella preannunciata da Anderson) che, lungi dal darci risposte, apre scenari terribili: che cosa ne sarà di quanti finiranno nella lista nera?

E' cruciale, dunque, saper rendere ragione della sostanziale *vacuità delle classificazioni e correlazioni non guidate da ipotesi, così come della falsa oggettività dei* 'profili automatici' che, lungi dall'essere necessari, ci sembrano la (labile) maschera di una dittatura moderna: gli elementi minimali di statistica qui forniti dovrebbero far comprendere al lettore come l'appartenenza a un 'cluster ottimale' di un lavoratore o la correlazione tra una sua caratteristica e la produttività non sono altro che manipolazioni camuffate da una (falsa) aura scientifica.

Tags: Big Data Progresso digitale Statistica