

ANTONIETTA MIRA

Istituto Lombardo Accademia di Scienze e Lettere
Università degli Studi dell'Insubria, Università della Svizzera italiana

La Scienza dei Dati: come Estrarre Informazioni dai Big Data

I big data sono enormi “data base” caratterizzati da 3 “V”: Volume, Velocità e Varietà.

Siamo noi stessi che contribuiamo a generare i big data ogni volta che affidiamo le nostre vite alla tecnologia. Pensate ai telefoni intelligenti, i motori di ricerca, Twitter, Facebook, Instagram, Whatsup, Waze, Googlemaps, Booking.com, Tripadvisor . . .

Ma non solo le interazioni fra persone-persone e persone-cose generano dati. Recentemente si sono aggiunte le interazioni fra cose-cose, la cosiddetta “internet of things”.

Per fare un esempio, i più di 2000 sensori nelle ali e nei motori di un aeroplano comunicano al computer di bordo se ci sono pezzi che si stanno usurando e questo computer, che sa dove stiamo volando, a sua volta ordina il pezzo da sostituire direttamente al magazzino dell'aeroporto, se le scorte, pure monitorate, sono sufficienti, o al produttore che lo fa trovare direttamente sul posto all'atterraggio allertando in contemporanea il personale che dovrà sostituirlo. Questo è un esempio di internet delle cose che è già una realtà per i motori della Rolls Royce.

Tutte queste nuove tecnologie generano dati intesi in senso allargato, non solo numeri, ma anche immagini, video, coordinate gps e altro ancora.

I dati vengono organizzati e conservati in banche dati e questo termine sottolinea il Valore che i dati contengono. È questa la quarta “V” che caratterizza i big data. Il valore dei big data risiede nelle informazioni che contengono.

La scienza dei dati (Data Science) si occupa di estrarre informazioni e quindi valore dai big data.

È una nuova disciplina legata alla rivoluzione post-industriale che stiamo vivendo, "rivoluzione guidata da dati digitali, computazione e automazione", per dirla con le parole di un recente comunicato del parlamento europeo.

L'informazione viene estratta dai dati attraverso un approccio multidisciplinare con sofisticati strumenti metodologici quali la statistica, la probabilità, la fisica, l'intelligenza artificiale e tecniche di machine learning e data mining.

Ma la catena del valore generata dai big data non termina qui.

Dall'informazioni si deve passare alla conoscenza attraverso il supporto di esperti di dominio, ("knowledge domain experts"). La conoscenza deve poi supportare decisioni che a loro volta si trasformano in azioni, deve cioè diventare "actionable knowledge". Le decisioni e le azioni generano a loro volta dati analizzabili in un percorso migliorativo e un processo che si autoalimenta fino all'ottimizzazione dello stesso. Il processo non può però prescindere dall'intervento umano anche se alcune fasi dello stesso si possono automatizzare.

Il Report MC Kinsey Global Institute del 2011 prevede che negli Stati Uniti, dal 2018 ci potrebbe essere un eccesso di domanda di persone con competenze analitiche e di scienza dei dati fra 140 e 190 mila persone.

La statistica e la scienza dei dati sono state definite da Hal Varian, Google chief executive, "the sexiest job of the the 21st century" in un articolo comparso sul New York Times nel 2009.

A voi, studenti del 21esimo secolo, la sfida di imparare ad estrarre informazione e valore dai dati.