

# Big Data e Formazione

Sergio Margarita<sup>1</sup>, Riccardo Beltramo<sup>1</sup>

<sup>1</sup> Università di Torino, Dipartimento di Management, Torino, Italia  
(sergio.margarita, riccardo.beltramo)@unito.it

**Abstract.** Big Data è un termine che viene spesso associato a grandi volumi di dati, impossibili da trattare con strumenti e metodologie tradizionali. Questa visione troppo restrittiva dimentica l'importanza degli aspetti qualitativi e la rilevanza strategica del Big Data, insita nella sua capacità di estrarre valore da grandi masse di dati disomogenei e non strutturati, che lo fanno ritenere un pilastro della "Quarta rivoluzione industriale" disegnata nel progetto tedesco Industrie 4.0. In questo contributo si vuole porre l'accento sulla necessità e le difficoltà di formazione al Big Data, nei vari livelli, dalla Scuola fino alla formazione dei *data scientist* e dei *data analyst*. Si presenta inoltre un approccio formativo multi-disciplinare e multi-livello che supera alcune di queste difficoltà e si illustra la sua messa in opera in un progetto pluriennale che sta coinvolgendo Scuola e Università in una iniziativa di alternanza scuola-lavoro ante litteram promossa dal Dipartimento di Scienze Merceologiche dell'Università di Torino.

**Keywords:** Formazione·Big Data·Alternanza scuola-lavoro

## 1 Introduzione

Ad aver richiamato l'attenzione del management sul Big Data nel 2011 è stato il Rapporto McKinsey [1]. Doug Laney [2] ha definito le tre dimensioni del fenomeno, che ne sottolineano gli aspetti qualitativi, note come le "3 V": il Volume dei dati, la Varietà delle loro fonti e rappresentazioni nonché la Velocità di crescita. Dati voluminosi e provenienti da fonti eterogenee quali database dell'impresa, reti sociali, Internet of Things (IoT); dati non strutturati rappresentati in modi diversi, la cui abbondanza cresce a velocità sempre maggiore. Si tratta di dati che necessitano di elaborazioni sempre più veloci per "estrarne valore" in tempo reale e per alimentare processi decisionali man mano più complessi. E' proprio la capacità di estrazione di valore da masse enormi di dati disomogenei che conferisce al Big Data caratteristiche strategiche e fa ritenere che il Big Data sia un pilastro della "Quarta rivoluzione industriale", risultando strategico per le imprese e i paesi che ne fanno una base fondante per il proprio sviluppo industriale, primo fra tutti la Germania con il progetto Industrie 4.0 [3].

L'importanza del fenomeno porta a riflettere in questo articolo sulla formazione al Big Data, accennando a due casi, e un progetto Scuola-Università in essere da diversi anni che costituisce su questi temi un'iniziativa di alternanza scuola-lavoro ante litteram, promossa dal Dipartimento di Scienze Merceologiche dell'Università di Torino.

## 2 Dalla Business Intelligence al Big Data

Alla fine degli anni '80, il linguaggio SQL, quale standard per i DBMS (DataBase Management System) di tipo relazionale, ha portato alla Business Intelligence (BI), dando un forte impulso all'analisi delle informazioni nell'impresa e fornendo ai manager veri e propri cruscotti per estrarre informazioni significative (*data mining*) tramite strumenti statistici [4]. A sua volta, la BI ha portato al Big Data.

Come molti strumenti di BI, quelli su cui si basa il Big Data sono Open Source: il principale, Hadoop ([hadoop.apache.org](http://hadoop.apache.org)), è un framework di calcolo distribuito per il trattamento di grandi insiemi di dati ripartiti su cluster di centinaia o migliaia di server.

Semplificando, possiamo dire che è da una doppia esplosione dei dati della BI che è nato il Big Data: un'esplosione quantitativa (reti sociali, smartphone, IoT, sensori) e una qualitativa (dati grezzi non strutturati e multimediali, raccolta di qualunque tipi di dati al massimo livello di dettaglio, indipendentemente dall'uso che se ne farà).

## 3 I Casi Walmart e Industrie 4.0

Il caso Walmart (che citiamo soltanto) è particolarmente significativo in tema di Big Data. Wal-Mart Stores, Inc. (Walmart) è un'impresa americana della grande distribuzione, nata nel 1962, da diversi anni in prima posizione per fatturato nella classifica Fortune Global 500. Impiega 2,3 milioni di persone, ha un fatturato (2016) di 482 miliardi di dollari, con più di 11.500 negozi in 28 paesi e 260 milioni di clienti [5].

Si ritiene che Walmart disponga di dati completi su 150 milioni di cittadini americani. Nel 2014, Walmart dichiarava di gestire circa 30 petabyte di informazioni sugli acquisti dei suoi clienti (un petabyte è un milione di miliardi di byte).

Un fattore riconosciuto del successo di Walmart è l'asset strategico costituito dal sistema informativo e dalle metodologie Big Data messe in atto dall'impresa prima ancora che il termine stesso si diffondesse. In altre parole, si tratta della visione strategica che l'impresa ha del proprio sistema informativo.

Il Big Data è un pilastro di quello che sembra annunciare la quarta rivoluzione industriale: il progetto Industrie 4.0. Promosso nel 2010 dalla Germania, Industrie 4.0 è un progetto chiave nella strategia delle alte tecnologie del governo tedesco. Esso intende sviluppare una nuova organizzazione della produzione che tocca tutta la catena del valore, con l'introduzione nella fabbrica dell'IoT. Robot e sistemi intercomunicanti accedono in tempo reale a tutti i dati per reagire il più velocemente possibile ad eventi, incidenti o situazioni di non conformità, con strumenti tipici del Big Data.

Con l'adozione crescente di questo modello, si rischia che si crei una sorta di digital divide fra i paesi [6], con un'ulteriore fuga dei cervelli verso quelli che hanno integrato in modo strategico queste nuove tecnologie nella propria politica industriale.

## 4 Le Nuove Sfide per la Formazione al Big Data

La velocità di evoluzione delle tecnologie ha sempre creato difficoltà di aggiornamento e adeguamento in tutti i contesti formativi, creando spesso disallineamento tra offerta e domanda di figure professionali. Il Big Data non fa eccezione e addirittura fa

nascere sfide di notevole difficoltà: novità nelle competenze richieste (le competenze informatiche, statistiche, matematiche e manageriali richieste sono nuove), estrema multidisciplinarietà del Big Data (queste nuove competenze devono essere integrate, diversamente da quanto propone il sistema universitario italiano), forte impatto sulle imprese (mancano una cultura *data-driven* e una visione strategica dei dati).

Le figure professionali emergenti, più critiche in termini di formazione, sono quelle note come *data scientist* (formazione statistica, matematica e informatica) e *data analyst* (formazione gestionale quantitativa in management e marketing). La preparazione di queste figure è compito di un sistema di formazione superiore, purché segua un approccio integrato e si adatti ad un modello *data-driven*, in mancanza dei quali si avrà una discordanza fra le professionalità che escono dal sistema formativo e quelle richieste dal mondo del lavoro, come già avviene in altri casi.

## 5 Le Difficoltà nella Formazione

Nel Big Data emerge la necessità di una formazione fortemente multidisciplinare che prepari al dialogo tra specialisti. Siamo abbastanza lontano da questo obiettivo e diverse difficoltà si presentano nella formazione al Big Data. Due domande per rendersene conto: quanti sono i laureati in discipline manageriali che hanno una buona conoscenza di statistica e di matematica? Quanti sono i matematici e statistici che hanno una buona conoscenza dell'impresa? Ricollegandoci alle nuove sfide, fra le principali difficoltà si possono identificare: elevato numero e varietà delle discipline coinvolte, ridotta continuità fra livelli di formazione, formazione poco orientata alle applicazioni.

In compenso, il Big Data presenta caratteristiche che facilitano la formazione: il software per il Big Data è Open Source, multi-piattaforma, rivolto a sistemi a basso costo, le soluzioni commerciali si basano su questi stessi strumenti Open Source, l'intervento formativo è coerente con le richieste del mondo del lavoro.

Proponendo un approccio integrato, si possono definire interventi di formazione, erogati a diversi gradi dell'istruzione scolastica per promuovere la continuità formativa, che non siano solo a beneficio di singoli livelli formativi (Scuola, Università) ma che li coinvolgano in un modello di scambio reciproco di competenze.

## 6 Un Progetto di Alternanza Scuola-Lavoro Ante Litteram

L'approccio appena descritto richiama alcune delle caratteristiche del modello dell'alternanza scuola-lavoro che si ritrovano nel progetto Scato18<sup>®</sup> ([www.scato18.net](http://www.scato18.net)).

Nato nel 2011 presso il Dipartimento di Scienze Merceologiche dell'Università di Torino, si sviluppa nel quadro del Progetto Interreg V.E.T.T.A., con la Regione Piemonte. Ideato dal prof. Riccardo Beltramo, Scato18<sup>®</sup> è un dispositivo di rilevazione di variabili ambientali e gestionali e trasmissione ad un server per l'elaborazione e la visualizzazione, in un percorso di sensibilizzazione sul tema della sostenibilità. Sin dal 2011 si avvia una collaborazione con l'Istituto Tecnico Industriale Ettore Majorana di Grugliasco (TO) che vede la partecipazione di allievi alla progettazione, realizzazione e installazione dei sistemi. Dal 2012, il sistema Scato18<sup>®</sup> è installato in contesti quali rifugi, miniera, imprese agricole. La collaborazione, durata tre anni, ha per-

messo a studenti dell'Istituto di applicare sul campo quanto studiato in laboratorio e di prendere contatto con una realtà simile a quella dell'impresa. Nel periodo 2012-2014 si affiancano le attività legate al progetto promosso dalla Regione Piemonte "Il risparmio energetico comincia dalla scuola" per la sensibilizzazione dei ragazzi verso una nuova consapevolezza di rispetto dell'ambiente. Il progetto vede l'installazione di sistemi in diverse scuole secondarie di primo grado e centri di formazione Piemontesi.

Per sperimentare altre modalità di integrazione fra i percorsi formativi, a partire dal 2014 si avviano collaborazioni con tirocinanti del Politecnico di Torino, sempre sullo stesso modello. In parallelo si svolgono attività di divulgazione, sensibilizzazione e formazione nelle scuole. I sistemi ancora in attività in vari contesti hanno raccolto ad oggi (aprile 2017) circa 15 milioni di rilevazioni. Le attività in corso e future prevedono, a scopo didattico, la progettazione e costruzione di un cluster per l'analisi dei dati con metodologie e strumenti tipici del Big Data [7].

## 7 Conclusione

Il Big Data richiede alle imprese un vero salto organizzativo per diventare delle *data-driven organization*, le quali spesso hanno solo una visione strumentale anziché strategica dell'informatica. Lo stesso adattamento culturale è richiesto a livello di paese, anche per non perdere il treno sul quale alcuni paesi sono già saliti.

Nella formazione, il Big Data fa nascere nuove sfide, fra le quali l'elevato numero e la varietà delle discipline coinvolte o la necessità di una loro forte integrazione nel processo formativo. Queste si scontrano con l'attuale sistema formativo poco orientamento alle applicazioni e che procede lungo percorsi isolati, sia fra le discipline, sia fra i gradi delle scuole. In compenso, l'uso di soluzioni Open Source rappresenta un forte catalizzatore per la sperimentazione didattica e assicura una forte coerenza della formazione con le figure richieste.

L'approccio metodologico di formazione-lavoro che è stato presentato, impostato sullo scambio di competenze fra diversi gradi e strutture del sistema dell'istruzione e applicato da più di sei anni in progetti nel campo delle Scienze Merceologiche ha dimostrato di riuscire a ridurre le difficoltà insite nel processo formativo.

## Bibliografia

1. Manyika, J., Big data: The next frontier for innovation, competition, and productivity, McKinsey Global Institute Report (2011)
2. Laney, D., 3D Data Management: Controlling Data Volume, Velocity, and Variety, META Group Inc. Application Delivery Strategies, 949 (2001)
3. Kagermann, H., Wahlster, W., Helbig, J., Recommendations for implementing the strategic initiative Industrie 4.0: Final report of the Industrie 4.0 Working Group (2013)
4. Kimball, R., Ross, M., The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling - 3rd Edition, John Wiley & Sons, USA (2013)
5. Walmart Annual Report 2016, <http://stock.walmart.com> (2017)
6. Gartner Inc., Hype Cycle for Emerging Technologies 2015, Gartner Inc., Stamford (2015)
7. Mönning, C., Schiller, W., A Hadoop data lab project on Raspberry Pi, <http://scn.sap.com/> (2015)