Data scientist: la professione più interessante del XXI secolo

Alberto Daprà

Vice Presidente Lombardia Informatica

alberto.dapra@lispa.it

1. Introduzione

Hal Varian, il Chief Economist di Google è noto per aver dichiarato " *The sexy job in the next 10 years well be statisticians. People think I'm joking, but who would've guessed that computers engineers would've been the sexy job of the 1990s ?"* Dove naturalmente il termine "sexy" deve essere inteso nella accezione "business" cioè interessante, divertente, appassionante,…etc..etc..
Ebbene questa profezia si sta iniziando ad avverare : nelle organizzazioni sta emergendo una figura professionale nuova e di grande valore il "data scientist". La continua crescita dei dati in formato digitale, sia in forma strutturata che non strutturata, disponibili nelle moderne organizzazioni, ha fatto emergere il ruolo strategico di una nuova figura professionale in grado di analizzare, interpretare, gestire, visualizzare queste enormi quantità di dati e renderli utilizzabili per supportare i processi decisionali all'interno delle organizzazioni.
I "data scientist" assumeranno, in un prossimo futuro, nelle organizzazioni ruoli simili a quelli dei cosiddetti "quants" che, negli anni 80 e 90, hanno innovato profondamente la cultura e la pratica della finanza.
I "quants" sono persone con background nelle scienze fisiche o matematiche o statistiche che hanno aiutato le grandi organizzazioni finanziarie a creare nuovi prodotti (fino ai vituperati "derivati"), nuovi algoritmi di gestione per analizzare ed interpretare i dati finanziari ed il comportamento degli investitori.
Sono già molte le aziende che già utilizzano con successo le professionalità dei "data scientist" per il miglioramento dei processi e dei prodotti (e.g. Linkedin, Google, GE, Amazon, Zynga, Netflix, Yahoo…etc..etc) ma la carenza di queste professionalità sul mercato sta diventando una criticità in molti settori.

2. Chi è il "data scientist"

Il "data scientist" è una nuova figura professionale che deve avere caratteristiche ibride : capacità di trattamento dei dati unite a notevoli capacità di comunicazione e di analisi. Una delle caratteristiche chiave di questa nuova figura professionale, oltre alle solide competenze tecniche (statistica, informatica..) è la capacità di comunicare i risultati dell'analisi a tutti gli "stakeholders" e di supportare i senior managers delle aziende nelle decisioni critiche.
Ma, forse, la caratteristica principale di queste figure deve essere la "curiosità" : cioè l'attitudine, tipica di chi utilizza il metodo scientifico, di analizzare in profondità i fenomeni non fermandosi alle apparenza ed identificare un insieme di ipotesi da verificare ed esplorare con l'analisi e lo studio dei dati. Da questo punto di vista si può paragonare il lavoro del "data scientist" a quello del fisico sperimentale che spesso deve progettare gli strumenti, condurre gli esperimenti, analizzare i dati ottenuti e presentare i risultati a tutta la comunità scientifica oltre che ai responsabili delle ricerche.
Dato che attualmente non esistono percorsi di formazione universitaria specifici in "Data Science" (anche se numerose università in USA li stanno pianificando…) il "data scientist " va costruito partendo da altri background universitari che possono comprendere oltre alla statistica e alla informatica, le scienze matematiche e fisiche e le scienze economiche e sociali. Infatti è fondamentale che, oltre a solide competenze in statistica e informatica, il "data scientist" conosca i processi di business e l'organizzazione aziendale. Un'altra competenza molto importante del "data scientist" riguarda i processi di gestione della qualità del dato e le competenze di analisi di dati non strutturati (testi, video, foto…) che nel tempo diventeranno una componente fondamentale dei "big data" disponibili nelle organizzazioni.

3. "Data scientist" e processi decisionali.

Una ricerca recente condotta su oltre 300 aziende nord-americane dallo MIT Center for Digital Business, ha dimostrato che le aziende che utilizzano processi decisionali "data-driven" sono più profittevoli e più produttive rispetto alla media. La spesso citata massima di P.Drucker "*you can't manage what you don't measure*" oggi, con l'enorme ed in continua crescita, quantità di dati digitali disponibili all'interno delle organizzazioni sta spingendo verso un cambiamento significativo della cultura manageriale che deve sempre più diventare "data-driven" in tutte le decisioni critiche. La disponibilità di "big data" non solo nelle aziende che interagiscono con il cliente solo con la Rete (e.g. Amazon, Google, Yahoo, ..) ma anche nei settori tradizionali può portare ad un eccezionale miglioramento dei processi di gestione e di previsione e di gestione del cliente. Nella ricerca citata riportata in (1) si fa cenno alla società PASSUR Aerospace che, utilizzando e gestendo una quantità enorme di informazioni (comprese quelle metereologiche), è in grado di prevedere con precisione i tempi di atterraggio degli aerei, consentendo cosi alle società che gestiscono gli aereoporti una gestione ottimizzata delle risorse con risparmi significativi e miglioramento dei servizi ai viaggiatori.
Nelle aziende migliori si sta affermando una nuova cultura dei processi decisionali che, senza rinunciare ai fattori di leadership, giudizio ed intuizione sarà sempre più guidata dall'analisi dei dati.
La domanda prima da porsi di fronte alla interpretazione di un fenomeno non dovrà più essere "che cosa pensiamo ?" ma "che cosa conosciamo ? " e i "data scientist" saranno protagonisti di questo cambiamento culturale nei processi decisionali che inizia ponendosi le giuste domande : " che cosa dicono i dati ?" "qual è la fonte dei dati?" "che tipo di analisi sono state fatte ?" "Qual è il livello di confidenza nei risultati ?" etc..etc..

4. Caratteristiche principali del "data scientist"

Le caratteristiche principali di questa nuova figura professionale possono essere cosi sintetizzate :

- Competenze tecniche profonde in : statistica, informatica, trattamento dei dati strutturati e non strutturati, gestione della qualità dei dati e tecniche di visualizzazione
- Conoscenza delle tecnologie e delle piattaforme disponibili per la gestione ed il trattamento dei dati. Si tratta di un settore in evoluzione a grande velocità è quindi essenziale un continuo aggiornamento.
- Conoscenze dei processi aziendali e del funzionamento delle organizzazioni.
- Capacità di interazione e comunicazione interpersonale.
- Attitudine personale alla "curiosità" ed all'utilizzo del metodo scientifico-sperimentale nell'approccio dei problemi.
- Conoscenza delle problematiche e delle sfide del dominio settoriale (industria, finanza, settore pubblico, telecomunicazioni,…) di interesse

5. Conclusioni

Vi sono forti evidenze, sia di studio che sperimentali, che in futuro le organizzazioni che sapranno trasformare la propria cultura manageriale adottando sempre più un processo decisionale "data-driven" miglioreranno in modo significativo le proprie performance e la loro produttività. Per questo saranno sempre più richieste le figure del "data scientist" in grado di interpretare e gestire grosse moli di dati. Data l'attuale carenza di queste figure professionali e la crescente domanda del mercato, il sistema universitario ha una grande opportunità di contribuire attivamente alla formazione di competenze chiave per il miglioramento delle performance delle organizzazioni.

Riferimenti Bibliografici

(1) Getting Control of Big Data, Harvard Business Review, Ottobre 2012
A.McAfee, Enterprise 2.0, Harvard Business School Press, 2009
E. Brynjolfsson, Race Against The Machine, Digital Frontier Press, 2012
T.H. Davenport, Judgement Calls, Harvard Business Review Press, 2012