

Alessandro Giaume

#DATA SCIENTIST

Tra competitività
e innovazione



FrancoAngeli

Informazioni per il lettore

Questo file PDF è una versione gratuita di sole 20 pagine ed è leggibile con



La versione completa dell'e-book (a pagamento) è leggibile con Adobe Digital Editions. Per tutte le informazioni sulle condizioni dei nostri e-book (con quali dispositivi leggerli e quali funzioni sono consentite) consulta [cliccando qui](#) le nostre F.A.Q.



Professioni Digitali

Le professioni di domani, raccontate dai protagonisti di oggi

Direzione di Alberto Maestri

Il paradigma digitale ha aperto opportunità straordinarie, per chiunque. Innovazione, dati, omni-canalità sono solo alcune delle keyword alla base di questa profonda rivoluzione: per i professionisti di oggi e domani diventa fondamentale rimanere aggiornati e competenti in uno scenario così dinamico, fluido, stimolante. In questo contesto Professioni Digitali propone una collezione di guide pratiche raccontate dai protagonisti di oggi: autori che hanno saputo fare la differenza nel proprio settore diventando fonte di ispirazione per tanti. Una Collana dedicata a consulenti, freelancer, professionisti che desiderano aggiornare le proprie competenze e a quanti hanno da poco intrapreso la via del digitale. Libri agili, pratici e concreti, ricchi di consigli, casi studio, testimonianze e contributi di grandi esperti nazionali e internazionali, pensati per approfondire competenze specifiche e le metodologie più innovative.

Il dialogo continua su...



blog.francoangeli.it/professionidigitali



FrancoAngeliDigitale



I lettori che desiderano informarsi sui libri e le riviste da noi pubblicati possono consultare il nostro sito Internet: www.francoangeli.it e iscriversi nella home page al servizio "Informatemi" per ricevere via e.mail le segnalazioni delle novità.

Alessandro Giaume

#DATA SCIENTIST

Tra competitività
e innovazione

Prefazione di Marco Icardi

Progetto grafico della copertina: Gianni Camusso
In copertina: © Shutterstock

1a edizione. Copyright © 2017 by FrancoAngeli s.r.l., Milano, Italy

L'opera, comprese tutte le sue parti, è tutelata dalla legge sul diritto d'autore. L'Utente nel momento in cui effettua il download dell'opera accetta tutte le condizioni della licenza d'uso dell'opera previste e comunicate sul sito www.francoangeli.it.

Indice

Prefazione , di <i>Marco Icardi</i>	pag.	9
1. Promuovere la cultura digitale	»	11
2. Apertura e integrazione	»	12
3. Competenze per la trasformazione	»	13
4. Una sfida per la crescita dell'Italia	»	14
Introduzione	»	17
1. Piccola storia moderna dei dati: la nascita del Data Scientist	»	17
2. Dai file VSAM ai Big Data	»	23
1. Big Data in azienda	»	29
1. Dati strutturati e non strutturati: i Big Data nella realtà	»	29
2. Data Governance: la potenza è nulla senza controllo	»	37
3. Big Data e mercato	»	42

Nuovi scenari. Il Data Scientist a sostegno delle città: il governo degli ambienti urbani è una questione di Big Data

di Michele Vianello pag. 46

2. Da grande voglio fare il Data Scientist » 50

1. Una carriera come Data Scientist.
Quale background è premiante? » 50

Approfondimento – Strategy Execution e Big Data » 54

2. Il lato umano dei Big Data » 57

3. Thinking... » 62

4. Quattro modi di essere Data Scientist in azienda » 65

I (data) designer di una nuova specie
di Cosimo Accoto » 69

3. Anche il Data Scientist ha un suo “profilo” » 72

1. The *why* and the *how* » 72

2. Il Data Scientist è sexy » 84

3. Skill e competenze/1: quali avere e come svilupparle » 90

Approfondimento – Il metodo scientifico » 97

4. Skill e competenze/2: tra formazione ed esperienza » 101

Quattro chiacchiere con un Data Scientist
di e con Andrea Casati » 106

4. Questa Job Description mi sta stretta » 110

1. Esplorando il mondo dei dati » 110

2. L'età dei “perché?” (e i rischi connessi) » 116

3. Ti raccomando, fa' la cosa giusta	pag.	121
<i>Dove l'impresa incontra il Data Scientist</i>		
di Luca De Biase	»	127
5. L'importanza dello strumento giusto	»	129
1. Tools...	»	129
2. La tecnologia del nuovo millennio: quanto tempo costa?	»	132
3. La frontiera del Cognitive Computing	»	137
4. L'Apollo XIII rientra grazie a un cronografo manuale: l'importanza degli strumenti "analogici"	»	138
<i>La nuova pelle dell'informatica e la nuova figura del Data Scientist</i>		
di Pietro Leo	»	140
6. Big Data e oltre: social engineering, algorithm economy e COBOT	»	145
1. Si può guarire dall' <i>influenza</i> ?	»	145
2. La algorithm economy, sempre più basata sulle interconnessioni di un ecosistema pervasivo	»	147
3. Che cosa succede quando una macchina intelligente prende una decisione al nostro posto?	»	150
<i>Big Data e tecnologie esponenziali</i>		
di David Orban	»	152
I 5 proverbi Zen del Data Scientist	»	157
Bibliografia	»	159
Ringraziamenti	»	161

Prefazione

*di Marco Icardi**

Viviamo in un'epoca di trasformazione accelerata delle nostre modalità di interagire, di utilizzare le cose, dei processi di produzione e della fornitura di servizi, nemmeno immaginabile fino a pochi anni fa, che cambia in maniera profonda le nostre abitudini e la nostra quotidianità. Parliamo di trasformazione digitale che, in modo pervasivo e a livello globale, si inserisce nella nostra esperienza di vita personale e anche negli oggetti che utilizziamo.

Digitale è sinonimo di numerico e digitalizzare significa convertire una certa grandezza fisica in cifre, traendo valore dai dati. Per farlo è necessario coinvolgere le persone, digitalizzare la cultura e la mentalità all'interno delle organizzazioni. E sono proprio le persone l'asset su cui fare leva e da cui partire per attuare questa rivoluzione.

Sul valore del dato e su come estrarlo con le competenze necessarie si concentrerà gran parte di questo libro: dall'informazione alla conoscenza, dalla conoscenza alle decisioni, dalle decisioni all'execution. L'obiettivo finale, infatti, è utilizzare le informazioni ottenute dai dati per prendere decisioni efficaci ed elaborare strategie utili al business delle organizzazioni. Processi facili a dirsi, ma che si rivelano impegna-

* CEE Regional VP, SAS Institute e Amministratore Delegato SAS Institute Italia.

tivi da mettere in atto, secondo modalità che non possono avere i tempi, le verifiche, la valutazione dei ritorni che eravamo abituati a utilizzare in passato. Ma andiamo con ordine.

L'onda crescente della digital transformation ha spinto le aziende a indirizzare gli investimenti IT nelle quattro principali tecnologie della Terza Piattaforma che in un acronimo si fanno chiamare SMAC (Social, Mobile, Analytics & Big Data, Cloud) per iniziare a ridisegnare l'esperienza dei clienti, gestire i dati come veri elementi di differenziazione, creare inediti modelli di business, invece che limitarsi a ottimizzare processi già esistenti.

L'Internet of Things (IoT) sarà il più grande cambiamento a cui assisteremo dopo la nascita stessa di Internet. Non c'è nessun altro fenomeno che cambierà più profondamente l'economia, le aziende e anche la vita di tutti i giorni. Gli oggetti connessi possono fare qualunque cosa, soprattutto automatizzare una serie di azioni ripetitive aggiungendo una precisione che non è patrimonio dell'uomo, ma delle macchine sì.

Lo IoT è associato anche al concetto di Industria 4.0, un contesto all'interno del quale macchine intelligenti, dispositivi e persone sono tra di loro collegate: questa interconnessione conduce alla possibilità di migliorare il processo decisionale attraverso ampi Database condivisi e analytics avanzati che favoriscono una rapida evoluzione verso nuovi modelli di business.

Per esempio, l'Harbor Research ipotizza che i nuovi player focalizzati su spazi di mercato IoT non perseguono il tradizionale approccio di business focalizzato sulla centralità del prodotto, ma modelli basati sulla collaborazione, lo sviluppo partecipato del prodotto e l'innovazione di servizi sempre più intelligenti. Si produce un contesto in cui chiunque può avere accesso alla conoscenza e realizzare il proprio business anche con capitale limitato: non a caso l'economista Jeremy Rifkin parla di "terza rivoluzione industriale a costo marginale zero".

1. Promuovere la cultura digitale

Siamo tutti consapevoli di quanto il livello della cultura digitale di un Paese influenzi il suo livello di innovazione. Più in profondità, ogni ambito della cultura digitale – che sia fatto da cittadini, specialisti o da altri tipi di lavoratori – ha le sue complessità, i suoi specifici bisogni, le sue azioni da intraprendere. Ma una cosa è certa: senza una larga azione sulle persone, sugli studenti di tutte le scuole, su tutti i lavoratori, senza un reale salto di qualità della cultura digitale del Paese, l'innovazione resta problematica.

Il digitale introduce cambiamenti in tutti i contesti; rendere disponibili i fondamenti di cultura e di competenza digitale, senza limitarsi agli aspetti tecnici, è fondamentale per creare la consapevolezza necessaria ad affrontarli. Da qui la necessità di parlare di “cultura digitale per il lavoro”, o meglio di Data Scientist, concetto più esteso e diverso dalla cittadinanza digitale e dalle sole competenze necessarie ai professionisti ICT. Inoltre, fornendo a tutti dei fondamenti di “cultura digitale” si apre, in lunga prospettiva e al termine di un impegnativo percorso di crescita, l'opportunità per alcuni di acquisire competenze di e-leadership. L'e-leadership caratterizza una figura che arricchisce la cultura digitale di particolari attitudini riconducibili alla capacità di immaginare il cambiamento, a volte anche in maniera radicale e di contestualizzarlo nell'organizzazione in cui opera.

La cultura digitale si può paragonare allora alla cassetta degli attrezzi ormai necessaria in qualsiasi lavoro, e coloro che eccellono nel suo utilizzo rappresentano gli “e-leader”.

La vera, e per certi versi nuova, sfida del Paese consiste nella capacità di intervento nel mondo del lavoro, di tutti i lavori, in cui la capacità elementare di uso del computer non basta e il possesso professionale della tecnologia non serve, ma serve l'attitudine al cambiamento accompagnata dalla consapevolezza che il digitale può esserne una formi-

dabile leva. Non che tutti i lavoratori debbano o possano diventare degli e-leader, ma che a tutti sia data la strumentazione culturale (digitale) necessaria, con la speranza che in ogni piccola (e media e grande) impresa, e non ultima la Pubblica Amministrazione, ne arrivi almeno uno.

Oggi la cultura digitale deve essere un patrimonio di tutti. Si tratta di un'azione capillare, endemica, virale e deve iniziare dalla scuola ed essere sostenuta dalle aziende e dalla Pubblica Amministrazione, dando la possibilità di innescare un processo di “democratizzazione” e condivisione della conoscenza, proprio a partire dall'interno dell'ecosistema aziendale. Se prima i business analytics erano appannaggio esclusivo del Chief Information Officer, del Top Management e del Chief Marketing Officer, oggi l'accesso alle informazioni si sta estendendo anche verso la base della piramide. Si assiste a un'evoluzione culturale in cui la conoscenza del patrimonio informativo aziendale abilita un circolo virtuoso sia in termini di produttività interna sia di customer satisfaction (il servizio guadagna in qualità, velocità, puntualità). Della cultura digitale questo libro delinea i contenuti e le modalità, per svilupparla e mantenerla sempre attiva nelle organizzazioni.

2. Apertura e integrazione

La cultura digitale e la capacità dei Data Scientist di utilizzare il valore dei dati comporta apertura, non solo mentale, ma anche delle aziende a integrare fra loro processi produttivi e di servizio, a condividere competenze, a utilizzare esperienze di settori diversi per massimizzare la capacità di innovazione in tempi brevi.

Anche la transizione della Pubblica Amministrazione verso il digitale può essere accelerata puntando su Open Government, Open Data e Interoperabilità in una logica di trasparenza, collaborazione estesa PA-cittadini-imprese,

migliore comunicazione tra enti e valorizzazione degli investimenti ICT.

Il patrimonio di dati e di informazioni open è enorme. Per primi, i dati relativi alle dimensioni economiche prodotti dal funzionamento stesso delle città: i bilanci economici, le statistiche infrastrutturali, che sempre più spesso vengono restituiti sotto forma di Open Government Data. Poi ci sono i dati raccolti dai sensori che troviamo in qualsiasi dispositivo: dagli ascensori, ai semafori fino allo smartphone che abbiamo in tasca. È l'Internet of Things che produce informazioni continue in tempo reale. Sempre più importanti sono poi le informazioni, aggiornate ogni secondo in tutto il mondo, che scaturiscono dalle interazioni sui social media. Informazioni che possono essere elaborate e utilizzate per modelli predittivi o per capire i gusti e le opinioni del pubblico, dei cittadini e dei consumatori: è la cosiddetta Sentiment Analysis.

Ritengo che non solo i manager del futuro, ma anche quelli di oggi, debbano essere in grado di governare i processi tecnologici attivi all'interno delle loro aziende, oltre a comprenderne le problematiche di business, perché, in assenza di persone con queste conoscenze, il settore produttivo rimarrà nella nicchia che occupava in passato e che dovrebbe invece aver già superato. L'obiettivo è quindi colmare questo gap ed è possibile farlo solo parlando il più possibile del valore del dato, aprendosi anche con le aziende considerate competitor e diffondendo le conoscenze in nostro possesso. Oggi il potere non è tenere per sé le informazioni, la conoscenza è diffusa.

3. Competenze per la trasformazione

Tra tre anni la maggior parte degli investimenti aziendali in tecnologie informatiche sarà destinata ai processi digitali. Secondo le previsioni di IDC, la spesa mondiale in

tecnologie per la digital transformation supererà i duemila miliardi di dollari nel 2019, andando a pesare per oltre il 50% sul totale della spesa ICT mondiale attesa per quell'anno.

IDC definisce digital transformation quel processo continuo attraverso il quale un'azienda apporta o si adatta a un cambiamento radicale nel proprio mercato (la cosiddetta "disruption") facendo leva su tecnologie e competenze digitali per innovare modelli di business, prodotti e servizi, migliorando nel contempo l'efficienza operativa e le prestazioni aziendali.

Tra le aziende europee, rivela la medesima indagine di IDC, l'80% ha già individuato internamente un leader, ovvero un responsabile, per le attività di digital transformation. Nel 23% dei casi si tratta del Chief Information Officer, nel 19% del CTO e nel 14% del CEO in prima persona. Seguono con percentuali più basse il CFO, e il nuovo Chief Data Officer, che in Italia, secondo uno studio di Boston Consulting Group, dalle attuali 200 persone con questo titolo, passeranno a 2.000 nel 2018 e arriveranno fino a 10.000 nel 2020.

4. Una sfida per la crescita dell'Italia

Se le cifre ci dicono che, seppure lentamente, l'Italia sta uscendo da una crisi che è durata otto anni, ora più che mai è necessario agire per uscire diversi da come siamo entrati e con altre prospettive e priorità. In questo contesto non possiamo che constatare che non esiste più una old economy e una new economy: il digitale è l'aria che tutti noi respiriamo e il digitale è quindi il principale settore in cui investire per una crescita possibile e uno sviluppo equo e sostenibile.

Penso all'Italia come l'ambiente ideale per valorizzare le possibilità offerte dalla rivoluzione digitale e dai nuovi

modelli di business. Il Piano per la banda larga e il Piano Industria 4.0, un progetto di politica industriale finalmente centrato sull'innovazione della nostra base manifatturiera, pongono le basi per poter innovare le nostre aziende e rilanciare il paese.

Ma credo anche alla valorizzazione della nostra biodiversità, la più alta al mondo, della cultura e arte, diffusa su tutto il territorio, e delle eccellenze intellettuali nei nostri atenei e centri di ricerca, affermati a livello mondiale. Le filiere produttive sviluppate nei diversi distretti industriali sono il terreno fertile su cui sviluppare questa trasformazione e investire per la crescita.

L'Industria 4.0 rappresenta una grande occasione per il rilancio del settore manifatturiero italiano: oggi la manifattura italiana vale solo il 16% del valore aggiunto sul PIL, mentre era il 20% nel 2000. Eppure, le nostre imprese manifatturiere hanno tutte le qualità per essere all'avanguardia in un mondo che cambia così velocemente: abili nell'innovazione integrata, agili nella gestione, flessibili nell'adattarsi al cliente, interessate a coprire nuove nicchie di mercato.

Un'intelligenza diffusa, quindi, che richiede messa in campo di un'efficace sinergia tra strumenti, conoscenze, facoltà e visioni, per poterne ricavare valore, per poter riconoscere, tra i rumori di fondo, le effettive opportunità e accelerare gli investimenti.

Noi allora possiamo agire, e il momento per farlo è ora. Le occasioni anche in Italia non mancano e per coglierle occorre lavorare sulle competenze in modo concreto, come fa questo libro, che indirizza su un percorso che delinea bene i profili necessari, le competenze richieste e l'apertura intellettuale a cui predisporre.

Introduzione



Quando i dati hanno cominciato ad assumere importanza per la nostra vita? E quando la capacità di aggregare dati in informazioni ha dato il via all'economia della conoscenza?



1. Piccola storia moderna dei dati: la nascita del Data Scientist

I dati.

Dal latino, la parola “datum” viene tradotta letteralmente con “fatto”. Ovvero la descrizione elementare di un avvenimento, un fenomeno occorso, o anche semplicemente un'entità.

Dalla combinazione di più dati possiamo giungere alla conoscenza di una informazione, più o meno complessa e articolata, e la combinazione di più informazioni ci consente di affrontare e risolvere complesse operazioni e problemi o di esprimere un'opinione, sulla cui base giungere a prendere decisioni.

I dati, per come oggi li conosciamo, sono rappresentati da cifre e lettere, composti o meno in numeri e parole, immagini, video, suoni, esperienze.

Per quello che ne possiamo sapere, oggi i dati sono ovunque. Ed è possibile disporre in maniera relativamente semplice, in ambito industriale o, più in generale, di business.

Quello di cui non è semplice disporre sono invece le informazioni, di base o complesse, su cui fondare l'interpretazione di un fenomeno di mercato o politico, o di una combinazione di questi, dando loro un senso compiuto, utile a indicarci quali decisioni prendere.

E questo è tanto più vero quanto più i volumi di dati disponibili aumentano a un ritmo talmente elevato da richiedere l'utilizzo di una nuova unità di misura: lo Zettabyte (ZB), ovvero 1 trilardo di byte.

Per avere un'idea pratica della quantità di dati che 1 Zettabyte esprime, basti pensare che questo corrisponde a circa un milione di volte le informazioni depositate in tutte le biblioteche degli Stati Uniti.

Uno studio IDC del 2014¹ afferma che il volume di dati impiegati annualmente erano 4,4 ZB nel 2013 e che raggiungeranno i 40 ZB entro il 2020, con una crescita del 40% anno su anno, grazie agli oltre 2 miliardi di persone connesse oggi a Internet e ai milioni di imprese che oggi utilizzano Internet per i propri affari, ma anche e soprattutto grazie alla rapida diffusione dell'Internet of Things.

Appare chiara a questo punto la definizione che Gartner ha dato di Big Data:

Big data is high-Volume, high-Velocity and high-Variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation.

¹ "The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things", April 2014.

Dicevamo, dare un senso ai dati. È una lunga storia, lunga come quella del computer.

Ecco alcune tappe significative.

Nel 1962 viene pubblicato su *The Annals of Mathematical Statistics* un lungo articolo a firma di John Tukey dal titolo “The future of Data Analysis”, che probabilmente rappresenta il primo documento su quella che sarebbe poi diventata la Data Science. Il concetto primario che questo testo² porta all’attenzione dei lettori è: quale e quanto vitale deve essere il ruolo svolto dai programmi di electronic computing?

Dobbiamo attendere fino al 1974 per avere una raccolta³ dei programmi di Data processing più utilizzati negli Stati Uniti e in Svezia, organizzata attorno al concetto di dati come definiti nella *Guide to Concepts and Terms in Data Processing* presentata nel 1968 da parte della IFIP (International Federation for Information Processing) durante un congresso intitolato *Datalogy, the science of data and of data processes and its place in education*. E per la prima volta compare il termine Data Science, largamente impiegato nel testo.

Nel 1977 viene finalmente fondata la IASC (International Association for Statistical Computing): “It is the mission of the IASC to link traditional statistical methodology, modern computer technology, and the knowledge of domain experts in order to convert data into information and knowledge”. Questa la mission, ben più che una dichiarazione d’intenti, piuttosto un piano già piuttosto preciso.

“Companies are collecting mountains of information about you...” Così inizia l’articolo di copertina pubblicato nel settembre del 1994 da *Business Week*. Per la prima volta si descrive come utilizzare grandi masse di dati per “prevedere un possibile comportamento” e sulla base di questo decidere come indirizzare le azioni di marketing.

² <http://projecteuclid.org/euclid.aoms/1177704711>.

³ Naur P., *Concise Survey of Computer Methods*, Petrocelli/Charter, New York, 1974.