



UNIVERSITÀ DEGLI STUDI DI PADOVA

**DIPARTIMENTO DI SCIENZE ECONOMICHE E AZIENDALI
“M. FANNO”**

CORSO DI LAUREA IN ECONOMIA E MANAGEMENT

PROVA FINALE

***“Big Data e Data Analytics come supporto ai processi decisionali:
definizione e analisi dei vantaggi competitivi per le aziende”***

RELATORE:

CH.MO PROF. MICHELE MORETTO

LAUREANDO: NICOLA PORRU

MATRICOLA N. 1113116

ANNO ACCADEMICO 2017 – 2018

INDICE

INTRODUZIONE	4
1 <i>BIG DATA</i>	6
1.1 IL PROBLEMA DELLA DEFINIZIONE	6
1.2 L'ANALISI DI <i>BIG DATA</i> : <i>DATA SCIENCE</i> O <i>DATA ANALYTICS</i> ?	9
2 UNA NUOVA REALTÀ TECNOLOGICA	11
2.1 LE FONTI DI <i>BIG DATA</i>	11
2.2 ACQUISIZIONE E IMMAGAZZINAMENTO.....	13
2.3 <i>KNOWLEDGE DISCOVERY IN DATABASES</i> (KDD).....	14
3 IL CONTESTO DI RIFERIMENTO	17
3.1 TIPOLOGIE DI STRUMENTI OFFERTI	18
3.2 GARTNER <i>MAGIC QUADRANT</i>	20
3.3 <i>MAGIC QUADRANT FOR ADVANCED ANALYTICS</i>	21
4 <i>BIG DATA</i>: SFIDE E OPPORTUNITÀ IN AZIENDA	25
4.1 I SETTORI INDUSTRIALI	27
4.2 LE AREE FUNZIONALI	34
4.3 CASI DI STUDIO	39
5 IL TEMA DELLA PRIVACY	42
5.1 GDPR (REG. UE 2016/679)	42
5.2 COSA CAMBIA PER LE AZIENDE	44
5.3 PRIVACY E <i>BIG DATA</i>	46
CONCLUSIONI	48
BIBLIOGRAFIA	51

Introduzione

“*Information is the oil of the 21st century, and Analytics is the combustion engine*”. Peter Sondergaard, Executive Vice President di Gartner, azienda leader mondiale nella consulenza strategica (2011).

Siamo sottoposti a un diluvio di dati ininterrotto, che ci accompagna quotidianamente e rappresenta un tesoro inestimabile, tanto da definirlo appunto il nuovo petrolio, in grado di cambiare profondamente comportamenti individuali e collettivi, con importanti conseguenze economiche e sociali.

Ci sono diverse ragioni che spiegano questa esplosione di informazioni. Sicuramente l’incremento della popolazione mondiale è stato accompagnato da un aumento del numero di persone che possiedono dispositivi elettronici, con capacità che migliorano nel tempo (soprattutto smartphone e computer), attraverso cui interagiscono con altri soggetti. L’evoluzione tecnologica ha spinto notevolmente questa esplosione grazie a software sempre più efficienti e sistemi d’immagazzinamento sempre più capienti che hanno ridotto notevolmente il costo di trasmissione e acquisizione dei dati condivisi. Infine tali informazioni vengono applicate in diversi ambiti, più frequentemente e ampiamente che in passato: le persone le usano per regolare certi aspetti della propria vita; le aziende le aggregano per ottimizzare numerose attività e trasformare i propri settori; lo Stato le analizza per guidare le scelte politiche e comprendere meglio i trend macroeconomici attuali e futuri.

Il presente lavoro, diviso in cinque capitoli, ha lo scopo di capire e approfondire il tema dei *Big Data*, un termine recentemente introdotto per identificare queste grandi quantità di dati, e dei *Data Analytics*, strumenti per analizzarli ed estrarne valore. In particolare si sofferma sull’esaminarne le conseguenze economiche, soprattutto dal punto di vista delle aziende, ossia come possano sostenere un vantaggio competitivo.

Nel primo capitolo darò una definizione quanto più univoca possibile di questi due termini tecnici, tenendo conto del fatto che ad oggi non ne esiste una e che la confusione è ancora tanta. Nel secondo disegnerò un quadro generale di come possano essere trattati, partendo dall’individuazione delle principali fonti e chiudendo con la loro acquisizione e pre-elaborazione.

Successivamente viene descritta la situazione di offerta attuale attraverso il *Magic Quadrant*, strumento di analisi strategica dell’azienda di Gartner, che ne evidenzia i *player* più importanti, e presentata una breve classificazione dei *Data Analytics*, sulla base delle ricerche di Forrester, altra famosa società di consulenza.

Il quarto capitolo riguarda l'analisi delle performance delle imprese che usano gli strumenti precedentemente individuati: una tabella esaminerà quali sono le applicazioni principali di questi ultimi e in quali settori sono diffuse, a cui seguirà un elenco degli utilizzi possibili per ciascuna funzione aziendale. Queste riflessioni saranno sostenute da una serie di casi empirici, a titolo di esempio e di spiegazione.

Infine tratterò del tema della privacy, questione molto attuale e delicata perché i dati che le aziende immagazzinano e sfruttano sono spesso di natura personale. La trattazione, tuttavia, sarà breve, poiché la letteratura di riferimento è ampia, complessa e tocca ambiti giuridico-legislativi che esulano dall'obiettivo di questo elaborato.

1. *Big Data*

I *Big Data* sono ovunque, in ogni business o funzione industriale e aumentano ad un tasso estremamente rapido. Un articolo di Forbes (2015) afferma che nel 2020 ogni essere umano produrrà 1.7 megabyte di dati ogni secondo e che ogni azienda, indipendentemente dalle dimensioni, farà uso di una qualche tecnica di analisi delle informazioni per migliorare le proprie performance. I *Big Data*, infatti, diventeranno la chiave della competizione, essendo ormai un fattore di produzione importante tanto quanto le risorse umane e il capitale. Sono e saranno la base di nuovi fenomeni di crescita esponenziale e innovazione, in grado, se ben gestiti e interpretati, di incrementare notevolmente la produttività di molti settori, in particolare la pubblica amministrazione, la sanità, la finanza, l'industria dell'elettronica e dell'informatica. Non tutti, però, sanno cosa sono effettivamente questi *Big Data*...

In questa prima parte cercherò quindi di darne una definizione quanto più chiara possibile, concludendo con un breve paragrafo per esplicitare le differenze fra i diversi termini utilizzati, spesso confusi gli uni con gli altri.

1.1 Il problema della definizione

Secondo il glossario di Gartner¹ “*Big Data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation*”.

Non esiste ancora una definizione rigorosa del termine *Big Data*. Nel 2011, il Teradata Magazine dell'omonima azienda statunitense² afferma che “Un sistema di *Big Data* eccede/sorpassa/supera i sistemi hardware e software comunemente usati per catturare, gestire ed elaborare i dati in un lasso di tempo ragionevole per una comunità/popolazione di utenti anche massiva”. Un'ulteriore definizione di *Big Data* è stata data dal McKinsey Global Institute nel 2011: “Un sistema di *Big Data* si riferisce a *dataset* la cui taglia/volume è talmente grande che eccede la capacità dei sistemi di database relazionali³ di catturare, immagazzinare, gestire ed analizzare”.

¹ Guida curata dall'azienda per la spiegazione di terminologie e definizioni tecnologiche

² Teradata è un'azienda leader mondiale nell'*Enterprise Data Warehousing* (strumenti utilizzati per raccogliere, integrare e distribuire i dati, a supporto delle decisioni dell'organizzazione)

³ Un database è relazionale se basato su tabelle strutturate che indicano le relazioni fra gli elementi di cui sono composte

Per poter parlare di *Big Data* il volume dei dati deve essere correlato alla capacità del sistema di acquisire le informazioni così come arrivano dalle differenti sorgenti di dati che sono adoperate. Quindi un sistema diventa *big* quando aumenta il volume dei dati e, allo stesso tempo, aumentano la velocità e il flusso d'informazioni per secondo che il sistema deve acquisire e gestire. Non esiste, tuttavia, una dimensione di riferimento poiché varia a seconda dei settori e degli strumenti software a disposizione. Inoltre aumenterà sicuramente col tempo, poiché le macchine sono sempre più veloci e i *dataset* sempre più grandi (McKinsey Global Institute, 2011). Negli ultimi due anni c'è stato, infatti, un incremento del 90% dei dati prodotti nel mondo. Il presidente del consiglio di amministrazione di Google, ex amministratore delegato, Eric Schmidt ha affermato, in una conferenza del 2010, che la mole di dati creata ogni due giorni nel mondo era pari a quella prodotta in tutta la storia dell'umanità fino al 2003.



Fig.1: Le 5 V dei Big Data (Dreamstime.com, s.d.)

Il termine *Big Data* è relativamente nuovo. Comincia a diffondersi all'inizio del 2001 quando Doug Laney, analista di Gartner, formula la definizione delle tre V dei *Big Data*, all'interno di un modello di crescita tridimensionale:

- **Volume**

Le organizzazioni raccolgono enormi quantità di dati da una grande varietà di sorgenti tra cui transazioni finanziarie, social media, database tradizionali, e-mail, eventi, sensori – per esempio Wal-Mart, il gigante americano della grande distribuzione, gestisce 2,5 Pb di dati ogni ora, sulla base delle transazioni con i suoi clienti (McAfee & Brynjolfsson, 2012), equivalente a oltre 150 volte i dati contenuti in tutti i libri della *American Library of Congress*, la biblioteca più grande del mondo.

- **Velocità**

I dati sono generati e fluiscono a velocità senza precedenti, in alcuni casi rendendoli obsoleti in tempi ridotti. Risulta dunque strategico presidiare e controllare anche il loro ciclo di vita, per soddisfare la sempre più forte necessità di gestire questi fiumi di dati in tempo reale, o quasi, al fine di effettuare analisi tempestive su di essi – si stima per esempio che nel 2015 Twitter generasse 12 Tb di dati ogni giorno (MIUR, 2016).

- **Varietà**

I dati vengono generati, collezionati e utilizzati in qualsiasi tipo di formato. In passato si tendeva a prendere in considerazione per le analisi principalmente dati strutturati e numerici (tabelle o dati provenienti da database tradizionali) e la loro manipolazione veniva eseguita mediante uso di modelli relazionali poiché presentavano uno schema ben formato. Oggi, per avere analisi più accurate, è necessario prendere in considerazione anche dati non strutturati (ad esempio file di testo generati dalle macchine industriali, log di server web⁴ o dei firewall, e-mail, video, audio, transazioni finanziarie, dati GPS) e semi strutturati (file compilati con sintassi XML, grafi) per la cui elaborazione sono necessari tecnologie e database non convenzionali.

In molti casi questo modello è ancora valido, nonostante esso sia stato successivamente esteso tramite l'aggiunta di altre due V:

- **Veridicità**

È fondamentale assegnare un indice di veridicità ai dati su cui si basano le analisi, in modo avere una misura dell'affidabilità, per poter prendere le giuste decisioni. Infatti, considerando la varietà delle fonti e la velocità alla quale i dati possono variare, è evidente che se i dati alla base delle analisi sono poco accurati, i risultati delle stesse non saranno migliori;

- **Valore**

Si riferisce alla capacità di trasformare i dati in valore. Prima di avviare un progetto *Big Data* è importante valutare e documentare quale sia il valore effettivo portato al business, poiché la raccolta e l'analisi delle informazioni richiedono ingenti investimenti in termini di tecnologie da implementare e nuove figure professionali da assumere. Non solo la quantità, ma anche la qualità e la veridicità dei dati diventano fondamentali affinché possano davvero creare valore.

⁴ Elenco cronologico delle attività e delle operazioni svolte da un utente, un sistema operativo o da altri programmi

Con il passare del tempo, nonostante non siano sempre tenute in considerazione, ulteriori dimensioni si sono aggiunte al modello, quali ad esempio:

- **Variabilità**

I flussi di dati possono essere altamente inconsistenti e con picchi periodici. Gestire i picchi di dati giornalieri, stagionali o innescati da eventi, soprattutto in caso di dati non strutturati, può essere una vera sfida.

- **Complessità**

I dati arrivano da molteplici fonti eterogenee, rendendo così difficile collegare, ripulire e trasformare i dati trasversali. Tuttavia, è necessario connettere le relazioni, individuare gerarchie e collegamenti al fine di estrarre il massimo valore dalla raccolta dei dati. Inoltre, maggiore è la dimensione del *dataset*, maggiore è la difficoltà di gestione dei dati.

1.2 L'analisi di *Big Data*: *Data Science* o *Data Analytics*?

A complicare l'ambiguità della definizione di *Big Data*, vi è anche la confusione che caratterizza le terminologie utilizzate per identificare le strumentazioni di analisi di queste enormi moli di dati. Le differenze sono, infatti, poco chiare e molto tecniche, difficili da comprendere se non si è esperti nel campo. Cerchiamo di esplicitarle in maniera semplice.

Data Science è la scienza dei dati, intesa come un ambito che raccoglie tutte le discipline che riguardano la pulizia, la preparazione e l'analisi dei dati (non necessariamente deve trattarsi di *Big Data*) per estrarne informazioni di valore altrimenti non evidenti. Utilizza, accanto a una serie di tecniche matematiche, statistiche e di programmazione, algoritmi di *machine learning* e intelligenze artificiali.

I *Data Analytics* (o, più semplicemente, *Analytics*) sono strumenti che si basano sull'inferenza statistica per esaminare in maniera approfondita dati grezzi e le conoscenze a disposizione per individuare correlazioni, trend o verificare teorie e modelli esistenti. Rispondono a domande precise e partono da ipotesi formulate sin dall'inizio, focalizzandosi su particolari settori, con lo scopo di ottenere le best practises che portano ad un miglioramento del business. Possono essere distinti in *Business Analytics* (o *Business Intelligence*) e *Advanced Analytics* (anche detti *Big Data Analytics*).

La *Business Intelligence* si focalizza sul fornire alle persone dati e informazioni in tempo reale, per rendere le operazioni e i processi più efficienti. Utilizza la statistica descrittiva per misurare qualcosa, rilevare tendenze, fornire report a partire da *dataset* limitati (campioni), dati puliti e modelli semplici ricavati dallo storico dell'organizzazione.

Gli *Advanced Analytics*, invece, consentono di effettuare previsioni future e simulazioni *what-if* per verificare gli effetti di determinati cambiamenti sul business, nonché di ottenere analisi più dettagliate e approfondite. Attraverso l'uso di tecniche e strumenti più sofisticati sono in grado di analizzare *dataset* molto più grandi.

Questi ultimi sono gli strumenti che esaminerò più nel dettaglio nella terza parte.

2. Una nuova realtà tecnologica

L'evoluzione tecnologica degli ultimi anni ha permesso di ottenere più facilmente ed economicamente grandi moli di dati su ogni tipo di fenomeno d'interesse, appartenente a qualsiasi settore. Questo grazie alle sempre maggiori capacità automatiche di rilevazione e di calcolo, nonché al potenziamento dei sistemi di immagazzinamento e memorizzazione delle informazioni. Tali innovazioni hanno contribuito a creare la cosiddetta "società dell'informazione" (Azzalini, Scarpa, 2004), in cui il numero degli oggetti digitali connessi cresce esponenzialmente, con il crollo dei costi e l'aumento del valore di connettere, memorizzare ed elaborare informazioni.

Questi insiemi di dati sono raccolti in *dataset* di grandi dimensioni che possono essere gestiti, conservati e analizzati solo facendo uso di tecnologie avanzate. Una trattazione analitica e tecnica dei nuovi strumenti statistici e informatici che tale nuova realtà tecnologica richiede esula dallo scopo di questo elaborato, ma ritengo sia comunque utile fornire una panoramica generale. In particolare cercherò di rispondere alle tre domande seguenti: quali sono le fonti di *Big Data*? Come si acquisiscono e archiviano le informazioni raccolte? Quali sono le prime elaborazioni da svolgere su di esse?

2.1 Le fonti di *Big Data*

Ogni telefono cellulare registra continuamente una grande quantità d'informazioni sul luogo in cui si trova, sulla velocità dei suoi spostamenti, sulla cella più vicina alla quale si può collegare, sugli altri telefoni con i quali comunica, sui servizi applicativi che gli utenti adottano per svolgere le più svariate attività. Milioni di automobili raccolgono e registrano in ogni momento

Fig.2: La varietà di fonti che generano Big Data (Pixabay.com, 2016)



dati sui percorsi, le soste, gli stili di guida, lo stato del motore e dei freni con sensori, computer di bordo, scatole nere. Ogni contatore intelligente dell'energia elettrica, ogni macchina agricola, ogni termostato moderno raccoglie e comunica dati.

Nelle fabbriche ogni macchina è dotata di sensori che registrano e comunicano informazioni per migliorare i prodotti, per la manutenzione, per l'efficientamento della produzione. Nei campi i sensori avvertono i contadini dei cambiamenti climatici o della necessità di annaffiatura.

Anche con questi pochi esempi s'intuisce come, oltre alla varietà di formati e strutture, i *Big Data* presentano anche una

grande varietà di fonti. I dati da esse generati sono spesso classificati in:

- *human generated*
- *machine generated*
- *business generated*

Tra le sorgenti di dati *human generated* si trovano in particolare le piattaforme di social network (Facebook, LinkedIn), blogging (Blogger, Wordpress) e micro-blogging (Twitter, Tumblr), social news (Digg, Reddit), condivisione multimediale (Instagram, Flickr, YouTube), nonché le wiki⁵ (Wikipedia), i siti di domande e risposte (Yahoo Answers), i siti di recensioni (Yelp, TripAdvisor), i portali di e-commerce (eBay, Amazon), i *click stream*⁶ da siti web. Questo tipo di fonti è anche chiamato *user-generated content*, poiché i contenuti sono prodotti in gran parte dagli utenti.

I dati *machine generated* sono prodotti da sorgenti quali sensori GPS, dispositivi IoT, RFID⁷, centrali di monitoraggio di eventi meteorologici, contatori delle utenze energetiche, strumenti scientifici, sistemi di *High Frequency Trading*⁸ dei mercati finanziari, dispositivi biomedicali, macchinari industriali, smartphone.

Infine, per dati *business generated* si intendono tutti i dati, *human* o *machine generated*, prodotti internamente ad un'azienda relativi alle attività *data-driven* dei processi di business aziendali, ossia a quelle attività le cui decisioni sono determinate a partire da considerazioni sui dati a disposizione. Molti di essi sono dati storici, memorizzati staticamente e normalmente archiviati separatamente nei database relazionali della relativa area funzionale.

Un'impresa può acquisire dati utili da numerose fonti interne, quali il marketing, che fornisce informazioni sugli aspetti demografici e psicografici della clientela, sui comportamenti d'acquisto e sulle visite al sito web; il servizio clienti, che verifica il livello di soddisfazione del cliente e registra eventuali problemi di assistenza; la contabilità, che redige i bilanci e rileva sistematicamente dati di vendita, costi e flussi di cassa; la produzione, che fornisce dati sui volumi di produzione, sulle spedizioni e sulle scorte; il personale di vendita, che riporta le reazioni dei rivenditori e le attività dei concorrenti (Kotler et al., 2015). Ulteriori fonti sono le

⁵ Applicazione web che consente la creazione, la modifica e l'illustrazione di pagine di siti Internet

⁶ Il percorso compiuto da un singolo utente all'interno di un sito web

⁷ Radio-Frequency IDentification: tecnologia che permette di identificare, memorizzare e scrivere informazioni su oggetti, animali o persone grazie a particolari etichette elettroniche

⁸ Modalità di investimento che consiste nel mettere in atto un alto numero di negoziazioni di breve durata per lucrare su margini minimi, ma in grandi quantità

e-mail, le app e la *landing page*⁹ aziendali, gli *Open Data* (diffusione gratuita e in formato aperto dei dati riguardanti il funzionamento di organizzazioni pubbliche e private) e i dati sulle transazioni nei punti vendita riferiti dai partner.

2.2 Acquisizione e immagazzinamento

L'acquisizione di *Big Data* dalle fonti sopra citate può avvenire in diversi modi:

- accedendo ad API¹⁰ messe a disposizione dai servizi Web, grazie alle quali è possibile interfacciarsi ad essi per esaminarne i contenuti. Un esempio sono le Twitter API, Facebook Graph API e le API fornite da motori di ricerca come Google e Bing;
- utilizzando software di *web scraping* che, attraverso operazioni di *crawling*, *parsing* ed *entity extraction*¹¹, raccolgono automaticamente dati da siti web e documenti presenti in Internet;
- sfruttando strumenti di ETL (*Extract, Transform, Load*) che estraggono i dati dalle sorgenti, li trasformano rendendoli più omogenei e strutturati e li memorizzano in adeguati sistemi di immagazzinamento;
- catturando flussi continui di dati di *log* tramite particolari software che li salvano nei database e li elaborano in tempo reale;

Una volta acquisite, le informazioni vengono raccolte nei database operativi delle aziende. Esse sono, infatti, dotate di complessi sistemi informativi che contengono database, detti appunto operativi, con scopi differenti (raccolta delle fatture dei clienti e delle paghe dei dipendenti, elenco dei fornitori, assortimento prodotti ecc.). Gli operatori possono utilizzarli per l'inserimento di dati o per interrogazioni, relative a singole registrazioni o informazioni specifiche. Gli strumenti che svolgono queste funzioni si chiamano *OnLine Transaction Processing* (OLTP) e si basano sullo *Structured Query Language* (SQL), un linguaggio standardizzato che consente di comunicare, gestire e amministrare i vari database, oltre che di leggere e modificare i dati. Un'alternativa sono i modelli NoSQL (*Not only SQL*), di maggiore

⁹ Pagina web specificamente strutturata che il visitatore raggiunge dopo aver cliccato su un link o una pubblicità. Ha il preciso scopo di trasformarlo in cliente.

¹⁰ Acronimo di interfaccia di programmazione di un'applicazione: insieme di procedure per l'espletamento di un determinato compito all'interno di un programma

¹¹ Rispettivamente si intende analisi dei contenuti, della sintassi e classificazione degli elementi di un testo

semplicità di sviluppo e scalabilità: poiché sono privi di uno schema rigido, possono essere modificati più facilmente e hanno una velocità di risposta più alta. Essi sono maggiormente adatti a grandi volumi d'informazioni, che possono essere rappresentati con documenti, grafi, in formato colonnare o in chiave-valore in memoria¹².

A questo punto i dati ottenuti sono ancora grezzi e necessitano di ulteriori elaborazioni per poterne estrarre il massimo valore e fornire supporto alle strategie decisionali in ambito aziendale. In particolare si tratta di eseguire tutte le fasi del processo di *Knowledge Discovery in Databases* (Azzalini, Scarpa, 2004).

2.3 Knowledge Discovery in Databases (KDD)

I database operativi raccolgono informazioni molto diverse fra loro e spesso per fini differenti dall'analisi; inoltre l'attività strategica andrebbe ad interferire nelle attività quotidiane di lavoro su di essi. La prima fase di questo processo, quindi, consiste nel creare database strategici specificamente destinati al supporto decisionale, chiamati *Data Warehouse* (DWH). In essi confluiscono i dati raccolti dalle varie fonti dopo essere stati filtrati per mezzo dell'obiettivo strategico assegnato e ripuliti da ridondanze, incompletezze e incongruenze, al fine di ridurre il disordine, minimizzare lo spazio di archiviazione richiesto e facilitare future elaborazioni. Normalmente si costruiscono mediante l'aggregazione successiva di *data mart*, ovvero archivi informatici più piccoli focalizzati su particolari soggetti o fenomeni (per esempio su una determinata divisione commerciale), per poi essere continuamente aggiornati coi nuovi flussi di dati provenienti dai database operativi.

I DWH possono essere utilizzati in vari modi, di cui due sono i principali. Si possono creare nuovi *data mart* a partire dalla ricombinazione di quelli originari oppure si può cominciare ad estrarre informazioni utili attraverso un'analisi essenzialmente descrittiva (*OnLine Analytical Processing*, OLAP), che rappresenta la seconda fase del KDD. Tale attività viene generalmente eseguita grazie al supporto di una tabella ad entrata multipla in cui ognuna di essa corrisponde a una variabile e ogni cella contiene un indicatore ad esse riferito. Per esempio, un'azienda di carte di credito potrebbe individuare le abitudini di acquisto dei propri clienti (le unità statistiche), distinguendoli per caratteristiche demografiche come l'età, il sesso e la provenienza (le variabili) e indicando le frequenze di incrocio di tali caratteristiche (il contenuto delle celle). I DWH lavorano bene su dati strutturati e ancora meglio su piccoli, costanti e prevedibili

¹² Quando i carichi di lavoro in lettura o di elaborazione sono particolarmente gravosi, le cache delle informazioni vengono archiviate in una memoria con una bassa velocità nel rendere disponibili i dati, non appesantendo così le applicazioni.

carichi di lavoro. Nel caso di grandi ammontare di dati semi o non strutturati deve essere supportato da altri sistemi più potenti, che ne garantiscono la pulizia e l'elaborazione anche in tempo reale.

Fra le tante tecnologie adottabili, il sistema migliore è Apache Hadoop. Ideata nel 2005 da Doug Cutting e Mike Cafarella, è una piattaforma open-source costituita da due elementi essenziali: Hadoop Distributed File System si occupa dell'immagazzinamento e della conservazione dei dati, mentre MapReduce li processa per ricavare le informazioni ricercate. Lavora su diverse migliaia di server paralleli, detti nodi, che formano un cluster gestito da un terzo componente fondamentale, YARN (*Yet Another Resource Negotiator*); questo coordina le attività da eseguire, controllando la CPU, la memoria e lo spazio di archiviazione. Il successo della piattaforma è garantito dalla sua flessibilità, poiché l'immagazzinamento non è vincolato dal tipo di dato, dalla sua efficienza nel processare grandi quantità di dati rapidamente, dalla sua tolleranza al guasto, in quanto se anche un nodo si guastasse gli altri continuerebbero a funzionare, e dalla sua economicità rispetto agli altri sistemi. È, infatti, utilizzata da numerose aziende, fra cui, per citarne solo alcune, IBM, Facebook, Spotify e Yahoo!.

Infine, nella terza e ultima fase, si può procedere con analisi ed elaborazioni più complesse per costruire modelli predittivi, individuare relazioni significative, caratteristiche o configurazioni particolari dei dati. A questo proposito occorre tenere sempre in considerazione l'obiettivo della ricerca o, in sua assenza, fare attenzione alle variabili *leaker* (surrogati delle variabili d'interesse) e a non trovare qualcosa in maniera forzata pur di rendere fruttuoso il lavoro: Ronald H. Coase, premio Nobel 1991 per l'economia, disse saggiamente "If you torture data long enough, Nature will always confess".

È inoltre frequente, come già ricordato precedentemente, che i dati siano stati raccolti con scopi diversi da quelli delle analisi che vengono effettuate sugli stessi, come nel caso delle aziende che registrano prevalentemente dati contabili o gestionali per poi soddisfare obiettivi commerciali o di marketing. In tali circostanze è necessario verificare la validità degli elementi raccolti e degli strumenti statistici e l'effettiva possibilità di riciclaggio dei dati, per esempio dimostrando che possano essere un campione rappresentativo anche del nuovo fenomeno.

Tuttavia, per poter sfruttare le interessanti opportunità dell'analisi dei *Big Data*, bisogna affrontare ulteriori problemi non trascurabili. In primo luogo la numerosità di dati, ossia il numero di casi o unità statistiche su cui si effettuano le rilevazioni, rendono inutili gli strumenti di analisi semplici, come i diagrammi di dispersione, e complica gli aspetti computazionali (talvolta il flusso di dati è continuo, dunque la numerosità può anche essere infinita). In secondo luogo la dimensionalità dei dati, ossia il numero di caratteristiche o variabili rilevate su

ciascuna unità, aggiunge complessità non solo di calcolo, ma anche di concettualizzazione e interpretazione di eventuali modelli (Azzalini, Scarpa, 2004).

Fra gli strumenti adatti a svolgere adeguatamente questa fase di analisi approfondita troviamo sicuramente i *Data Analytics*, in particolare gli *Advanced Analytics*.

3. Il contesto di riferimento

Oggi le aziende hanno un'esigenza crescente di estrarre *insight* dall'analisi di enormi quantità di dati di tipologie differenti, andando oltre quella di *query*¹³ e *reporting*¹⁴, garantita dalla più tradizionale *Business Intelligence*. Analizzare una mole di dati maggiore consente di estrarre più informazioni rispetto a quelle che si potrebbero ottenere da piccoli *dataset*. Per questo motivo, la domanda di *Advanced Analytics* è in aumento e il tasso di crescita continua a salire, superando quello della *Business Intelligence*, il cui valore di mercato è però di gran lunga superiore (Gartner, 2016).

Possiamo osservare un analogo trend in Italia: nel 2015 il mercato degli *Analytics* valeva complessivamente 790 milioni di euro, di cui l'84% era relativo alla *Business Intelligence* e il 16% agli *Advanced Analytics*, ma, mentre per i secondi era prevista una crescita del 34%, per la prima questa si arrestava all'11% (Osservatorio Big Data Analytics & Business Intelligence, 2015). Nel 2017 il valore complessivo di mercato dei *Data Analytics* era di 1,1 miliardi euro, grazie a una crescita del 22% nell'ultimo anno. Ad essa hanno contribuito maggiormente le grandi imprese (87% della spesa totale), ma gli investimenti delle PMI in questo campo sono comunque aumentate del 18%. Il 42% della spesa per i *Data Analytics* è rivolta ai software (database, strumenti e applicativi per acquisire, visualizzare e analizzare i dati), il 33% ai servizi (personalizzazione dei software, integrazione con sistemi informativi aziendali e riprogettazione dei processi), il 25% alle infrastrutture abilitanti (capacità di calcolo, server e storage). Anche in Italia cresce tale mercato e la consapevolezza delle aziende delle opportunità offerte: molti CIO vedono la *Business Intelligence* e gli *Advanced Analytics* come le principali priorità di investimento e si orientano sempre più verso strategie guidate dai dati (non solo dall'esperienza o dall'intuizione) e processi e servizi automatizzati. Quasi un'impresa su due, infatti, ha già inserito nel proprio organico uno o più *data scientist*. Tuttavia, il processo di trasformazione delle tradizionali imprese italiane in "*Big Data enterprise*" è ancora lungo: soltanto il 17% ha raggiunto un buon livello di maturazione (contro l'8% del 2016), mentre il 26% si trova in una fase di riconfigurazione dei propri processi organizzativi e il 55% è rimasto legato a un modello organizzativo tradizionale, in cui le singole unità di business analizzano i dati di propria competenza senza una visione aziendale complessiva (Osservatorio Big Data Analytics & Business Intelligence, 2017).

¹³ Interrogazione di un database per estrarne o aggiornarne i dati che soddisfano un determinato criterio di ricerca

¹⁴ La rilevazione periodica dei dati relativi alla gestione e all'andamento di un'azienda

Questi nuovi bisogni hanno portato ad un cambiamento dell'offerta nell'ambito delle piattaforme di *Business Intelligence* e *Analytics*. I fornitori stanno cercando da un lato di incorporare i classici strumenti di reporting, *dashboard*¹⁵ e analisi interattive nei processi di business e dall'altro di inserire gli *Advanced Analytics* all'interno della piattaforma di *Business Intelligence*. Per questo stanno introducendo le capacità di *data-discovery* intelligente, individuazione di correlazioni e integrazione di dati multistrutturati, provenienti da fonti interne o esterne alle aziende, al fine di condurre nuovi tipi di analisi come la *Location*, la *Sentiment* e la *Graph Analysis* (Gartner, 2016).

L'offerta è molto ampia per entrambi i tipi di piattaforma, ma le soluzioni che consentono di estrarre il maggior valore dai *Big Data* sono quelle riguardanti gli *Advanced Analytics*. Dopo aver esaminato le tipologie di strumenti analitici disponibili, prenderò in considerazione il Gartner *Magic Quadrant* corrispondente, uno strumento ampiamente riconosciuto che evidenzia i fornitori più importanti nel mercato di riferimento.

3.1 Tipologie di strumenti offerti

Molte delle aziende incluse nel *Magic Quadrant* forniscono soluzioni o piattaforme in grado non solo di analizzare enormi mole di dati, ma anche di raccogliervi e organizzarli. Ancora prima di poterne estrarre il valore, le aziende devono individuare e adottare tecnologie che permettano di strutturare e ordinare le informazioni e, solo successivamente, implementare strumenti di gestione e analisi delle stesse.

Possiamo quindi dividere gli strumenti disponibili nel mercato dei *Big Data* in tre grandi macro aree di offerta (Forrester, 2015):

- **Aggregatori**

Raccolgono e organizzano i dati, sia aziendali sia soprattutto provenienti da fonti esterne, in modo che possano essere usati dagli utenti business per il loro lavoro. Talvolta vengono aggiunti anche servizi di *Data Management*, pulizia e arricchimento. Rientrano in questa categoria strumenti di *Data Warehousing*, che raccolgono dati provenienti da svariate fonti su archi temporali estesi, li integrano, conservano e distribuiscono come supporto ai processi decisionali; sistemi tradizionali di ETL, che estraggono i dati, li puliscono e trasformano per essere utilizzati; iPaas (*Integration platform as a service*), ossia servizi in cloud per governare e connettere i flussi di dati; tutte le *Descriptive Analytics*, che condensano i dati in pacchetti ordinati, grafici

¹⁵ Una schermata che permette di monitorare in tempo reale l'andamento dei report e dei dati aziendali desiderati

sintetici e indicatori di prestazione. Queste soluzioni sono necessarie soprattutto per le aziende che vogliono entrare in mercati nuovi o sconosciuti, hanno bisogno di gestire varie tipologie di *Big Data* spesso “sporchi” (cioè ridondanti, ambigui, incerti...), fonti diverse, o la cui struttura dati sulla clientela presenta aree scoperte. Tuttavia, ormai sostanzialmente tutte le organizzazioni ne fanno uso.

- **Arricchitori**

Potenziano e completano i dati a disposizione sfruttando elementi di diverse fonti, principalmente notizie e *clickstream* raccolti dal Web e dai social network. Molti strumenti pre-elaborano tali dati per estrapolare informazioni mirate a soddisfare i bisogni dell'azienda. Fra questi si annoverano i software di *Sentiment Analysis*, che effettuano analisi di contenuti e di linguaggio per estrarre informazioni soggettive (opinioni, preferenze, interessi, sentimenti) da documenti o testi; i software che gestiscono e processano dati relativi alla clientela e alle attività commerciali (CRM, *Customer Analytics*); applicazioni analitiche per ottenere informazioni dal web e dai social (*Web, Social e Content Analytics*); le *Geospatial Analytics*, che esaminano le proprietà topografiche, geometriche e geografiche. Questi strumenti vanno utilizzati da chi vuole perfezionare la segmentazione del mercato, la profilazione della clientela e dei competitor o fare marketing diretto con messaggi personalizzati e interagire con specifici clienti.

- **Modellatori**

Applicano ai dati algoritmi e distribuzioni di probabilità che ne evidenziano gli schemi e le correlazioni, con l'obiettivo di poter costruire modelli di previsione e comprendere gli andamenti futuri. Fanno parte di questa classe tutti gli strumenti di *Predictive Analytics*, che utilizzano tecniche statistiche (regressioni, studio di serie temporali, modelli predittivi), e di *Prescriptive Analytics*, che si servono di simulazioni, *machine learning*, reti neurali e analisi grafiche per proporre soluzioni strategiche o operative ottimizzate. Sono adatte a società già abbastanza esperte nell'analisi dei dati, testimoniato dal fatto che, rispettivamente, solo il 73% e il 33% delle imprese ne usufruisce.

Vi sono poi le poco utilizzate *Automated Analytics* (11%), procedure algoritmiche automatizzate in grado di effettuare analisi a partire dall'inserimento di parametri arbitrari e di implementare autonomamente la soluzione risultante. Una di queste è la *Behavioural/Gestural*

Analytics che sfrutta telecamere e sensori per tracciare movimenti e gesti nelle attività umane, individuare e comprendere comportamenti e intenzioni.

3.2 Gartner *Magic Quadrant*

Il *Magic Quadrant* è uno strumento grafico che esprime i risultati di una serie di ricerche e report svolte da Gartner. Fornisce un'analisi qualitativa, aggiornata ogni 1/2 anni, di un determinato mercato tecnologico indicandone i trend, la maturità e i principali partecipanti. Di questi ultimi caratterizza il posizionamento competitivo nel settore e ne definisce la strategia nei confronti del consumatore finale. Inoltre fornisce una comparazione fra i punti di forza e debolezza di ciascuna tecnologia esaminata (Gartner, 2018).

In particolare i fornitori vengono valutati sulla base di due criteri: l'abilità di esecuzione, ossia la capacità di rendere la loro vision una realtà di mercato valida che si differenzi dalle altre per *customer experience*, competenze nell'assistenza, qualità del prodotto; la completezza della vision, ossia la capacità di comprendere come le varie forze che agiscono nel mercato possano essere sfruttate per creare valore per l'azienda e per il consumatore. Il grafico è quindi composto di quattro quadranti, rappresentativi delle tipologie dei fornitori, individuati a seconda dei punteggi ottenuti nei due parametri di valutazione:

- ***Leader***

Sono i migliori nel mercato. Hanno un'ottima abilità di esecuzione e sono ben posizionati sugli sviluppi futuri. Normalmente si tratta di aziende mature di grandi dimensioni che hanno la credibilità, le competenze e le risorse finanziarie per guidare i processi di innovazione e di assorbimento delle nuove tecnologie. Non necessariamente, tuttavia, corrispondono alle necessità e ai bisogni di tutti i buyer;

- ***Challenger***

Eseguono bene, sostengono una continua crescita e possono ottenere il controllo di una larga fetta di mercato, ma mancano di una vision di lungo periodo, così come dell'influenza dei leader. Anche in questo caso si tratta di grandi organizzazioni;

- ***Visionary***

Riconoscono i trend di mercato e hanno una vision orientata al cambiamento e all'innovazione che permette loro di soddisfare nuovi bisogni della clientela o di risolvere eventuali problemi, ma non hanno ancora le capacità o le risorse per

mantenere una profittabilità stabile. Sono generalmente piccole aziende in crescita, golosi bersagli di imprese già affermate;

- *Niche Player*

Si concentrano prevalentemente su un piccolo segmento, una nicchia di mercato e sono inferiori rispetto ai competitor in termini di performance e innovazione. Si parla di piccole aziende o grandi aziende che stanno estendendo la propria attività in nuovi settori, tipicamente appena aggiunte al *Magic Quadrant*.

3.3 *Magic Quadrant for Advanced Analytics*

Considerando il settore degli *Advanced Analytics*, l'abilità di esecuzione è alta se i prodotti sono di ottima qualità e forniscono un ampio set di funzionalità, e i servizi sono offerti con professionalità e competenza; un importante aspetto per determinare tali caratteristiche è l'esperienza vissuta dai clienti, positiva se sono soddisfatti del prodotto e della sua integrazione con altri già esistenti. La completezza della vision è elevata se i fornitori riescono a comprendere i bisogni, i desideri dei clienti e a soddisfarli coerentemente alla loro visione, riuscendo anche a seguire l'evoluzione e l'innovazione del mercato (Gartner, 2015).

Per essere inclusi in questo *Magic Quadrant* i fornitori devono però rispettare una serie di requisiti:

- offrire le capacità di *Advanced Analytics* come prodotto unico o come set di prodotti, utilizzabili in diverse aree all'interno dell'organizzazione;
- offrire supporto per le seguenti funzionalità: accesso, preparazione ed esplorazione dei dati; alcune tecniche di *Predictive Analytics* come la regressione lineare e logistica¹⁶, gli alberi decisionali, le reti neurali¹⁷ e le macchine a vettori di supporto¹⁸; una serie di tecniche di *Descriptive Analytics* tra cui il *k-means clustering*¹⁹ e l'analisi della varianza; *testing*, validazione e sviluppo dei modelli;

¹⁶ Particolare regressione lineare utilizzata per le variabili dicotomiche

¹⁷ Modello matematico/informatico di calcolo basato sulle reti neurali biologiche

¹⁸ Un modello di apprendimento supervisionato per la regressione e la classificazione di modelli

¹⁹ Algoritmo che permette di classificare un insieme di oggetti in K gruppi in base ai loro attributi

- aver raggiunto un fatturato globale di almeno 3 milioni di dollari negli ultimi 12 mesi, derivanti da licenze software relative agli *Advanced Analytics*.



Fig.3: Magic Quadrant per le Advanced Analytics 2016 (Gartner, 2016)

Fra i migliori leader di mercato troviamo SAS e IBM mentre, confrontando il grafico con quello del 2015, si può notare come Dell non sia più un *Challenger*. Angoss e Predixion si spostano, rispettivamente, nei *Challenger* e nei *Visionary* dal quadrante dei *Niche Player*. In quest'ultimo quadrante si sono registrati i maggiori movimenti, con fornitori nuovi entranti e altri che sono usciti. Esaminiamo più da vicino un esempio di azienda per ogni quadrante.

SAS

SAS (*Statistical Analysis System*) nasce nel North Carolina come progetto per analizzare ricerche sull'agricoltura. Al crescere della domanda per tale software, viene fondata nel 1976 quella che sarebbe diventata la più grande azienda leader del settore della *Business Intelligence*

e degli *Advanced Analytics* con il 33% della quota di mercato, più del doppio del competitor più vicino. Ha clienti in 149 stati e oltre 83000 installazioni, incluse 96 delle prime 100 aziende secondo la classifica 2017 Fortune Global 500.

Offre soluzioni software per il supporto decisionale rivolte a tutte le aree strategiche dell'azienda e a tutti i settori di mercato. In particolare SAS/STAT permette sofisticate analisi statistiche su qualsiasi ammontare di dati; SAS Analytics Pro fornisce analisi statistiche, report, mappe e grafici che offrono risultati analitici facilmente comprensibili; SAS Enterprise Miner dà modelli descrittivi e predittivi che individuano schemi innovativi e promettenti per migliorare le decisioni; SAS Data Management aiuta a migliorare, integrare e gestire i dati delle imprese; SAS Customer Intelligence si occupa di dare analisi predittive di marketing, di coinvolgere e interagire con il cliente per personalizzarne l'esperienza.

L'ampia gamma di prodotti è di alta qualità, con un elevato grado di scalabilità e integrazione con altri prodotti, rendendo la piattaforma flessibile e stabile (SAS, 2018).

SAP

Fondata nel 1972 da cinque ex-impiegati di IBM a Weinheim, in Germania, SAP (*System Analysis and Programm development*, tradotto dal tedesco) aveva la vision di produrre software di analisi dati in tempo reale. Oggi è un'azienda leader nelle soluzioni informatiche per le imprese, in particolare nel settore degli ERP, con oltre 338000 clienti in più di 180 stati.

I suoi prodotti relativi agli *Analytics* sono: SAP Predictive Analytics che fornisce e gestisce modelli predittivi per anticipare comportamenti e risultati futuri; SAP Analytics Cloud che permette di lavorare in cloud con *Business Intelligence* e strumenti di pianificazione; SAP Analytics Hub che diminuisce l'incertezza, monitora indicatori strategici e ottiene valide informazioni sui clienti; SAP Business Planning and Consolidation che dà capacità di previsioni, budgeting, creazione del business plan nonché di informazioni finanziarie; SAP SuccessFactors Workforce Planning per analizzare e gestire dati sulle risorse umane.

La qualità dei prodotti è garantita da sistemi di *quality management* che assicurano il raggiungimento dei più alti standard, mentre la rete di entità e connessioni copre ogni campo di possibile innovazione (SAP, 2018).

ALTERYX

Alteryx è un'azienda relativamente giovane, nata nel 1997 a Irvine, California, con l'idea di aiutare le organizzazioni a ottenere migliori *insight* dalle proprie attività. Inizialmente si chiamava SRC LLC, ma nel 2010 cambia il suo nome con quello del prodotto più importante,

la prima piattaforma di *Business Intelligence* che consentiva di analizzare e manipolare dati con un'intelligenza spaziale (da cui il nome Alter - modificare - yx - gli assi cartesiani). Opera a livello globale fornendo una piattaforma integrata e scalabile che offre quattro diverse funzionalità: Alteryx Designer combina preparazione, analisi di dati (previsioni, statistiche, spaziali) e *data blending*²⁰ in un'unica interfaccia; Alteryx Server consente la condivisione dei dati e la collaborazione dei team; Alteryx Connect ricerca informazioni, ricava risultati e osservazioni per migliorare i processi decisionali; Alteryx Promote crea e gestisce modelli predittivi velocemente e con facilità.

Questo prodotto regala alle imprese la flessibilità analitica in grado di superare gli ostacoli all'utilizzo di grandi quantità di dati (Alteryx, 2018).

FICO

FICO è un'azienda di software analitici che supporta business in più di 90 paesi tramite l'utilizzo di algoritmi matematici e di *Big Data*. Fornisce numerosi strumenti diretti a gestire il rischio, ottimizzare le operations, soddisfare i regolamenti governativi, combattere le frodi, stabilire relazioni profittevoli con il cliente. Nasce nel 1956 grazie all'investimento dell'ingegnere Bill Fair e del matematico Earl Isaac, i quali credevano fermamente nelle opportunità di miglioramento dei business grazie all'utilizzo dei dati. I prodotti che usano i *Big Data* sono: Data Distillery identifica i dati ad alto valore, li struttura in una roadmap analitica e li arricchisce proponendo il miglior strumento analitico per estrarre il maggior valore possibile; FICO Decisions Management Platform è una piattaforma integrata e scalabile che consente di raccogliere e gestire grandi quantità di dati, connetterli agli strumenti analitici e ricavare importanti informazioni in tempo reale; FICO Marketing Solutions Suite semplifica la creazione di modelli e l'integrazione del flusso continuo di dati dalle transazioni coi clienti al fine di migliorare la *customer experience* e le campagne di marketing personalizzate (FICO, 2018).

²⁰ Processo che raggruppa dati provenienti da sorgenti multiple in un unico *dataset*

4. *Big Data*: sfide e opportunità in azienda

Un nuovo aggiornamento della *Worldwide Semiannual Big Data and Analytics Spending Guide* (International Data Corporation²¹, marzo 2017) aveva previsto che, nel 2017, i ricavi mondiali derivanti dai *Data Analytics* avrebbero raggiunto i \$150,8 miliardi, registrando una crescita pari al 12,4% rispetto al 2016. Gli acquisti di hardware, software e servizi relativi aumenteranno ad un tasso CAGR (*compound annual growth rate*) costante pari a 11,9% fino al 2020, anno in cui i ricavi raggiungeranno la cifra significativa di \$210 miliardi. Tuttavia, secondo le stime di IDC, solo un terzo dei circa 40 zettabyte che verranno prodotti entro il 2020, verrà adeguatamente analizzato e potrebbe portare valore alle aziende.

I dati generati dai clienti e da attori di altro tipo non sono più risorse scarse, e l'accessibilità si è notevolmente semplificata: ciò che creerà il valore differenziale per le aziende saranno il possesso e l'applicazione delle tecniche e delle competenze in grado di aggregare i dati in maniera unica e di estrarre informazioni utili al processo decisionale del business. L'importanza strategica dei dati e degli strumenti di analisi, che consentono di prendere decisioni basate su prove, esperimenti e previsioni accurate anche in tempo reale, li rende uno dei pilastri fondamentali per la trasformazione digitale delle imprese di qualsiasi settore. Nonostante questa sia una conclusione evidente e condivisa, molte faticano a sviluppare una corretta strategia *data-driven*, ampliando il divario fra i leader e i *laggard* (The Age of Analytics, McKinsey, 2017). I problemi principali sono di carattere organizzativo: tale strategia risulta complessa da inserire in un'ottica di attività quotidiana e vi è una carenza significativa di figure professionali, in particolare di *data scientist*, persone in grado di tradurre i risultati ottenuti dalle analisi e applicarli nei processi di business.

La ricerca del 2017 dell'Osservatorio Big Data Analytics & Business Intelligence conferma che il 51% delle aziende rileva la mancanza di competenze e figure organizzative interne come *Chief Data Officer* e *data scientist*. Invece non sembrano essere elementi di freno i software poco usabili o le soluzioni obsolete (12%), così come la capacità di reperire dall'esterno professionalità con l'adeguato mix di competenze (22%). Tra gli altri ostacoli vengono segnalati la necessità di investimenti troppo elevati e la difficoltà nello stimarne i benefici, soprattutto per le aziende che per la prima volta si avvicinano a soluzioni di *Data Analytics* (39%), e le problematiche di security e privacy (27%). Forrester (2015) rileva due ulteriori problemi. In primo luogo bisogna considerare che ormai quasi tutte le organizzazioni possiedono una *Business Intelligence* di supporto alle operations e alla gestione finanziaria, col

²¹ Azienda che fornisce informazioni e statistiche strategiche, report e servizi di consulenza nel settore dell'Information and Communication Technologies

rischio che queste posticipino le analisi predittive al momento in cui è perfettamente dimostrato che l'attuale processo che dalla transazione porta al guadagno sia efficiente. In secondo luogo il management potrebbe non voler rischiare di affrontare un cambiamento importante, poiché potrebbe non essere capace di gestirne le conseguenze sulla cultura organizzativa e sui nuovi metodi operativi.

Nelle piccole realtà l'analisi dei dati rimane una prerogativa dell'IT o un argomento complesso da dover richiedere in modo sistematico la consulenza di società esterne specializzate.

Una volta acquisita consapevolezza sul ruolo e l'importanza che i *Big Data* e i modelli analitici ricoprono per il proprio business, le aziende dovranno implementare una *data strategy* condivisa da tutte le business unit. SAS la definisce come un piano per migliorare il modo in cui i dati sono acquisiti, memorizzati, condivisi e gestiti (*The five essentials components of a data strategy*, SAS, 2018). Tale progetto si compone di 5 passi da seguire:

1. Identificare e catalogare i dati, indipendentemente dalla loro struttura;
2. Memorizzare i dati in strutture con spazi di archiviazione adeguati e che consentano un facile accesso;
3. Organizzare i dati in modo che possano essere condivisibili e creare delle linee guida per un utilizzo ordinato;
4. Trasformare i dati grezzi in dati pronti da usare, anche in futuro e più volte;
5. Stabilire e comunicare regole e meccanismi per gestire i dati trasversalmente all'azienda, non solo dove vengono generati.



Fig.4: Le cinque componenti fondamentali della data strategy (SAS, 2018)

Le aziende che per prime sapranno affrontare e risolvere queste sfide saranno in grado di differenziarsi e ottenere un vantaggio competitivo sui propri concorrenti di settore. Uno studio svolto dall'Università di Verona, in collaborazione con la CUOA Business School e la Karlsruhe Institute für Technologie, ha dimostrato, infatti, che in settori competitivi, con alta dinamicità in termini di cambiamenti nei bisogni dei clienti e delle tecnologie, l'accesso differenziale alle informazioni può portare ad un vantaggio competitivo sostenibile: utilizzando

informazioni prodotte in tempo reale nei processi decisionali, è possibile capire e reagire tempestivamente ai cambiamenti esterni ottenendo performance superiori. Poiché i *Data Analytics* possono essere sfruttati per supportare diverse altre attività interne, il conseguente impatto diretto e positivo può essere generalizzato, indipendentemente da quello sulla capacità di risposta delle organizzazioni. Dai risultati ottenuti nello studio dell'Osservatorio si evince che la totalità delle imprese intervistate ha migliorato l'engagement con il cliente, il 91% ha incrementato le vendite, il 78% ha ridotto il tempo che intercorre fra l'ideazione e commercializzazione del prodotto, il 73% ha ottimizzato l'offerta per aumentare i margini di guadagno, il 67% ha ampliato l'offerta di prodotti e servizi, il 56% ha contenuto i costi e il 38% ha esplorato nuovi mercati. Quasi due su tre danno un giudizio positivo: in particolare, il 13% giudica i risultati *disruptive*, il 21% un grande successo, il 29% un moderato successo, mentre soltanto il 37% ritiene che sia prematuro esprimere una valutazione e nessuna considera un fallimento questo tipo di iniziative.

I prossimi paragrafi sono quindi dedicati a esplicitare quali siano le applicazioni economiche più significative per l'analisi dei *Big Data* in ambito aziendale e a metterle in relazione con i settori industriali e le aree funzionali interne. Infine presenterò alcuni casi empirici di imprese che hanno utilizzato con successo determinate tipologie di *Data Analytics*.

4.1 I settori industriali

L'innovazione dei *Big Data* può coinvolgere e generare valore, potenzialmente, in qualsiasi settore, benché le opportunità e le sfide che presenta variano notevolmente. Alcuni ne hanno estratto un maggior valore o sono risultati più veloci nell'adottare le tecniche adeguate, altri invece faticano a tenere il passo con i repentini cambiamenti delle tecnologie e hanno ancora ampi margini di miglioramento.

È interessante evidenziare quali siano le differenze settoriali delle applicazioni economiche di *Data Analytics* messe in atto.

APPLICAZIONI SETTORI	Ottimizzazione prodotti	Customer intelligence	Financial management	Gestione risorse umane	Security	Forecasting	Mass-customization	Predictive maintenance	Brand reputation
Manifatturiero	X					X		X	X
Banche e assicurazioni		X	X	X	X		X		
Telco e media		X			X		X	X	X
Sanità e farmaceutico	X	X	X	X					
Energia e utenze	X	X				X	X	X	
PA e forze dell'ordine				X	X				X
Automotive	X	X						X	X
Trasporti e infrastrutture	X			X		X		X	X
Retail e consumer goods	X	X		X		X	X		X

Tabella 1: Applicazioni economiche di Data Analytics in alcuni settori industriali

- I settori presenti in tabella sono quelli in cui i *Big Data* sono maggiormente utilizzati, o quelli in cui hanno più impatto
- Le applicazioni elencate sono quelle tecnologicamente più mature e più frequentemente adottate dalle aziende
- Ottimizzazione prodotti (in senso lato, ossia i prodotti o i servizi che una azienda offre): tecniche che migliorano la gestione dei prodotti, tramite tempestive verifiche della qualità (*quality assurance*) o adattamenti del prezzo a determinate variabili in tempo reale (*dynamic pricing*)

- *Customer intelligence*: tutte le analisi effettuate sui clienti al fine di migliorarne l'esperienza d'acquisto, mantenerne la fedeltà, quantificarne il valore e ricavarne informazioni
- *Financial management*: si occupa di predire i rischi connessi ad azioni o eventi (*risk modeling*) e di valutare i benefici di investimenti e operazioni finanziarie secondo determinati parametri
- *Security*: tutti gli strumenti di protezione dei dati e d'identificazione di frodi o attività illegali
- *Forecasting*: le attività che ricercano nuove possibilità in termini di mercati, business, innovazioni di prodotti o servizi ed effettuano previsioni sull'andamento della domanda
- *Mass-customization*: attività di segmentazione della clientela, vendita di prodotti diversi o di qualità superiore (*cross e up-selling, recommendations*), micromarketing

Notiamo come il settore che sfrutta il maggior numero di applicazioni sia quello del retail e dei prodotti consumer con sei applicazioni. Seguono con cinque banche e assicurazioni, telecomunicazioni e media, energia e utenze, trasporti e infrastrutture.

Le applicazioni più diffuse (in sei settori) sono invece: ottimizzazione prodotti, soprattutto nei settori orientati al prodotto fisico; *customer intelligence*, in settori in cui la relazione col cliente porta maggior valore; *brand reputation*, dove la forza della marca traina le vendite. Appena dietro (presenti in cinque settori) la gestione risorse umane e la *predictive maintenance*, rispettivamente nelle industrie *human* e *capital intensive*. Riguardo al *financial management* e al *forecasting*, si tratta evidentemente di attività che tutte le imprese svolgono, ma ho voluto segnalare solo i casi in cui sono effettuate con maggiore complessità e frequenza.

MANIFATTURIERO

Nel settore manifatturiero la combinazione tra *Internet of Things* e *Data Analytics* potrebbe diminuire notevolmente i costi, aumentare la vita utile degli asset e la loro produttività, con un evidente incremento di profitti. I sensori intelligenti connessi a Internet e i software che consentono la comunicazione fra le macchine generano grandi quantità di dati sullo stato e le condizioni ambientali degli stabilimenti e dei macchinari. Dalle analisi si individuano i potenziali guasti prima che accadano, per effettuare così delle manutenzioni predittive o riparazioni in tempo reale (cioè quando ce n'è effettivamente bisogno). In questo modo si abbattano i costi delle ispezioni e del magazzino delle parti di ricambio; si riduce l'incertezza, poiché i componenti o i macchinari che presentano schemi di malfunzionamento casuali non

provocano più interruzioni del processo produttivo; si migliora la fase post vendita con la fruizione di servizi di riparazione intelligente (Dimension Data²², 2016). In questo ambito il futuro sarà la costruzione di un sistema di macchine che si auto monitora e gestisce, prevedendo potenziali problemi di prestazione e programmando le manutenzioni autonomamente.

Tra le attività di *forecasting* troviamo la previsione dell'andamento della domanda, in modo da coordinare la produzione alle quantità richieste dal mercato e la ricerca di nuovi prodotti sulla base delle nuove esigenze dei clienti. Le tecniche di *financial management* vengono impiegate per valutare diversi progetti di investimento in nuove tecnologie o macchinari, dei quali viene calcolata la profittabilità tramite l'indice finanziario *Return On Investment*.

BANCHE E ASSICURAZIONI

I *Big Data* sono molto importanti nel settore finanziario, non solo per la valutazione degli investimenti, dei mercati azionari o dei tassi d'interesse, ma anche per l'identificazione di attività di riciclaggio di denaro, la rapida risoluzione dei reclami e l'investigazione di frodi. La compagnia assicurativa sudafricana Santam ha implementato un sistema che classifica tutti i reclami secondo cinque categorie che spaziano dalla semplice lamentela fino ai casi di frode, separando i casi a basso rischio da quelli ad alto rischio. È riuscita così a migliorare notevolmente la rapidità di risoluzione di questi problemi, nonché a riallocare meglio le sue risorse e i suoi agenti, diminuendo del 6% il premio annuale dei propri clienti, un aumento che prima era stato necessario per far fronte a questi atti illegali (*Analytics: the real world use of Big Data*, IBM, 2012).

Le banche e le assicurazioni si servono anche di tecniche di segmentazione della clientela per offrire pacchetti e prodotti migliori nel momento ottimale, basandosi sul *Customer Lifetime Value*, sulla scoperta di schemi di comportamento di particolari segmenti e sulle singole esigenze d'investimento.

TELECOMUNICAZIONI E MEDIA

A causa di un'aspra competizione dovuta all'abbattimento dei costi di fornitura e a player come Skype e Whatsapp che consentono di chiamare e mandare messaggi a costo zero, l'utilizzo di *Data Analytics* in questo settore si focalizza soprattutto sul fornire servizi individuali ai singoli clienti, specializzando l'offerta per migliorare la profittabilità della relazione per entrambe le

²² Azienda sudafricana specializzata in servizi di *information technology*

parti. Netflix, per esempio, ha creato 10 trailer di House of Cards indirizzati a diversi tipi di audience, assegnati a ciascun consumatore sulla base delle proprie visioni precedenti. Si ricorre poi ad analisi dei social network e del web per ottenere informazioni sulla propria reputazione, far leva sulle qualità gradite dai consumatori per mantenerne la fedeltà e individuare e risolvere eventuali problemi o insoddisfazioni per evitare abbandoni (*churn prevention*).

La raccolta delle informazioni sulla clientela e sulle loro abitudini d'acquisto è fondamentale per ricavare importanti feedback, ma si scontra con delicate questioni di privacy di cui tutte le aziende del settore dovranno tenere conto.

SANITÀ E FARMACEUTICO

I costi sempre maggiori e la necessità riconosciuta di una migliore e più ampia copertura sanitaria a livello pubblico rendono le tecnologie di *Big Data* indispensabili. McKinsey aveva rilevato nel 2011 che il settore della sanità degli Stati Uniti avrebbe potuto crescere di più di 300 miliardi di dollari ogni anno, ma un successivo report del 2016 evidenzia come solo il 20% delle opportunità individuate sono state sfruttate. Lo studio di Zillner et al. (2014) afferma che il problema principale da risolvere è quello di raccogliere e rendere disponibili dati che provengono da numerose fonti, spesso in forme non strutturate, come dati biometrici, immagini di esami medici (radiografie, ecografie ecc.), ricerche farmaceutiche, dati di monitoraggio dei pazienti, database che raccolgono dati statistici su malattie e trattamenti, dati amministrativi e contabili. La *Customer Intelligence* in questo caso si declina nell'ottenere informazioni sulle caratteristiche dei pazienti da tutte queste sorgenti, in modo da supportare le decisioni e le diagnosi dei medici.

Il *financial management* invece si occupa di verificare quali siano i migliori dispositivi medici da acquistare mentre la qualità del servizio medico viene garantita tramite una continua analisi delle procedure e delle performance al fine di tagliare gli sprechi, ottimizzare i tempi e automatizzare i processi.

ENERGIA E UTENZE

I dati provengono da grandi quantità di sensori, contatori, impianti, raffinerie, pompe. Anche in questo settore la *predictive maintenance* ha un impatto significativo sui costi delle aziende grazie alle contromisure prese in tempo reale per verificare e riparare anomalie nel servizio. Importante è pure l'attività di previsione del fabbisogno energetico di ciascun cliente e del meteo, così da poter ottimizzare le operations e la produzione. Tuttavia, ciò che permette di crescere e ottenere ampie quote di mercato è la gestione della *customer experience*, poiché i

margini stanno diminuendo e le barriere all'entrata si intensificano (Rusitschka, Curry, 2016): sulla base del livello di consumo e una dettagliata segmentazione demografica o geografica si possono creare tariffe precise adatte alle esigenze di ciascun cliente; attraverso continue innovazioni che migliorano l'erogazione del servizio, anche dal punto di vista della sostenibilità ambientale, la fedeltà del consumatore viene rafforzata.

PUBBLICA AMMINISTRAZIONE E FORZE DELL'ORDINE

L'amministrazione governativa può individuare più facilmente eventuali frodi, evasioni fiscali o errori di inserimento, ma l'applicazione più interessante si ha in ambito politico: la campagna presidenziale del 2008 di Barack Obama ha mostrato come il web e le piattaforme multimediali abbiano supportato con successo la sua elezione. In questo caso non si parla di *brand reputation*, ma il concetto alla base è sempre lo stesso: effettuare ricerche sui social, su blog di discussione politica, sulle campagne promozionali online per ottenere informazioni sulle opinioni e i sentimenti dei propri elettori (Chen et al., 2007).

I *Data Analytics* consentono inoltre alle forze dell'ordine di identificare schemi di attività illegali, ottenere informazioni accurate su criminali effettivi o potenziali, aiutare a stanare eventuali fuggitivi rintracciandone le azioni, la posizione e le connessioni con altre organizzazioni criminali. Un altro aiuto importante riguarda la gestione delle risorse umane per far fronte alle sempre più frequenti richieste di intervento, con pressanti esigenze di rapidità, e al sempre minor numero di poliziotti in servizio sul campo (Wyllie, 2013).

AUTOMOTIVE

La qualità di una macchina è fondamentale non solo per una questione di prestazione o di estetica, ma anche per la sicurezza dell'utilizzatore: negli stabilimenti produttivi si usano quindi sofisticati sistemi di analisi che garantiscono il raggiungimento e mantenimento di elevati standard qualitativi collegati a software che comunicano eventuali anomalie su cui intervenire.

Si cerca sempre più, inoltre, di innovare e reinventare l'esperienza del consumatore attraverso la personalizzazione delle funzionalità e degli accessori di ciascuna automobile, creando app per la gestione da remoto di varie attività (per esempio trovare parcheggio, gestire l'allarme o l'illuminazione di casa) o per il car-sharing (*Driving digital destiny: digital reinvention in automotive*. IBM, 2017).

TRASPORTI E INFRASTRUTTURE

Con l'utilizzo di tecnologie IoT e di *Data Analytics* si possono creare le cosiddette *Smart Cities*. In esse, ad esempio, i flussi del traffico automobilistico e del trasporto pubblico vengono controllati in tempo reale, utilizzando dati provenienti da sensori magnetici e telecamere sui semafori o su altri supporti, da social media e dalla meteorologia, diminuendo così le congestioni; i cestini e i cassonetti vengono svuotati solo quando sono effettivamente pieni, grazie ad appositi sensori che lo segnalano (Marr, 2018).

Per quanto riguarda i trasporti, le compagnie aeree sfruttano il *dynamic pricing* per ridurre al minimo il surplus del consumatore e massimizzare il profitto per ogni biglietto, modulandone il prezzo sulla base di informazioni quali le condizioni atmosferiche, il carburante necessario, i giorni di permanenza del cliente, il numero di posti già assegnati, l'anticipo nell'acquisto.

Sempre più frequenti sono anche le attività di *forecasting* per la ricerca di nuovi business con l'obiettivo di riciclare e riutilizzare i dati ottenuti sulla mobilità, vendendoli, per esempio, a terze parti.

RETAIL E PRODOTTI CONSUMER

In questi settori il focus principale è la relazione col cliente. A questo proposito tre sono gli ambiti principali per l'analisi di *Big Data*: risorse umane, brand, prodotti.

Il personale di vendita è indispensabile per migliorare l'esperienza del consumatore e non può (ancora) essere sostituita dalle macchine, grazie alle competenze e ai servizi di consulenza o di semplice supporto nella scelta del prodotto; vengono così regolati i turni e il numero di persone in negozio in maniera ottimale, analizzando dati storici dei ricavi, orari di apertura dei negozi (anche dei concorrenti), previsioni meteorologiche, festività ed eventi, abitudini di acquisto.

Il brand deve comunicare determinati messaggi e indurre un'identificazione nel cliente con l'obiettivo di conquistarne la preferenza e la fedeltà (Kotler et al., 2015); le aziende ricorrono quindi all'analisi di social network, social media, siti web, quotidiani, riviste per analizzare le opinioni e i sentimenti dei consumatori con cui valutare il valore e la reputazione della marca.

Infine l'attenzione al prodotto è massima: dall'ottimizzazione dei prezzi alla ricerca di nuovi prodotti e nuovi trend, dalla gestione dell'assortimento a quella del design e del *visual merchandising*. Altri elementi che curano l'esperienza del cliente sono le attività di micromarketing, l'omnicanalità, le *recommendations* sui prodotti acquistati, le email indirizzate personalmente.

4.2 Le aree funzionali

Vediamo ora, attraverso un esame delle principali funzioni aziendali, in che modo si può usufruire dei *Big Data* per migliorare le performance e i processi decisionali. Nello studio dell'Osservatorio (2015) la funzione aziendale che utilizza maggiormente soluzioni di *Analytics* è quella del marketing e vendite, seguita da amministrazione, finanza e controllo, produzione e operations, logistica, risorse umane mentre si colloca ultima quella di ricerca e sviluppo.

MARKETING E VENDITE

L'utilizzo di alcuni tipi di *Analytics* in questa area funzionale potrebbe apportare un valore significativo alle performance aziendali. In particolare le *Geospatial Analytics* permettono il marketing mobile, con il quale si comunica col consumatore attraverso i dispositivi elettronici per la connettività mobile: in prossimità dei punti vendita le imprese inviano messaggi promozionali, video pubblicitari, sconti per invogliarli alla visita e all'acquisto. Le *Behavioural Analytics* agiscono invece al loro interno facendo uso di telecamere e sensori per tracciare e analizzare il comportamento dei clienti (per esempio il percorso nel negozio o il movimento degli occhi e delle mani sugli scaffali), al fine di disporre al meglio l'assortimento dei prodotti. *Web e Social Analytics* si occupano di fare operazioni di *crawling* su siti internet, social media e social network per supportare le analisi sia delle opinioni e dei sentimenti sul brand e suoi prodotti che dei desideri e delle lamentele della clientela, con l'obiettivo di soddisfarli rapidamente.

A queste analisi le aziende aggiungono altri dati sui consumatori, per ottenerne una visione a 360 gradi e profilare e segmentare il proprio mercato in maniera molto precisa:

- Dati demografici quali età, sesso, etnia;
- Dati psicosociali quali occupazione, reddito familiare, atteggiamenti e credenze;
- Dati comportamentali che riguardano il processo decisionale del consumatore, passate transazioni e il potere d'acquisto.

I benefici maggiori di questa visione così completa si hanno soprattutto per il potenziamento del *Customer Relationship Management* che Kotler et al. (2015) definiscono come un'attività che gestisce le ricche e dettagliate informazioni sui singoli clienti con lo scopo di creare e consolidare relazioni profittevoli, massimizzando la fedeltà. Altro vantaggio si ha nel marketing personalizzato raggiungendo forme di *mass-customization* molto complesse in cui le

offerte sono indirizzate in modi diversi a ciascun cliente: sulla base delle informazioni raccolte, che esprimono interessi e abitudini del consumatore, si possono proporre prodotti, promozioni e sconti personalizzati.

Queste considerazioni valgono anche per l'*account management*, ossia per i rapporti di vendita B2B: analizzando le operazioni precedenti e integrandole con informazioni su ciò che fanno i clienti al di fuori della transazione (fra cui fusioni, acquisizioni, investimenti, assunzioni, problemi legali o finanziari) si possono ottimizzare le offerte, incrociando gli obiettivi e le esigenze di entrambe le parti.

AMMINISTRAZIONE, FINANZA E CONTROLLO

Questa funzione, nello specifico, analizza investimenti, debiti, transazioni con i clienti o i fornitori per ridurre al minimo i rischi operativi e finanziari; pianifica il budget raccogliendo informazioni sui costi e le necessità di investimento delle altre aree aziendali; monitora continuamente i livelli di performance mediante precisi KPI in modo da non allontanarsi dalla strategia dell'organizzazione, grazie alla creazione di *dashboard* e *scorecard*²³ dinamiche.

In generale lo sfruttamento dei *Big Data* rende i flussi informativi snelli, fruibili con semplicità e trasparenti, portando grande valore all'azienda e agli stakeholder.

PRODUZIONE E OPERATIONS

In questo ambito le analisi di *Big Data* hanno l'obiettivo principale di massimizzare l'efficienza del processo produttivo, ossia minimizzare i costi attraverso un'ottimale allocazione delle risorse. Innanzitutto sono utilizzati strumenti di *Predictive Analytics*, che esaminano dati del sistema informativo aziendale, dati di mercato e dati di partner o altri stakeholder, per rendere prevedibile lo sviluppo dell'andamento della domanda, cercando di comprendere i trend e le propensioni di acquisto dei consumatori. Sulla base del piano di produzione si può poi regolare la strategia di spesa, sia in termini di riduzione di sprechi che di selezione dei migliori fornitori grazie ad un'analisi della loro reputazione e delle loro precedenti prestazioni.

All'interno del vero e proprio processo produttivo si può intervenire con l'analisi dei dati generati dalle macchine per svolgere manutenzioni predittive e individuare eventuali colli di bottiglia. In generale gli strumenti di *Data Analytics* automatizzano e ottimizzano diverse

²³ Strumento di supporto alla gestione dell'impresa che permette di tradurre la missione e la strategia in un insieme coerente di misure di performance, facilitandone la misurabilità

operazioni, alleggerendo notevolmente il carico di lavoro manuale, con evidenti risparmi in termini di risorse umane.

LOGISTICA

La funzione logistica ha come obiettivi la riduzione dei costi e il miglioramento della gestione del magazzino e delle spedizioni. Per risolvere il primo problema si possono effettuare analisi dei comportamenti di acquisto sulla base di studi demografici, previsioni meteorologiche, eventi o festività in modo da aggiustare i livelli delle rimanenze. Inoltre, grazie alla *predictive maintenance*, i pezzi di ricambio sono ridotti al minimo consentendo risparmi di spazio, affitto e gestione. Per quanto riguarda le spedizioni, i partner commerciali possono distribuire informazioni sul traffico globale di prodotti e sulle rotte di trasporto. Le *Geospatial Analytics* assolvono questo compito incrociando i dati territoriali e climatici con quelli relativi a incidenti, congestioni, zone di ZTL o centri storici per disegnare i percorsi più veloci. Il cliente può infine verificare in ogni momento il luogo e la tempistica della spedizione (Jeske, Grüner, Weiß, 2013).

RISORSE UMANE

Le aziende sfruttano le *Workforce Analytics* per analizzare approfonditamente non solo i costi delle proprie risorse umane, ma anche raccogliere tutta un'altra serie di informazioni come il tipo di contratto, le competenze professionali, i giorni di malattia, le caratteristiche psicologiche e fisiche. Tali dati vengono successivamente aggregati e visualizzati in grafici o statistiche di facile interpretazione e correlati alle job description di ciascuna posizione lavorativa per verificare le performance e sopperire ad eventuali mancanze (Hota, Ghosh, 2013).

Si hanno interessanti vantaggi anche nella fase di reclutamento poiché si possono calcolare tutti gli indici standard, come il tasso di abbandono, di assenteismo, di costo per assunzione e così via, e integrare le informazioni interne con quelle provenienti dal mercato del lavoro. Esistono anche aziende specializzate proprio nel reclutamento online che fungono da piattaforme di incontro fra i lavoratori e le imprese, aggregando dati di entrambe le parti per consentire l'assunzione delle migliori figure nelle posizioni più adatte a ciascuna di esse.

RICERCA E SVILUPPO

La ricerca e sviluppo vuole usare i *Big Data* per dare vita a innovazioni, anche radicali, in un'ottica di riduzione del Time to Market²⁴, fondamentale soprattutto nei settori in cui i prodotti e le tecnologie diventano obsoleti rapidamente. Si tratta di individuare preventivamente quali siano i driver alla base dell'evoluzione del mercato di riferimento, raccogliendo informazioni dai social media e dai desideri espressi (o non espressi) dei consumatori, al fine di sviluppare azioni di adeguamento o nuovi modelli di business che mirino a garantire un vantaggio competitivo sostenibile. Tali ricerche possono essere indirizzate anche al di fuori del proprio mercato, per esempio ad un'altra nicchia di clienti o anche verso settori completamente diversi.

Oltre che i nuovi punti di arrivo, la rivoluzione riguarda anche i punti di partenza. Le imprese possono condurre progetti di ricerca esclusivamente all'interno dell'organizzazione (*close innovation*), costosi da gestire e mantenere ma su cui si hanno il controllo e la segretezza; oppure possono adottare il concetto di *open innovation*, ossia acquisire idee innovative che provengono dalla comunità, sostanzialmente gratuite, a parte per un compenso in denaro, o esternalizzare quelle prodotte internamente (Chesbrough, 2011).

Attraverso il *crowdsourcing* idee, prototipi, prodotti e scenari sono discussi, sviluppati e aggiornati dalla comunità dei partecipanti che confrontano dati storici e attuali. La piattaforma online Kaggle, per esempio, fa da collegamento tra le imprese e le persone interessate alle competizioni che l'azienda mette in atto: le prime forniscono i *Big Data* e determinate richieste su cui le seconde lavorano con *Data Analytics* per soddisfarle (Marr, 2015).

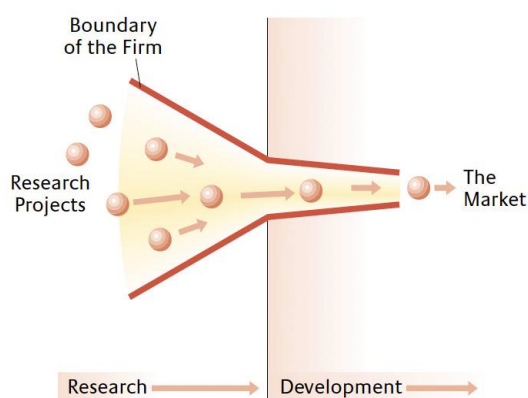


Fig.5: *Close innovation* (Chesbrough, 2011)

²⁴ Tempo che intercorre tra il concepimento di un prodotto e la sua effettiva disponibilità di commercializzazione

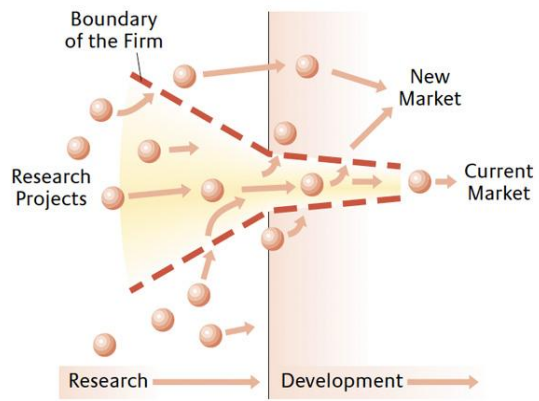


Fig.6: Open Innovation (Chesbrough, 2011)

4.3 Casi di studio

A sostegno della parte teorica, tratterò ora casi empirici interessanti di aziende che hanno sfruttato con successo particolari applicazioni dell'analisi di *Big Data*.

LA PREDICTIVE MAINTENANCE IN PORSCHE

Porsche è tra le principali case automobilistiche del mondo, specializzata in auto sportive di alto livello e di proprietà del gruppo Volkswagen, con un fatturato di €23,49 miliardi nel 2017. La sua strategia per il 2025 è incentrata sul ridefinire il proprio portfolio prodotti all'insegna di innovazioni tecnologiche, della digitalizzazione e connettività, della sostenibilità ambientale e elettromobilità, pur rimanendo fedeli alla storia e ai valori dell'azienda (Porsche, 2018).

Certamente un recente progetto che va in questa direzione è quello elaborato dal Porsche Digital Lab in collaborazione con la start-up high-tech iNDTact, nel febbraio 2018. La *predictive maintenance* è una componente fondamentale per una trasformazione digitale in ottica di Industria 4.0, così hanno sviluppato un sistema innovativo con un design modulare costituito da sensori solidi e a banda larga, sensibili a rumori, pesi, vibrazioni di ciascun elemento dell'automobile. Un'intelligenza artificiale analizza i dati registrati per individuare anche le più insignificanti anomalie o deviazioni rispetto a un comportamento standard. Svolgendo test di carico e sforzo e determinate analisi tecniche sui segnali ricevuti è in grado, inoltre, di identificare le probabili cause delle irregolarità. Grazie ai continui report è possibile ridurre al minimo le manutenzioni preventive e affidarsi invece a manutenzioni predittive, ossia svolte in caso di effettiva necessità per evitare danni imminenti: l'azienda ritiene che le ispezioni di qualità automatizzate possano ridurre fino al 60% le risorse necessarie, aumentare la qualità del prodotto fino al 20% e incrementare la produttività del 30% (grazie ai minori tempi d'interruzione per malfunzionamenti). Prima del *machine learning* e delle *Advanced Analytics*, infatti, le ispezioni richiedevano investimenti costosi in attrezzatura, lunghi tempi di setup e riavvio e minor precisione a causa della prevalenza di lavoro manuale (Porsche Consulting, 2017).

Questi risparmi si traducono in una migliore qualità del prodotto, in termini di minor peso, riduzione del rischio di guasti e in generale migliori prestazioni di guida, a totale beneficio del consumatore finale.

LE RECOMMENDATIONS DI NETFLIX E AMAZON

Il gigante del retail online Amazon ha fatturato il 35% delle sue vendite totali grazie al sofisticato algoritmo di *recommendations* integrato lungo tutto il processo di acquisto, dalla ricerca e selezione del prodotto fino al pagamento finale. Tale sistema analizza una serie di informazioni come gli storici degli acquisti dell'utente, i prodotti nel carrello e quelli valutati in maniera positiva, ciò che altri clienti hanno comprato o visualizzato. Elabora una lista di merce che potrebbe interessare con buona probabilità per effettuare campagne di *up* e *cross selling*, aumentando così la percentuale di conversione in acquisto e il suo valore medio. L'algoritmo riesce anche a superare due grossi problemi che un retailer di grandi dimensioni deve affrontare: per prima cosa, essendo il numero di prodotti altissimo e così pure quelli consigliati, è necessario capire quali di questi mostrare all'utente; successivamente servirà trovare un modo per ordinarli, mettendo in alto quelli che potrebbero essere maggiormente preferibili (Arora, 2014).

Anche Netflix, azienda leader nella distribuzione di film e serie tv via Internet con 125 milioni di iscritti, si basa molto sulle *recommendations*. Ha calcolato che circa il 75% delle visualizzazioni sono guidate da queste ultime e che permette di risparmiare circa 1 miliardo di dollari all'anno (Vanderbuilt, 2013). L'algoritmo in questo caso analizza le caratteristiche dei media visti (genere, attori, registi ecc.), lo storico dello streaming, le valutazioni inserite da tutti gli utenti con gusti simili e assegna una percentuale di correlazione per indicare la possibile preferenza. Questi dati vengono inseriti in un contesto, per esempio indicando il giorno e l'ora della settimana della visione, il dispositivo utilizzato, quante volte è stato rivisto o è stato mandato indietro/avanti.

Grazie a questo tipo di *Analytics* si migliora l'esperienza e la soddisfazione del consumatore perché tale personalizzazione mostra prodotti che piacciono ai consumatori e che altrimenti non sarebbero in grado di scoprire (Bulygo, 2018).

SENTIMENT ANALYSIS IN MACY'S

Macy's, importante dettagliante americano nel settore fashion, ha implementato una nuova tecnologia in partnership con SAP per conoscere meglio le preferenze e le abitudini di acquisto dei propri clienti. Il software caratterizza gli interessi di questi ultimi con dati relativi alla stagionalità, a caratteristiche demografiche, al colore, alla forbice di prezzo per poi effettuare una *Sentiment Analysis* che valuta i sentimenti positivi o negativi espressi dai consumatori sui social media e social network. I risultati ottenuti non riguardano solo la soddisfazione o meno

su un prodotto specifico, ma anche la comprensione di quali brand inserire o eliminare dal negozio e l'invio di e-mail personalizzate. Per esempio, l'azienda ha scoperto che coloro che condividono tweet sul termine "giacca", includono riferimenti a marche come Michael Kors o Louis Vitton. Questo investimento si è tradotto in un incremento dall'8 al 12% delle vendite online (Fiorletta, 2014).

Vale la pena citare anche l'analisi che Voices from the Blog²⁵ ha eseguito in occasione del lancio dell'iPad del 2012, anche se non è stata effettivamente utilizzata da Apple. La *Sentiment Analysis* può infatti essere integrata con *l'Opinion Analysis*, che va a individuare quali sono le motivazioni dietro ai sentimenti rilevati. La start-up, analizzando 40000 testi postati su Twitter, ha registrato un sentimento positivo sull'acquisto del prodotto (+52%), le caratteristiche più gradite (processore, retina display, nuove app) e le cause di insoddisfazione (nome, dock, porta USB). Emerge quindi che questo tipo di risultati possono supportare le campagne di marketing sui prodotti, individuandone punti di forza e di debolezza (Ceron, Curini, Iacus, 2014).

²⁵ Start-up innovativa nell'ambito *Big Data e Data Analytics*, nata dal Politecnico di Milano

5. Il tema della privacy

La rapidità dell'evoluzione tecnologica e la globalizzazione comportano nuove sfide per la protezione dei dati personali²⁶ che, nel mare dei *Big Data*, sono preda ambita delle organizzazioni. L'integrazione economica e sociale degli ultimi anni ha condotto a un considerevole aumento dei flussi di dati personali scambiati tra vari attori pubblici e privati, comprese persone fisiche, associazioni e imprese. Grazie alla tecnologia attuale, sia le imprese private sia le autorità pubbliche sono in grado di raccogliere e utilizzare tali dati, come mai in precedenza, nello svolgimento delle loro attività. Sempre più spesso, infatti, le persone fisiche rendono disponibili al pubblico su scala mondiale informazioni che li riguardano, la cui circolazione verso altri paesi o organizzazioni internazionali è ormai libera e semplice.

Tal evoluzione richiede un quadro più solido e coerente in materia di protezione dei dati nell'Unione, affiancato da efficaci misure di attuazione, vista l'importanza di creare il clima di fiducia che consentirà lo sviluppo dell'economia digitale in tutto il mercato interno. È opportuno che le persone fisiche abbiano il controllo dei dati personali che li riguardano e che la certezza giuridica e operativa sia rafforzata tanto per le persone fisiche quanto per gli operatori economici e le autorità pubbliche (Gazzetta Ufficiale dell'Unione Europea, 2016, p.2). Questa necessità si concretizza con l'introduzione del nuovo regolamento per la protezione dei dati, la GDPR. Vediamo in che cosa consiste e quali impatti avrà sulle attività delle aziende.

5.1 GDPR (Reg. UE 2016/679)

Il 25 maggio 2018, nei 28 paesi dell'UE, è diventato pienamente operativo il Regolamento Generale sulla Protezione dei Dati (GDPR), che entrerà nel sistema giuridico di tutti gli Stati membri e avrà il compito di armonizzarne le normative, prima esistenti soltanto a livello nazionale. Viene così abrogata la Direttiva 95/46/CE (la cosiddetta Direttiva Madre), il principale strumento legislativo nel campo della privacy a livello comunitario dal 1995, che aveva due obiettivi principali: salvaguardare il diritto fondamentale alla protezione dei dati e garantire la libera circolazione dei dati personali tra gli Stati membri. La direttiva, tuttavia,

²⁶ Qualsiasi informazione riguardante una persona fisica identificata o identificabile («interessato»); si considera identificabile la persona fisica che può essere identificata, direttamente o indirettamente, con particolare riferimento a un identificativo come il nome, un numero di identificazione, dati relativi all'ubicazione, un identificativo online o a uno o più elementi caratteristici della sua identità fisica, fisiologica, genetica, psichica, economica, culturale o sociale (Gazzetta Ufficiale dell'Unione Europea, 2016, p.33)

risale a un'epoca in cui Internet era ancora in uno stadio iniziale di sviluppo e i *Big Data* non erano ancora stati generati. Nonostante i suoi obiettivi e principi rimangano tuttora validi, la Direttiva Madre non ha impedito la frammentazione dell'applicazione della protezione dei dati personali nel territorio dell'Unione, né ha eliminato l'incertezza giuridica o la percezione, largamente diffusa nel pubblico, che soprattutto le operazioni online comportino rischi per la protezione delle persone fisiche. Tali problemi, ostacolando la libera circolazione dei dati, possono costituire un freno all'esercizio delle attività economiche dell'Unione, falsare la concorrenza e impedire alle autorità nazionali di adempiere agli obblighi loro derivanti dal diritto dell'Unione²⁷.

Per questo motivo la Commissione Europea ha proposto una riforma radicale del quadro dell'UE in materia di protezione dei dati, approvata il 27 aprile 2016. Il GDPR consacra il diritto alla protezione dei dati personali come diritto fondamentale e costituzionale configurandolo come diritto all'autodeterminazione informativa. In generale questo nuovo regolamento:

- muta l'approccio regolatorio da "formale e re-attivo" in "sostanziale e pro-attivo": il trattamento e la protezione dei dati personali evolvono acquisendo una propria e autonoma rilevanza all'interno dei processi organizzativi e gestionali di un ente o di un'azienda;
- consolida le garanzie e i diritti azionabili dall'interessato per il controllo delle proprie informazioni e l'esercizio dell'autodeterminazione informativa;
- accresce le responsabilità del titolare e del responsabile con la finalità di indurli a comportamenti e prassi virtuose che riducano i rischi di operazioni non conformi o non consentite;
- centralizza la governance e il controllo sul rispetto e la conformità dei trattamenti alla normativa, tramite la cooperazione e la valorizzazione delle Autorità di Controllo nazionali verso il Comitato europeo per la protezione dei dati, incoraggiando meccanismi di certificazione, ampliando il sistema di vigilanza e rafforzando quello sanzionatorio (CODAU, 2017).

²⁷ Il diritto dell'Unione impone alle autorità nazionali degli Stati membri di cooperare e scambiarsi dati personali per essere in grado di svolgere le rispettive funzioni o eseguire compiti per conto di un'autorità di un altro Stato membro (Gazzetta Ufficiale dell'Unione Europea, 2016, p.2)

Secondo Ira S. Rubenstein²⁸ (2012) tali riforme non bastano, in particolare per tre motivi: gli attuali modelli di business hanno già ripetutamente dimostrato che la vecchia regolamentazione era inadeguata e, nonostante introduca dei cambiamenti nella giusta direzione, anche la GDPR si basa sostanzialmente su di essa; le imprese raccolgono sempre più dati personali anche a vantaggio dei consumatori, ma è evidente che mantengono ancora il controllo del mercato e perseguono i propri interessi; i *Big Data* consentono la re-identificazione dei soggetti tramite dati non personali e affida importanti decisioni sulle persone a intelligenze artificiali. I legislatori dovrebbero, quindi, incoraggiare l'adozione di nuovi modelli di business che trasferiscano il controllo dei dati dalle aziende ai consumatori, offrendo in cambio incentivi, come una maggiore flessibilità delle norme attraverso codici di condotta creativi o sanzioni ridotte.

5.2 Cosa cambia per le aziende

Per diverse organizzazioni sono ancora molti i passi da compiere per adempiere ai nuovi obblighi di legge. Spesso manca la consapevolezza dell'effettiva portata del regolamento, che coinvolge i dati generati o processati da tecnologie, strumenti e prassi sottovalutati o non compresi a fondo da molte imprese. A questo proposito in un articolo di *CorriereComunicazioni* del 20 marzo 2018, Matteo Colombo, Amministratore Delegato di Labor Project, realtà specializzata nell'assistenza e nella formazione sui temi della privacy e della conformità normativa, oltre che Presidente di AssoDpo, l'associazione nazionale che tutela e promuove la professione del *Data Protection Officer*, spiega che i rischi sono molteplici. In particolare riguardano il corretto trattamento dei dati relativi ai lavoratori, come, per esempio, quelli di geolocalizzazione dei veicoli aziendali: le scatole nere che vengono installate al loro interno danno alle organizzazioni la possibilità di controllarne in qualsiasi istante tutti i movimenti, ricostruirne i percorsi e estrapolarne dati sulla performance, col rischio di ledere la privacy dei dipendenti a causa di un accesso indebito a informazioni personali.

Le persone fisiche, inoltre, possono essere associate a identificativi online prodotti dai dispositivi, dalle applicazioni, dagli strumenti e dai protocolli utilizzati, quali gli indirizzi IP, a marcatori temporanei (cookies) o a identificativi di altro tipo, come i tag di identificazione a radiofrequenza (RFID). Tali identificativi possono lasciare tracce che, in particolare se combinate con altre informazioni ricevute dai server, possono essere utilizzate per creare profili

²⁸ Membro del Consiglio e Professore Aggiunto di Legge all'Information Law Institute della New York University School of Law

dettagliati delle persone.

Un altro problema sono gli strumenti elettronici, come i tag RFID appena citati e i *wearable*²⁹, che vengono applicati direttamente ai lavoratori. A titolo di esempio, Amazon, per velocizzare la ricerca dei prodotti stoccati nei magazzini da parte dei dipendenti, ha brevettato il braccialetto wireless: è in grado di monitorare con precisione tutti i movimenti delle mani, vibrando per guidarle nella giusta direzione, e, per questo motivo, di sorvegliare i comportamenti dei dipendenti. “È necessario che chi sceglie di adottarli definisca attentamente cosa intende monitorare, in quanto esistono algoritmi in grado di stabilire, a partire dai suoi movimenti, l’effettiva produttività del collaboratore, da cui può dipendere a sua volta il prolungamento del contratto” continua Colombo.

Anche gli strumenti di ultima generazione per la sicurezza, se non opportunamente inquadrati nelle prescrizioni del GDPR, possono essere lesivi della privacy. La *cybersecurity*, analizzando i flussi informativi per individuare attività anomale o illegali, monitora la navigazione degli utenti e può rappresentare un illecito. I sistemi di videosorveglianza oltre a sfruttare il riconoscimento facciale possono essere anche dotati di microfoni, che rischiano di costituire un illecito di intercettazione.

Il concetto fondamentale sarà quello dell’*accountability*: chi gestisce dati a livello pubblico o privato non potrà semplicemente aderire alle norme, ma dovrà anche garantire misure per affrontare rischi derivanti dal trattamento di informazioni sensibili e dimostrare un comportamento responsabile e proattivo sotto questo punto di vista. Le organizzazioni devono instaurare un rapporto di fiducia con clienti, partner di business e dipendenti e quindi comunicare loro in modo chiaro le politiche che intendono implementare, in modo che i vari stakeholder sappiano quali informazioni su di loro sono in circolazione e possono essere potenzialmente utilizzate. La *privacy by design* delle nuove tecnologie (ossia la dicitura “conforme al GDPR” del prodotto, che garantisce la crittografia e l’anonimizzazione dei dati) non sarà più sufficiente. Le organizzazioni dovranno rispettare il regolamento, declinandolo secondo le proprie specificità, e comprendere appieno le conseguenze dirette o indirette dell’utilizzo delle tecnologie. Dovranno anche definire e implementare una strategia che includa la valutazione di tutti i rischi a cui sono sottoposte, dalla minaccia che gli hacker penetrino nei mainframe alla possibilità che gli operatori utilizzino i dati a cui hanno accesso per scopi che vanno contro quelli dell’organizzazione.

²⁹ Dispositivi elettronici indossabili che possono svolgere svariate funzioni

5.3 Privacy e Big Data

Cambridge Analytica, società di consulenza britannica che usava l'analisi dei dati e la comunicazione strategica per supportare le campagne elettorali americane dal 2014, il 2 maggio 2018 ha dichiarato bancarotta e cominciato il procedimento d'insolvenza. Nel marzo di quest'anno l'azienda è stata coinvolta un grosso scandalo con Facebook che l'ha portata al fallimento.

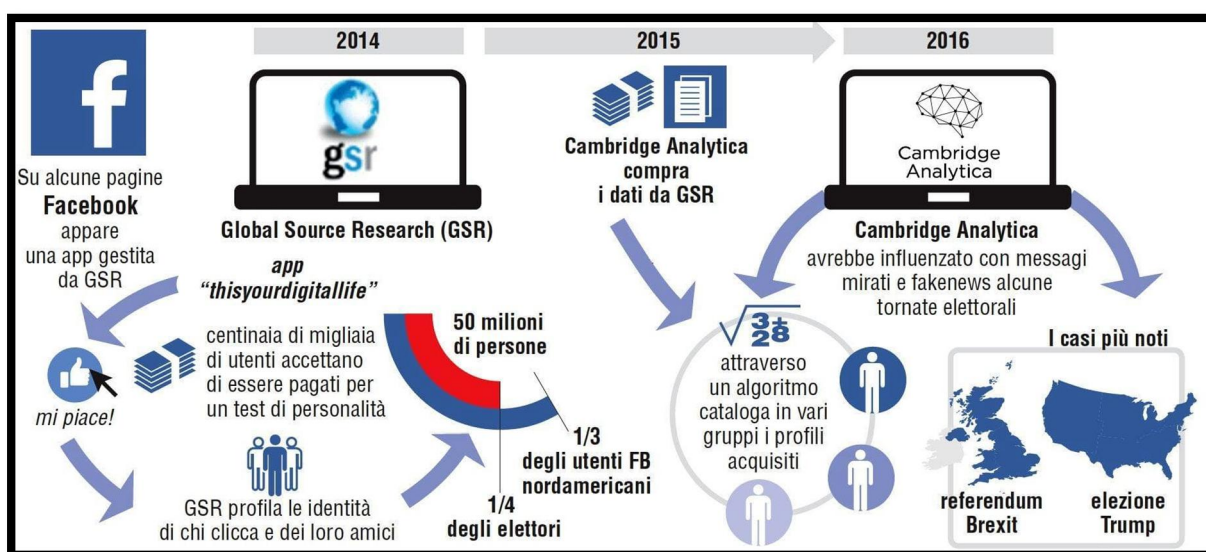


Fig.7 Il caso Facebook – Cambridge Analytica (Tgcom24.com, 2018)

L'app thisisyourdigitallife, creata dal ricercatore di Cambridge Alexander Kogan nel 2014, prometteva di riuscire a produrre profili psicologici e di previsione di comportamento dei propri utilizzatori, basandosi sulle attività svolte online. Accedendo con Facebook Login, le Graph API, le interfacce che terze parti possono usare per interagire con la piattaforma del social, consentivano di raccogliere enormi quantità di dati sugli utenti e, illegalmente, sui loro amici (secondo The Observer, il primo giornale a rivelare lo scandalo, pari a circa 50 milioni di americani e almeno un milione di inglesi), fra cui like, attività, foto, luoghi visitati, religione, interessi. Contro le privacy policy di Facebook, Kogan condivise tali informazioni con Cambridge Analytica, la quale le usò successivamente per creare profili psicologici e capire le inclinazioni politiche degli elettori americani, a cui inviare inserzioni personalizzate per influenzare gli esiti della campagna elettorale (The Guardian, 2018). L'episodio ha quindi puntato un faro su come Facebook, ma non solo, raccoglie, usa e condivide i dati per fare advertising personalizzato.

Fin dove le aziende possono spingersi nelle ricerche di mercato per approfondire la conoscenza dei bisogni dei propri clienti e offrire prodotti e servizi più soddisfacenti e coerenti con i loro interessi? Qual è il confine tra una raccolta dati legittima e una illegittima? La facile reperibilità delle informazioni potrebbe esporre i consumatori a degli abusi. È sufficiente il consenso, come rileva Daniel Solove, professore alla George University Law School, per permettere qualsiasi tipo di raccolta, uso e distribuzione dei dati. Inoltre, già sin dal momento della collezione dei loro dati, le implicazioni che ne conseguono risultano sconosciute alle persone, le quali non sono in grado di valutare i costi e i benefici dell'aver rivelato le informazioni né di sapere come e da chi verranno utilizzate. Aumenta così la diffidenza del consumatore a fornire dati su sé stesso, favorendo interventi sempre più intrusivi del legislatore che potrebbero limitare lo sfruttamento delle potenzialità dei *Big Data*.

Il caso Facebook - Cambridge Analytica è solo l'ultimo dei tanti che dimostrano il delicato trade-off tra gli enormi vantaggi dei *Big Data* e i rischi per la privacy. Il regolamento europeo ha il dovere di affrontare questo tema, alla ricerca di un difficile equilibrio. Sarà necessario raggiungere un punto di raccordo tra le divergenti esigenze di ricerca scientifica, salute pubblica, sicurezza nazionale, uso efficiente delle risorse, da un lato, e applicazione della legge e dei diritti di privacy, uguaglianza e libertà d'espressione degli individui, dall'altro. L'approccio puramente legislativo, inoltre, deve essere affiancato ad un approccio etico che vada oltre la sola difesa della riservatezza dell'individuo. Il Garante europeo della privacy Giovanni Buttarelli, in un articolo del Corriere della Sera (12 ottobre 2017), afferma: "Dobbiamo occuparci anche di cosa sia moralmente sostenibile e non soltanto tecnicamente fattibile, tenendo presente che lo sviluppo dei *Big Data* va messo al servizio dell'uomo e non viceversa. Nel momento in cui affidiamo al nostro smartphone tutte queste nostre informazioni personali, dobbiamo anche curarci della dignità e del rispetto dell'utente".

Bisognerà capire se, nei prossimi 15-20 anni di validità, il regolamento europeo sarà sufficiente per affrontare questi complessi temi.

Conclusioni

Questo elaborato ha evidenziato che, oggi, la quantità di dati prodotti e condivisi ogni momento è enorme e questo trend continuerà nel futuro, intensificandosi. Abbiamo capito che questo mare d'informazioni è talmente vasto da richiedere nuovi strumenti, processi e capacità per poterlo utilizzare efficacemente. Le ricerche di Forrester e dell'Osservatorio Business Intelligence & Big Data Analytics rilevano che assimilare queste novità non è semplice, poiché generano problemi di ricerca e inserimento delle figure, di modifica delle routine e dei processi operativi, di conciliazione con il modello culturale interno. Le organizzazioni che riconoscono le opportunità dei *Data Analytics* (e hanno una buona dose di abilità d'interpretazione), da dati che da soli sono poco significativi possono estrarre e ottenere importanti indicazioni e intuizioni, portando grande valore e vantaggi competitivi di lungo periodo al proprio business. È emerso che i settori in cui i *Big Data* hanno avuto più impatto sono il retail, dove le campagne di marketing si sono notevolmente potenziate (Macy's ha aumentato del 12% le sue vendite online grazie alla *Sentiment Analysis*); la finanza, dove l'identificazione delle attività illegali si è semplificata e velocizzata; le telecomunicazioni e i media, grazie alla possibilità di migliorare le relazioni col cliente con la *mass-customization*; i trasporti e le infrastrutture, per creare le cosiddette *smart city*; l'energia e le utenze, dove l'analisi dei dati porta a rilevanti risparmi di costo e maggiore sostenibilità ambientale.

Da un focus sulle aree funzionali abbiamo compreso che il marketing è quella che utilizza maggiormente soluzioni di *Analytics* e richiede l'inserimento di numerose figure specializzate come il *Chief Data Officer*. La personalizzazione di ciascuna relazione di vendita è l'applicazione più diffusa: Amazon e Netflix, con i loro sistemi di *cross* e *up-selling*, ne sono un chiaro esempio. Segue l'amministrazione, finanza e controllo che dall'analisi di *Big Data* ricava significativi vantaggi per la gestione e la trasparenza dei flussi informativi. La ricerca e sviluppo è invece ancora arretrata, anche se il *crowdsourcing* è la nuova frontiera dell'innovazione.

I benefici coinvolgono anche i clienti e le persone nella vita di tutti i giorni: innovazioni e rivoluzioni di prodotti, servizi, settori, soddisfano i bisogni comuni attuali, le esigenze più disparate, desideri ancora inespressi, in tempi brevi e costi bassi.

Qual è l'altra faccia della medaglia? Per le aziende e i consumatori, una questione delicata è quella della privacy. La condivisione sfrenata di dati personali, attività, interessi tramite dispositivi elettronici sempre più potenti mette tali soggetti in una situazione di conflitto: da un lato le aziende vogliono migliorare le proprie performance, ma hanno bisogno di tante

informazioni, anche private; dall'altro i consumatori desiderano prodotti e servizi personalizzati che li soddisfino, ma non gradiscono che le imprese conoscano e sorveglino tanti aspetti della loro vita. Il tema è tuttora aperto e dibattuto, ma con la rapida evoluzione tecnologica e i continui cambiamenti economici e sociali, la soluzione definitiva è ancora lontana.

Un altro problema su cui riflettere, cui accenno ora, è la conseguenza che la trasparenza delle informazioni ha sull'ottica di medio-lungo periodo delle organizzazioni. La ricca disponibilità di informazioni statistiche sull'andamento delle performance aziendali aumenta enormemente la pressione sui soggetti responsabili, come i vari manager, i quali vedono i probabili risultati delle proprie decisioni ancor prima che qualsiasi azione venga messa in atto. A livello governativo, anche i politici accorciano l'orizzonte delle proprie visioni, concentrandosi sull'accontentare le generazioni attuali allo scopo di guadagnarsi una facile rielezione. Questa miopia potrebbe compromettere o limitare le potenzialità dell'analisi dei dati, che invece mostrano i loro migliori vantaggi in un'ottica di medio-lungo periodo, con impatti negativi sul benessere socio-economico futuro.

La visione che abbiamo potuto rilevare da questo lavoro è complessivamente positiva: l'evoluzione tecnologica e i cambiamenti culturali, sociali ed economici messi in moto da questo diluvio di dati sono indiscutibili e ancora si promettono grandi innovazioni. A questo proposito, concludo con una notizia del momento: alcuni ricercatori di Google hanno elaborato un algoritmo per prevedere con una precisione del 95% la data della nostra morte, sulla base di dati di oltre 216 mila pazienti adulti ricoverati in ospedale per almeno 24 ore (Ansa, 2018). Quali applicazioni derivanti dallo sfruttamento di *Big Data* possono essere considerate buone innovazioni? Valutarne le conseguenze, da tutti i punti di vista, sarà una sfida complessa.ⁱ

Bibliografia

- Aliperto D., 2018. Gdpr, ecco quando il digitale può mettere a rischio la compliance. CorriereComunicazioni (online), 28 maggio. Disponibile su <<https://www.corrierecomunicazioni.it/privacy/gdpr/gdpr-ecco-quando-il-digitale-puo-mettere-a-rischio-la-compliance/>>. [Data di accesso: 30 giugno].
- Alteryx (s.d.). *Revolutionizing business through Data Science and Data Analytics*. Disponibile su <<https://www.alteryx.com/company/about-us>>. [Data di accesso: 30 maggio 2018].
- ANSA, 2018. Google scoprirà la data della morte con un algoritmo. Ansa.it (online) 20 giugno. Disponibile su <http://www.ansa.it/sito/notizie/tecnologia/software_app/2018/06/20/google-scoprirà-data-morte-con-algoritmo_4e06081e-8a96-4e28-860d-54d1bcd68e8d.html>. [Data di accesso: 5 luglio 2018].
- Apache Software Foundation, 2014. *What is Apache Hadoop?*. Disponibile su <<http://hadoop.apache.org/>>. [Data di accesso: 19 maggio 2018].
- Arora S., 2016. *Recommendation engines: how Amazon and Netflix are winning the personalization battle*. MarTech Advisor. Disponibile su <<https://www.martechadvisor.com/articles/customer-experience-2/recommendation-engines-how-amazon-and-netflix-are-winning-the-personalization-battle/>>. [Data di accesso: 26 giugno 2018].
- Azzalini A., Scarpa B., 2004. *Analisi dei dati e data mining*. Milano: Springer-Verlag.
- Bulygo Z., 2018. *How Netflix uses Analytics to select movies, create content and make multimillion dollar decisions*. Disponibile su <<https://neilpatel.com/blog/how-netflix-uses-analytics/>>. [data di accesso: 24 giugno 2018].
- Cadwalladr C., Graham-Harrison E., 2018. *Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach*. The Guardian (online), 17 marzo. Disponibile su <<https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>>. [Data di accesso 3 luglio 2018].
- Ceron A., Curini L., Iacus S. M., 2014. *Social Media e Sentiment Analysis – L’evoluzione dei fenomeni sociali attraverso la rete*. Milano: Springer.

Chen H., Chiang R. H. L., Storey V. C., 2012. *Business Intelligence and Analytics: from Big Data to big impact*. Management Information Systems Quarterly (online) dicembre, vol.36. Disponibile su <<https://ai.arizona.edu/sites/ai/files/MIS611D/chen-bi-december-2012.pdf>>. [Data di accesso: 8 giugno 2018].

Chesbrough H. W., 2011. *Top 10 lessons on the new business of innovation*. MIT Sloan Management Review, pp.35-41. Disponibile su <<https://sloanreview.mit.edu/files/2011/06/INS0111-Top-Ten-Innovation.pdf>>. [Data di accesso: 20 giugno 2018].

Convegno dei Direttori generali delle Amministrazioni Universitarie, 2017. Linee guida in materia di privacy e protezione dei dati personali in ambito universitario. Disponibile su <http://www.codau.it/sites/default/files/verbali/all_3_lineeguida_privacy_gdpr_ravera.pdf>. [Data di accesso: 1 luglio 2018].

Dezyre, 2015. *5 Big Data use cases - How companies use Big Data*. Disponibile su <<https://www.dezyre.com/article/5-big-data-use-cases-how-companies-use-big-data/155>>. [data di accesso: 26 giugno 2018].

Dimension Data, 2016. *Automation in manufacturing*. Disponibile su <<https://www2.dimensiondata.com/-/media/dd/corporate/content-images/pdfs/uncategorised/latest-thinking/automation-in-manufacturing.pdf>>. [Data di accesso: 4 giugno 2018].

FICO (s.d.). *FICO at a glance*. Disponibile su <http://www.fico.com/en/about-us - at _glance>. [Data di accesso: 30 maggio 2018].

Fiorletta A., 2014. *Macy's.com sees 12% boost in online sales with Predictive Analytics*. RetailTouchPoints. Disponibile su <<https://www.retailtouchpoints.com/features/news-briefs/macy-s-com-sees-12-boost-in-online-sales-with-predictive-analytics>>. [Data di accesso: 26 giugno 2018].

Foppa S. U. et al., 2017. Come fare Big Data Analysis e ottenere valore per le aziende. ZeroUno (online) 19 dicembre. Disponibile su <<https://www.zerounoweb.it/analytics/big-data/come-fare-big-data-analysis-e-ottenere-valore-per-le-aziende/>>. [Data di accesso: 24 maggio 2018].

Forrester, 2015. *Forrester's Global Business Technographics Data and Analytics Survey*.

Disponibile su

<<https://www.forrester.com/Global+Business+Technographics+Data+And+Analytics+Survey+2016/-/E-sus3014>>. [Data di accesso: 28 maggio 2018].

Forzinetti E., 2017. La privacy nell'era dei Big Data, il Garante UE della Privacy: «Ecco come cambierà la legislazione europea». Corriere della Sera Tecnologia (online), 12 ottobre.

Disponibile su <https://www.corriere.it/tecnologia/cyber-cultura/17_ottobre_05/privacy-nell-era-big-data-garante-ue-privacy-ecco-come-cambiera-legislazione-europea-a6f531a8-a9b0-11e7-b86b-3752ab14654a.shtml>. [Data di accesso: 30 giugno 2018].

Gartner IT Glossary (s.d.). *Big Data*. Disponibile su <<https://www.gartner.com/it-glossary/big-data>>. [Data di accesso: 10 maggio 2018].

Gartner, (s.d.). *Gartner Magic Quadrant*. Disponibile su

<https://www.gartner.com/technology/research/methodologies/research_mq.jsp>. [Data di accesso: 22 maggio 2018].

Gartner, 2016. *Magic Quadrant for Advanced Analytics Platforms*. Disponibile su

<<https://www.gartner.com/doc/3204117/magic-quadrant-advanced-analytics-platforms>>. [Data di accesso: 22 maggio 2018].

Gazzetta Ufficiale dell'Unione Europea N. L119, 04/05/2016.

Gulla J., 2017. *Trend watch: Business Intelligence and Business Analytics*. IBMSystems

magazine (online) 10 aprile. Disponibile su <<http://ibmsystemsmag.com/blogs/it-trendz/april-2017/trend-watch-business-intelligence-and-business-ana/>>. [Data di accesso: 11 maggio 2018].

Henke N. et al., 2016. *The age of Analytics: competing in a data-driven world*. McKinsey Global Institute. Disponibile su

<https://www.mckinsey.com/~/_media/McKinsey/Business%20Functions/McKinsey%20Analytics/Our%20Insights/The%20age%20of%20analytics%20Competing%20in%20a%20data%20driven%20world/MGI-The-Age-of-Analytics-Full-report.ashx>. [Data di accesso: 3 giugno 2018].

Hota J., Ghosh D., 2013. *Workforce Analytics approach: an emerging trend of Workforce Management*. AIMS International Journal of Management (online), vol.7. Disponibile su

<https://www.researchgate.net/profile/Debjani_Ghosh3/publication/261173868_Workforce_Analytics_ApproachAn_emerging_Trend_of_Workforce_Management/links/0c9605336ad2e13e9e000000/Workforce-Analytics-ApproachAn-emerging-Trend-of-Workforce-Management.pdf>. [Data di accesso: 19 giugno 2018].

International Data Corporation, 2017. *Worldwide Semiannual Big Data and Analytics Spending Guide*. Disponibile su <<https://www.idc.com/getdoc.jsp?containerId=prUS42371417>>. [Data di accesso: 3 giugno].

James D. et al., 2017. *Driving digital destiny: digital reinvention in automotive*. IBM Institute for Business Value. Disponibile su <https://public.dhe.ibm.com/common/ssi/ecm/gb/en/gbe03876usen/gbe03876usen-02_GBE03876USEN.pdf>. [Data di accesso: 13 giugno 2018].

Jeske M., Grüner M., Weiß F., 2013. *Big Data in logistics*. DHL Customer Solutions & Innovation. Disponibile su <<https://www.ssti.us/wp/wp-content/uploads/2014/03/Big-Data-In-Logistics.pdf>>. [Data di accesso: 19 giugno 2018].

Kotler P. et al., 2015. *Principi di Marketing*. 15^a edizione. Milano: Pearson Italia.

Levy E., 2018. *The five essentials components of a data strategy*. SAS. Disponibile su <https://www.sas.com/content/dam/SAS/en_us/doc/whitepaper1/5-essential-components-of-data-strategy-108109.pdf>. [Data di accesso: 3 giugno 2018].

Marr B., 2015. *Big Data case study collection*. Disponibile su <https://www.bernardmarr.com/img/bigdata-case-studybook_final.pdf>. [Data di accesso. 18 giugno 2018].

Marr B., 2015. *Big Data: 20 mind boggling facts everyone must read*. Forbes (online) 30 settembre. Disponibile su <<https://www.forbes.com/sites/bernardmarr/2015/09/30/big-data-20-mind-boggling-facts-everyone-must-read/#136378ad17b1>>. [Data di accesso: 11 maggio 2018].

Marr B., 2018. *How is Big Data used in practice? 10 use cases everyone must read*. Bernard Marr & Co. Disponibile su <<https://www.bernardmarr.com/default.asp?contentID=1076>>. [Data di accesso: 14 giugno 2018].

- Manyika J. et al., 2011. *Big Data: The next frontier for innovation, competition and productivity*. McKinsey Global Institute. Disponibile su <[https://www.mckinsey.com/~media/McKinsey/Business Functions/McKinsey Digital/Our Insights/Big data The next frontier for innovation/MGI_big_data_exec_summary.ashx](https://www.mckinsey.com/~media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Big%20data%20The%20next%20frontier%20for%20innovation/MGI_big_data_exec_summary.ashx)>. [Data di accesso: 12 maggio 2018].
- McAfee A., Brynjolfsson E., 2012. *Big Data: the management revolution*. Disponibile su <<http://tarjomefa.com/wp-content/uploads/2017/04/6539-English-TarjomeFa-1.pdf>>. [Data di accesso: 20 maggio 2018].
- Ministero dell'Istruzione dell'Università e della Ricerca, 2016. *Big Data: rapporto del gruppo di lavoro del Miur*. Disponibile su <<http://www.istruzione.it/allegati/2016/bigdata.pdf>>. [Data di accesso: 18 maggio 2018].
- Osservatorio Big Data Analytics & Business Intelligence, 2015. *Big Data: da data insight a data driven strategy*. Disponibile su <https://www.osservatori.net/it_it/osservatori/comunicati-stampa/big-data-da-data-insight-a-data-driven-strategy>. [Data di accesso: 27 maggio 2018].
- Osservatorio Big Data Analytics & Business Intelligence, 2017. *Big Data is now: tomorrow is too late*. Disponibile su <https://www.osservatori.net/it_it/osservatori/comunicati-stampa/big-data-is-now-tomorrow-is-too-late>. [Data di accesso: 27 maggio 2018].
- Piva A., 2018. Come impostare un progetto di Big Data Analytics?. Osservatori.net (online) 11 gennaio. Disponibile su <https://blog.osservatori.net/it_it/progetti-big-data-analytics>. [Data di accesso: 27 maggio 2018].
- Porsche (s.d.). *Porsche at a glance*. Disponibile su <<https://www.porsche.com/italy/aboutporsche/overview/>>. [Data di accesso: 22 giugno 2018].
- Porsche, 2018. *An eye for mistakes*. Disponibile su <<https://newsroom.porsche.com/en/porsche-digital/porsche-production-4-0-artificial-intelligence-ai-neural-networks-indtact-predictive-maintenance-tech-laboratory-digital-lab-berlin-14904.html>>. [Data di accesso 22 giugno 2018].
- Rubinstein I. S., 2012. *Big Data: the end of privacy or a new beginning?*. New York University Public Law and Legal Theory Working Papers.
- Rusitschka S., Curry E., 2016. *New horizons for a data-driven economy*. Springer International Publishing, pp.225-244.

SAP (s.d.). *Who we are*. Disponibile su <<https://www.sap.com/italy/about/careers/who-we-are.html>>. [Data di accesso: 30 maggio 2018].

SAS (s.d.). Big Data: cosa sono e perché sono importanti. Disponibile su <https://www.sas.com/it_it/insights/big-data/what-is-big-data.html>. [Data di accesso: 12 maggio 2018].

SAS (s.d.). *About SAS*. Disponibile su <https://www.sas.com/en_sa/company-information.html>. [Data di accesso: 30 maggio 2018].

Schroeck M. et al., 2012. *Analytics: the real-world use of Big Data*. IBM Institute for Business Value. Disponibile su <<https://public.dhe.ibm.com/common/ssi/ecm/gb/en/gbe03519usen/global-business-services-global-business-services-gb-executive-brief-gbe03519usen-20180209.pdf>>. [Data di accesso: 6 giugno 2018].

Solove D., 2013. *Privacy Self-Management and the Consent Dilemma*. Harvard Law Review.

Vanderbilt T., 2013. *The science behind the Netflix algorithms that decide what you'll watch next*. Wired (online) 7 agosto. Disponibile su <<https://www.wired.com/2013/08/qq-netflix-algorithm/>>. [Data di accesso 24 giugno 2018].

Wyllie D., 2013. *How Big Data is helping law enforcement*. PoliceOne (online) 20 agosto. Disponibile su <<https://www.policeone.com/police-products/software/Data-Information-Sharing-Software/articles/6396543-How-Big-Data-is-helping-law-enforcement/>>. [Data di accesso: 10 giugno 2018].

Ziegler M., Rossman S., 2017. *Digital machinery decoded*. Porsche Consulting. Disponibile su <https://www.porsche-consulting.com/fileadmin/docs/Startseite/News/tt1309/Porsche_Consulting_Digital_Machinery_Decoded.pdf>. [Data di accesso: 22 giugno 2018].

Zillner S. et al., 2016. *New horizons for a data-driven economy*. Springer International Publishing, pp.169-178.

BIBLIOGRAFIA DELLE FIGURE

Figura 1: Le 5 V dei *Big Data*. Disponibile su <https://www.dreamstime.com/stock-photography-big-data-v-words-image35236832>.

Figura 2: La varietà di fonti che generano *Big Data*. Disponibile su <https://pixabay.com/en/communication-internet-1927697/>.

Figura 3: *Magic Quadrant per le Advanced Analytics*. Gartner, 2016. *Magic Quadrant for Advanced Analytics Platforms*.

Figura 4: Le cinque componenti fondamentali della *data strategy*. SAS, 2018. *The five essentials components of a data strategy*.

Figura 5: *Close innovation*. Chesbrough H. W., 2011. *Top 10 lessons on the new business of innovation*.

Figura 6: *Open innovation*. Chesbrough H. W., 2011. *Top 10 lessons on the new business of innovation*.

Figura 7: Il caso Facebook – Cambridge Analytica. Disponibile su http://www.tgcom24.mediaset.it/mondo/cambridge-analytica-la-talpa-dietro-c-era-bannon-class-action-negli-usa_3129619-201802a.shtml.

Tabella 1: Applicazioni economiche di *Data Analytics* in alcuni settori industriali. Immagine elaborata dall'autore.

ⁱ Numero parole: 14834