

Dipartimento di Impresa e Management

Cattedra di Informatica

**“Big Data Analytics e gestione aziendale: il nuovo oro?
Focus sui cambiamenti apportati nel settore dei
trasporti aerei”**

RELATORE

Chiar.mo Prof.
Laura Luigi

CANDIDATO

Colone Benedetta
Matricola 220081

INDICE

| | |
|--|-----------|
| Introduzione | 5 |
| Capitolo 1: Big Data: perché sono così importanti? | 7 |
| 1.1 Una definizione non univoca | 8 |
| 1.1.1 L'evoluzione storica del concetto | 10 |
| 1.1.2 Le 5 V | 13 |
| 1.2 Qual è il loro valore? | 15 |
| 1.3 Tipologie di dati | 15 |
| 1.4 Modalità di analisi dei Big Data | 17 |
| 1.4.1 Strumenti per l'analisi | 18 |
| 1.4.2 Tipologie di <i>Analytics</i> | 21 |
| Capitolo 2: “Big Data, Big Dupe”: alcune critiche | 22 |
| 2.1 Difficoltà nella definizione | 23 |
| 2.2 Trattamento dei dati personali | 25 |
| 2.2.1 Cambridge Analytica..... | 27 |
| 2.2.2 General Data Protection Regulation | 31 |
| 2.3 Illusione o realtà? | 32 |
| Capitolo 3: Ambiti di applicazione nella gestione aziendale | 34 |
| 3.1 Management | 35 |
| 3.1.1 Processo di <i>decision-making</i> | 36 |
| 3.1.2 Supply Chain Management | 37 |
| 3.2 Marketing | 41 |
| 3.2.1 Customer Relationship Management..... | 44 |

| | |
|--|-----------|
| Capitolo 4: La trasformazione del settore dei trasporti aerei con l'analisi dei big data: alcuni case studies | 48 |
| 4.1 Vueling Airlines | 50 |
| 4.2 Delta Airlines nel monitoraggio dei bagagli | 50 |
| 4.3 EasyJet e nuove strategie di prezzo | 51 |
| 4.4 SouthWest Airlines, ottimizzazione dei consumi di carburante e sicurezza sugli aerei | 52 |
| 4.5 La funzione “Know me” di British Airways..... | 53 |
| 4.6 Raccomandazioni future sul settore trasporti aerei | 54 |
| | |
| Capitolo 5: Quali sviluppi per i Big Data in futuro? | 56 |
| 5.1 Big Data e l’Internet of Things | 58 |
| 5.2 Artificial Intelligence per valutare la qualità dei dati..... | 58 |
| 5.3 Data as a service (Daas)..... | 59 |
| 5.4 Come cambiano i modelli di BDA con l’avvento del Covid-19? | 61 |
| | |
| Conclusioni..... | 63 |
| | |
| Bibliografia | 65 |
| | |
| Indice delle figure..... | 68 |

INTRODUZIONE

Viviamo in un'epoca di sovraccarico informativo: noi umani creiamo continuamente molti più dati di quanti siamo in grado di processarne usando la nostra mente. In un minuto in media su Google si effettuano 3 milioni e mezzo di ricerche, vengono caricate più di 400 ore di nuovi video sulla piattaforma di YouTube. Passando a Facebook, negli stessi 60 secondi 300.000 persone aggiornano la loro bacheca e viene schiacciato circa 3 milioni di volte il tasto “like” per tutti i contenuti condivisi con gli amici.

Ma che valore possono generare questi dati apparentemente inutili? In questo elaborato si scopriranno più informazioni su tale aspetto e si capirà che non sono solo i big data che contano.

Big Data Analytics (BDA) è un termine frutto del connubio tra i *big data* e l'*analytics*. Rappresenta infatti la metodologia che utilizza l'analisi dei dati per trarne indicazioni vantaggiose in svariati ambiti applicativi oltre a quello economico.

L'obiettivo del seguente lavoro è capire quanto siano rilevanti ad oggi i nuovi modelli analitici e “misurare” la portata dell'innovazione. Si dice spesso che i big data sono il nuovo oro o anche il nuovo petrolio, ma l'accostamento in questo caso non è stato fatto con i “grandi dati”, bensì con tutto il processo della Big Data Analytics.

Nella prima parte dell'elaborato ci si concentrerà inizialmente sulle ragioni per cui le aziende hanno iniziato ad implementare la BDA. Seguirà poi una sezione in cui si argomenterà sulle caratteristiche generali dei big data e sull'evoluzione storica che questo termine ha avuto dai primi anni Quaranta del Novecento fino a diventare così usato. L'attenzione si sposterà quindi su un'analisi che approfondisce il concetto del valore da attribuirvi alla luce dei recenti progressi. Le ultime sezioni del primo capitolo verteranno sulla Big Data Analytics: saranno discusse le varie forme sotto cui le aziende possono trovare i big data e gli strumenti analitici richiesti per gestire ed elaborare le varie tipologie, traendone vantaggio dall'inserimento in modelli come Machine Learning e Data Mining.

Il secondo capitolo offrirà invece un punto di vista opposto rispetto a quello favorevole sui big data. Verrà quindi mostrata l'altra faccia della medaglia tramite la presentazione delle criticità osservabili quando si parla di determinate informazioni. Per prima si affronterà la debolezza legata all'inesistenza di un'unica definizione riconosciuta dagli studiosi e la conseguente ambiguità che ciò comporta nell'uso del termine big data. Il tema della privacy, di gran lunga quello che più attira l'attenzione di media, studiosi e curiosi, è un'altra sfida che il fenomeno presenta. Lo scandalo Cambridge Analytica ricoprirà la funzione di caso applicativo e subito dopo, anche seguendo una successione temporale, verranno spiegate le caratteristiche più importanti del Regolamento europeo emanato nel 2016 per la protezione dei dati personali. In ultimo, verranno affrontate due aspre critiche al sistema dei big data, secondo le quali si tratterebbe solo di un'illusione.

Il terzo capitolo spiega le modalità degli ambiti aziendali in cui questa strategia è stata implementata, concentrandosi però esclusivamente sulla parte gestionale. Il management e i relativi processi di *decision-making* sono i primi ad essere analizzati. Segue poi l'area del Supply Chain Management, in cui è presente un esempio applicativo di come la Big Data Analytics è stata capace di impattare su alcuni processi della catena di approvvigionamento. In secondo luogo si osserveranno le modalità d'impiego dell'analisi dei big data nella funzione marketing; il Customer

Relationship Management cercherà di approfondire come è cambiato il rapporto tra i clienti e le aziende con l'avvento dei nuovi strumenti digitali, spiegandolo meglio attraverso un secondo minicase.

Il quarto capitolo sarà di fondamentale importanza perché analizzerà il modo in cui la Big Data Analytics ha impattato su un intero settore economico: i trasporti aerei. Dopo una breve introduzione sulle motivazioni che hanno spinto le aziende ad effettuare queste scelte strategico-informatiche, si presenteranno una serie di applicazioni di successo. Ogni compagnia aerea ha sviluppato determinate competenze analitiche e le applica in modo tale da sfruttare l'Advanced Analytics per risolvere i problemi più rilevanti nell'azienda. Si procederà infine a dare alcune indicazioni e raccomandazioni sugli sviluppi futuri nel settore dei trasporti aerei una volta a conoscenza delle tecniche digitali utilizzate dalle compagnie e si possono trarre, seppur minime, delle conclusioni.

Ultimo ma non meno importante, il quinto capitolo esplora le tendenze future della Big Data Analytics, essenziali per capire l'andamento generale del fenomeno una volta delineate le maggiori caratteristiche e avuta una panoramica generale degli ultimi sviluppi. In particolare, per primi si studieranno gli andamenti futuri con i relativi miglioramenti da apportare alla tecnologia dell'Internet of Things. Successivamente verrà affrontata la tendenza di elementi come l'intelligenza artificiale di impattare sulla qualità dei dati archiviati dalle aziende. Un'ulteriore innovazione che in futuro acquisterà enorme importanza è il DaaS (Data as a Service), in cui si approfondiranno le possibili relazioni future tra Cloud e big data. come chiusura di questo capitolo si offre una piccola spiegazione del ruolo che ha giocato il Covid-19 sulla struttura dei modelli analitici della BDA e su come evitare che uno shock esterno alteri i processi analitici e di conseguenza le indicazioni ricavabili da questi.

Capitolo 1

Big data: perché sono così importanti?

Il termine big data descrive l'ampio volume di dati, sia strutturati che non strutturati, che attraversa un gran numero di organizzazioni quotidianamente. L'aspetto quantitativo dei dati non è più il punto chiave, ma ciò che le organizzazioni fanno con le informazioni che contano. I big data possono essere analizzati per osservazioni che portano a intraprendere decisioni migliori e mosse strategiche molto efficaci.

In numerosi settori, le imprese che competono tra di loro e perfino i nuovi entranti usano le strategie risultanti dai dati analizzati per competere, innovare, acquisire e creare valore. È possibile, per esempio, profilare i clienti di una compagnia non più basandosi su informazioni di natura meramente statistica, ma attraverso la raccolta e l'analisi di dati rinvenibili da commenti postati sui social, da recensioni condivise, dal comportamento dell'utente sui vari motori di ricerca e/o siti web o ancora dall'utilizzo di software di geolocalizzazione provenienti da smartphone e da altri dispositivi mobili che consentono di effettuare azioni pubblicitarie anche in tempo reale.

Ogni compagnia sfrutta i big data a disposizione nel modo che ritiene più opportuno; migliore sarà l'utilizzo che ne fanno, maggiore sarà il loro potenziale di crescita. Le aziende possono acquistare o ricavare i dati da innumerevoli risorse e analizzarli al fine di avere diversi vantaggi (Baumann, P., & Riedel, M., 2019):

- **Riduzioni di tempo:** l'elevata velocità di strumenti per la Big Data Analytics identifica facilmente nuove fonti di dati che ne permettono lo studio immediato e aumentano la capacità di prendere rapide decisioni basate sui risultati.
- **Miglioramento dell'offerta di marketing:** l'analisi dei dati aiuta ad implementare ad ogni livello le performance dell'impresa, specialmente nell'area marketing. Ciò implica incontrare le aspettative dei clienti, cambiare le linee di prodotto dell'azienda se necessario e assicurarsi che le campagne di marketing siano efficaci.
- **Riduzione dei costi:** alcuni meccanismi usati per analizzare i big data comportano vantaggi di costo alle imprese quando si tratta di archiviazione di dati e possono condurre all'identificazione di processi più efficienti per contenere i costi aziendali.
- **Controllo della reputazione:** esaminando le giuste informazioni si può misurare il *sentiment* online dei clienti e degli stakeholders in generale. Gli analisti saranno quindi a conoscenza di chi dice cosa sulla loro azienda e potranno monitorare in ogni momento i vari cambiamenti, fornendo feedback utili per migliorare la reputazione.
- **Driver per l'innovazione e lo sviluppo dei prodotti:** un ulteriore vantaggio deriva dalla capacità dei dati di contribuire all'innovazione e allo sviluppo di nuovi prodotti o di prodotti già esistenti per soddisfare le mutevoli esigenze dei clienti.
- **Incremento dell'acquisizione e della fidelizzazione dei clienti:** essendo i clienti tra le risorse più importanti da cui dipendono le aziende, uno degli obiettivi primari è costruire una valida e solida base dalla quale partire per ampliarsi e crescere. Allo stesso tempo bisogna riconoscere l'elevato tasso di concorrenza a cui tutte le aziende sono sottoposte; fondamentale è anche saper fidelizzare i propri clienti e offrire loro ciò che stanno. Lo studio dei big data consente alle imprese di osservare vari modelli e tendenze relativi ai clienti. Esaminare e saper interpretare i loro comportamenti è importante per innescare la lealtà.
- **Maggior comprensione delle condizioni di mercato:** Comprendendo le informazioni forniteci dalle analisi dei dati è possibile spiegare determinate condizioni del mercato in un dato periodo. Ad esempio, valutando i comportamenti di acquisto dei clienti, le aziende di

un certo settore possono scoprire i beni più venduti e produrre secondo questa tendenza. Questo comporterebbe un vantaggio competitivo rispetto ai concorrenti di notevole importanza.

1.1 Una definizione non univoca

I dati vengono creati con ogni clic, scorrimento, condivisione, ricerca e streaming, proliferando la domanda per il mercato dei big data e della loro analisi a livello globale. Ma come possiamo spiegare tutte queste proprietà e diversità in una sola frase? È un argomento sulla bocca di moltissime persone di cui però non si ha una precisa percezione. Uno studio (Favaretto, M. et al., 2020) dimostra come nemmeno tra esperti ci sia un consenso generale sulla definizione e l'interpretazione del concetto di big data. Vi è stata solo una diffusa opinione tra i ricercatori nel concordare su come tali informazioni fossero enormi quantità di dati digitali generati da dispositivi tecnologici, che richiedono specifici processi computazionali e algoritmici, per trovare risposte logiche alle varie ipotesi di ricerca. Famosa ed indicativa è l'espressione utilizzata nel 2013 da Dan Ariely, professore presso la Duke University: *“L'argomento big data è come l'argomento sesso per gli adolescenti: tutti ne parlano, nessuno sa davvero come farlo ma tutti pensano che chiunque altro lo stia facendo e, di conseguenza, tutti dichiarano di farlo”*.

Ad oggi non esiste una definizione riconosciuta univocamente dai ricercatori e dagli esperti. Esistono numerose descrizioni che sottolineano diversi aspetti caratterizzanti i big data. Una primordiale spiegazione di Douglas Laney nel 2003 presenta, come vedremo nei successivi paragrafi, una svolta sotto alcuni aspetti. Una definizione molto quotata ma non pienamente esplicativa venne diffusa tramite lo studio effettuato dalla società di consulenza McKinsey&Company: “set di dati le cui dimensioni vanno oltre la capacità dei tipici strumenti software di database di acquisire, archiviare, gestire e analizzare” (McKinsey&Co, 2011). Un ulteriore contributo viene dato dall'articolo accademico pubblicato su Gartner¹ che riprende il vecchio lavoro di Douglas Laney aggiungendo alcune novità; “i big data sono risorse informative ad alto volume, ad alta velocità e/o ad alta varietà che richiedono forme innovative e convenienti di elaborazione delle informazioni che consentono approfondimenti, processi decisionali e automazione dei processi” (Beyer, M., Laney, D., 2012). Nell'anno successivo Mayer-Schönberger e Cukier nel loro libro² offrono un loro punto di vista su ciò che si può fare con i big data, connotandoli secondo la “capacità della società di sfruttare le informazioni in modi nuovi per produrre intuizioni utili o beni e servizi di valore significativo” e “... cose che si possono fare su larga scala che non si possono fare su scala minore, per estrarre nuove intuizioni o creare nuove forme di valore” (Mayer-Schönberger, V., Cukier, K., 2013). Citando qualificazioni più recenti ma allo stesso modo non pienamente rappresentative, IBM³ caratterizza così questo fenomeno: “Big data è un termine applicato a set di dati la cui dimensione o tipo va oltre la capacità dei tradizionali database relazionali di acquisire, gestire ed elaborare i dati con bassa latenza. I big data hanno una o più delle seguenti caratteristiche: alto volume, alta velocità o alta varietà.” In ultimo, offrendo una definizione di SAS (Statistical Analysis System)⁴, si riferisce “a dati così grandi, veloci o complessi che è difficile o impossibile elaborarli utilizzando metodi tradizionali.”

¹ Beyer, M., & Laney, D. (21 giugno 2012). The Importance of 'Big Data': A Definition. Gartner Research. doi: <https://www.gartner.com/en/documents/2057415/the-importance-of-big-data-a-definition>

² Mayer-Schönberger V., & Cukier, K., (2013). Big Data: a Revolution That Will Transform How We Live, Work and Think. London: John Murray Publishers.

³ Big Data Analytics. Disponibile su <https://www.ibm.com/analytics/hadoop/big-data-analytics>

⁴ Big Data: What it is and why it matters. Disponibile su https://www.sas.com/en_us/insights/big-data/what-is-big-data.html

Come è possibile notare tutte le definizioni si concentrano su aspetti diversi ma ugualmente importanti per la qualificazione del termine e del fenomeno big data. Procediamo quindi a delineare alcune componenti essenziali del concetto (De Mauro, 2019), alcune delle quali introdotte da Andrea de Mauro, senza cui ci si troverebbe di fronte soltanto ad un semplice dataset.

- L'**informazione** è la prima parte poiché tutti i dati vengono prodotti e gestiti dalle aziende o dai ricercatori al fine di estrarne un contenuto chiarificatore.
- L'**impatto** dei servizi e prodotti deve essere rilevante, tale da influire sulle abitudini di vita, sul modo di lavorare, sui processi produttivi.
- La **tecnologia** è fondamentale vista la specificità di strumenti richiesta al fine di ottenere una corretta elaborazione dei dati disponibili.
- I **metodi** abbandonano la tradizione per concentrarsi sull'innovazione una volta intrapresa la fase di analisi delle informazioni.

Il modello ITMI (De Mauro, 2019) è composto dai quattro elementi presentati precedentemente e fornisce una serie di parole chiave che si ricollegano direttamente con i quattro elementi presentati precedentemente e di cui si parlerà nel corso di questa tesi.

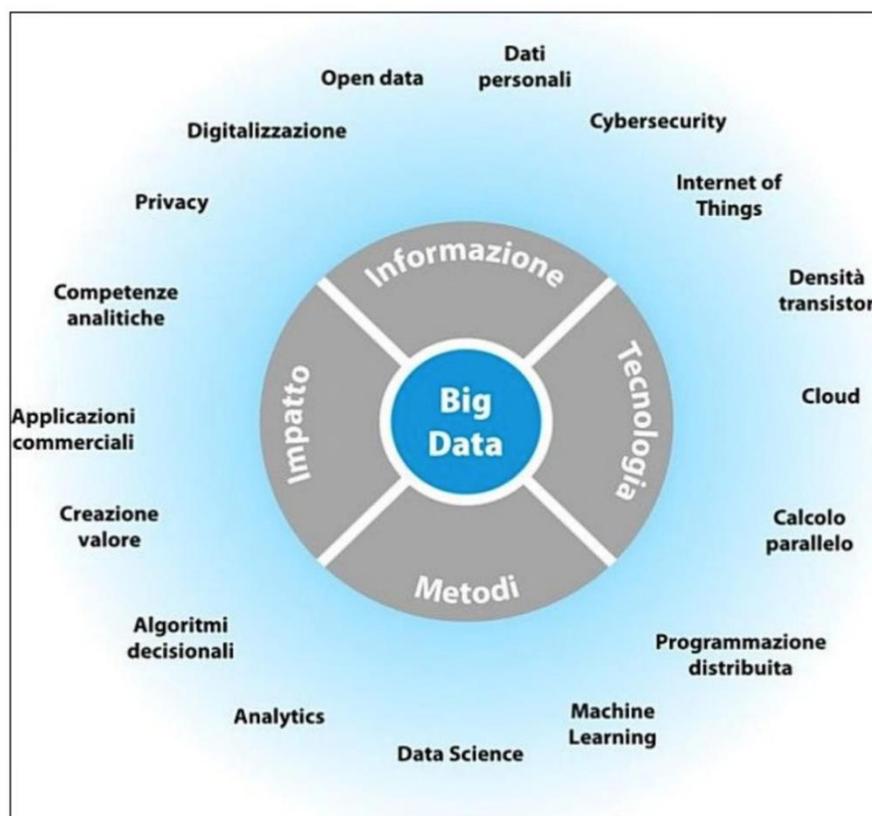


Figura 1: Modello ITMI (Informazioni, Tecnologie, Metodi, Impatto)

La **complessità** è infine un altro ingrediente fondamentale nella qualificazione di dataset come big data. Questa peculiarità deriva dal fatto che è molto difficile nel trasformare, scremare e collegare informazioni trasversali. Definire le gerarchie e stabilire efficaci connessioni tra i vari dati può

essere soltanto fatto mediante l'applicazione del modello ITMI: algoritmi complessi, tecnologie all'avanguardia sono l'arma vincente per affrontare questo aspetto.

1.1.1 L'evoluzione storica del concetto

La storia dei big data è iniziata molto prima che questo nome avesse per la prima volta spazio e risonanza negli articoli accademici scientifici (Press, G., 2013). In questo paragrafo pertanto si analizzeranno le varie tappe che hanno portato alla nascita e allo sviluppo di un fenomeno che ai giorni d'oggi rappresenta, almeno secondo molti informatici, economisti e scienziati, il futuro.

Già a partire dalla metà del ventesimo secolo si ebbe un primo riferimento al rilevante tasso di crescita nel volume dei dati, concetto racchiuso nell'idea di "esplosione delle informazioni" che sarebbe potuta avvenire negli anni successivi. Il volume quindi iniziava ad avere fin da quegli anni una certa importanza ed è un elemento che poi si riterrà essenziale nella definizione moderna dei big data. La pietra miliare fu posta da Fremont Ryder nel 1944, questi nel libro "*The Scholar and the Future of the Research Library*"⁵ stimò un raddoppio nella crescita delle dimensioni delle librerie americane ogni sedici anni. A questo tasso prevedeva una mole non indifferente di libri e articoli pubblicati che avrebbe avuto un grande impatto sul personale e sulle modalità di catalogazione dei documenti.

Seguendo l'idea nel 1961 venne pubblicato da Derek Price il libro "*Science Since Babylon*"⁶ in cui ci si ricollegava alla teoria di Ryder, che venne rafforzata attraverso la legge dell'incremento esponenziale, secondo la quale da ogni scoperta scientifica, e quindi da ogni *paper* accademico, si sarebbe generata una serie di innovazioni ad un tasso tendenzialmente costante che, a cascata, avrebbero poi portato ad ulteriori pubblicazioni. Da queste analisi si ebbe dunque la necessità di mantenere al minimo i requisiti di archiviazione delle informazioni: nel novembre 1967 si implementa questo processo grazie all'articolo "*Automatic data compression*"⁷ pubblicato da B. A. Marron e P. A. D. de Maine sulle *Comunicazioni dell'ACM (Association for Computing Machinery)*⁸. Ma proprio a proposito delle modalità di archiviazione e dello spazio disponibile I.A. Tjomsland tenne un discorso nel 1980 al Quarto Simposio IEEE sui Sistemi di Archiviazione di Massa (MSS). Egli utilizza un paragone con la prima legge di Parkinson⁹ per spiegare come i dispositivi di archiviazione stiano accumulando una grande quantità di dati, anche obsoleti, e non vi sia modo di identificare la loro bontà.

In una ricerca prodotta da Hal B. Becker (1986) si pone l'accento sulla densità di codifica dei dati¹⁰; nei secoli questa è notevolmente aumentata, ma la domanda che i ricercatori si ponevano era determinare gli spazi (fisici o digitali) e le modalità d'uso di tutte le informazioni prodotte negli anni. Infatti, attraverso la pubblicazione dell'articolo "*Saving*

⁵ Rider, F. (1944). *The scholar and the future of the research library, a problem and its solution*. N.Y.: Hadham Press.

⁶ Price, D. J. de S. (1975). *Science since Babylon*. New Haven: Yale University Press.

⁷ Marron, B. A., & Maine, P. A. D. D. (1967). *Automatic data compression*. *Communications of the ACM*, 10(11), 711–715. doi: 10.1145/363790.363813.

⁸ Nel documento viene descritto un compressore a tre parti rapido e completamente automatico che può essere utilizzato con un corpo di informazioni per ridurre notevolmente i requisiti di archiviazione esterna definiti lenti e per aumentare la velocità di trasmissione delle informazioni attraverso un computer.

⁹ Cyril Northcote Parkinson applicò un'osservazione prettamente burocratica a un'ampia varietà di altre circostanze, realizzando che quando le dimensioni di qualcosa aumentano, la sua efficienza diminuisce conseguentemente.

¹⁰ Becker stima che "la densità di ricodifica raggiunta da Gutenberg era di circa 500 simboli (caratteri) per pollice cubo - 500 volte la densità di tavole di argilla [4.000 a.C., civiltà dei Sumeri]. Entro il 2000, la memoria ad accesso casuale dovrebbe archiviare $1,25 \times 10^{11}$ byte per pollice cubo".

*All the Bits*¹¹ si esplicitano i numerosi interrogativi posti sulla questione. L'obiettivo dell'autore Peter J. Denning e degli scienziati in generale di salvare tutti i dati li portò ad affrontare un problema da loro considerato ai limiti dell'impossibile da risolvere: il volume e la velocità del flusso di informazioni tendono a travolgere le reti, i dispositivi di archiviazione e i sistemi di recupero, nonché la capacità umana di comprensione. È quindi possibile costruire macchine volte a monitorare il flusso di dati di uno strumento o ad analizzare, descrivere un database e proporre un riepilogo statistico di ciò che c'è? Si riuscirebbe a progettare e realizzare dispositivi in grado di riconoscere o prevedere schemi in dati senza comprendere il significato dei modelli? Tali macchine potrebbero essere abbastanza veloci da gestire dati di grandi dimensioni in tempo reale? Con una risposta positiva agli interrogativi citati precedentemente una diretta conseguenza è una significativa riduzione del numero di bit da salvare e del pericolo di perdere scoperte latenti a causa della loro sepoltura in un immenso database.

Un notevole passo in avanti per risolvere questi problemi fu fatto quando, nel 1996, l'archiviazione digitale diventò più economica per immagazzinare i dati rispetto alle forme cartacee¹² aumentando le probabilità di successo nella realizzazione degli obiettivi.

Un avvenimento essenziale per la nascita del concetto di big data si ha quando, nell'articolo pubblicato a fine 1997 da Michael Cox e David Ellsworth "*Application-controlled demand paging for out-of-core visualization*"¹³, troviamo per la prima volta il termine *Big Data* tra gli archivi della biblioteca digitale ACM. Si usa questa accezione per descrivere situazioni in cui "la visualizzazione rappresenta una sfida interessante per i sistemi informatici: i set di dati sono generalmente piuttosto grandi, gravando sulle capacità della memoria principale, del disco locale e persino del disco remoto" (Cox, M., & Ellsworth, D., 1997). Di conseguenza nel momento in cui le informazioni, che per gli autori sono diventati big data, non si adattano alla memoria principale o non si adattano al disco locale, acquisire ulteriori risorse rappresenta la soluzione più comune.

Questo aspetto viene approfondito dalla pubblicazione di Michael Lesk intitolata "*How much information is there in the world?*"¹⁴, in cui si quantificano in alcune migliaia di petabyte le informazioni già catalogate e si stima un raddoppio attraverso altri dispositivi di archiviazione (dischi, nastri ecc.) entro il 2000. In più, questi giunge a due conclusioni ritenute improbabili prima degli anni 90: (a) la capacità di poter salvare tutto - nessuna informazione dovrà essere tralasciata, e (b) l'informazione non sarà più guardata da un essere umano ma si dovrà automaticamente decidere la serie di dati che la preziosa risorsa dell'attenzione umana prenderà in considerazione.

Un secondo riferimento diretto ai big data si ha nel 1998 quando John R. Mashey presenta a un meeting USENIX un *paper* intitolato "*Big Data ... and the Next Wave of Infrastrress*"¹⁵. Anche qui si fa riferimento maggiormente al volume e alla crescente capacità di archiviazione, attraverso cui assistere ad un'esplosione di dati ampiamente accessibili per spostare, creare, comprendere informazioni. Questa esplosione, se troppo rapida, può però generare degli svantaggi quali instabilità, imprevedibilità, versioni multiple oppure hardware obsoleti prima del loro ammortamento. Bisogna quindi far attenzione alle

¹¹ Denning, P. J. (1990). Saving all the bits. Moffett Field, Calif.?: Research Institute for Advanced Computer Science, NASA Ames Research Center. doi: <http://denninginstitute.com/pjd/PUBS/AmSci-1990-5-savingbits.pdf>

¹² Morris, R. J. T., & Truskowski, B. J. (2003). The evolution of storage systems. IBM Systems Journal, 42(2), 205–217. doi: 10.1147/sj.422.0205

¹³ Cox, M., & Ellsworth, D. (1997). Application-controlled demand paging for out-of-core visualization. Proceedings. Visualization 97 (Cat. No. 97CB36155). doi: 10.1109/visual.1997.663888

¹⁴ Lesk, M. (1997). How Much Information Is There In the World? Disponibile su <http://www.lesk.com/mlesk/ksg97/ksg.html>

¹⁵ In questo caso il termine *Infrastrress* nasce dalla combinazione tra "*infrastructure*" e "*stress*" e comprende tutti gli effetti negativi di un cambiamento troppo rapido a livello di sistemi e loro utilizzo (CPU, memorie, dischi...) e anche a livello di infrastrutture sottostanti (larghezze di banda, sistemi operativi, backup...).

debolezze che questo sistema innovativo comporta e saperle gestire qualora si presentassero tali problemi.

Due anni dopo gli stessi autori dell'articolo pionieristico sui big data pubblicano insieme ad altri scienziati *"Visually exploring gigabyte data sets in real time"*¹⁶ nelle Comunicazioni dell'ACM. È il primo articolo CACM a utilizzare il termine Big Data (infatti il titolo di una sezione dell'articolo è "Big Data per la visualizzazione scientifica"). L'articolo si apre con la seguente dichiarazione: "i computer molto potenti sono una benedizione per molti campi di indagine. Sono anche una maledizione; calcoli veloci emettono enormi quantità di dati"¹⁷. I set di dati per singole simulazioni che prima erano considerati grandi in quel momento non lo erano più, comprendere i dati derivanti da calcoli di alta difficoltà è uno sforzo significativo e la loro interpretazione complica ulteriormente le cose.

Verso la fine del 2000 Francis X. Diebold presenta all'ottavo congresso mondiale della Società econometrica un documento intitolato *"'Big Data' Dynamic Factor Models for Macroeconomic Measurement and Forecasting"*¹⁸, in cui afferma che la scienza biologica e fisica durante quegli anni è stata costretta a confrontarsi con i Big Data e spesso ne ha beneficiato. In particolare, dà una prima definizione di questi ultimi riferendosi alla quantità (e talvolta alla qualità) dei dati disponibili e potenzialmente rilevanti, in gran parte risultato di progressi tecnologici nei meccanismi di registrazione e archiviazione dei dati.

Un nuovo elemento distintivo dei big data, seppure inconsapevolmente, venne reso pubblico da Douglas Laney nel febbraio 2001¹⁹. L'analista, infatti, presenta indirettamente il modello delle "3V"²⁰, sebbene il termine stesso non compaia esplicitamente nella sua nota. Nel decennio successivo però il modello verrà generalmente riconosciuto come caratterizzazione delle tre dimensioni tipiche dei big data, a cui se ne aggiungeranno in un secondo tempo altre due per migliorare la descrizione. Il significato e le implicazioni di questi termini verranno analizzati nel paragrafo successivo.

Sette anni dopo numerosi scienziati fanno il punto della situazione sui big data, sul loro sviluppo e sull'impatto che potranno avere nel futuro. Questi fanno un parallelismo tra il cambiamento apportato dai motori di ricerca e l'elaborazione dei big data, entrambi hanno trasformato le modalità di accesso alle informazioni. Queste novità saranno capaci di trasformare le attività di aziende, ricevitori laboratori di ricerca, operazioni di difesa e intelligence all'interno di una nazione. "Il big-data computing è forse la più grande innovazione nell'informatica dell'ultimo decennio. Abbiamo appena iniziato a vedere il suo potenziale di raccolta, organizzazione ed elaborazione dei dati in tutti i settori." (Bryant, R. E. et al., 2008)²¹

Da queste solide basi si sviluppa un sempre più frequente utilizzo del termine *"Big Data"* accostato soprattutto al concetto di analisi, che a partire dalla seconda decade degli

¹⁶ Bryson, S., Kenwright, D., Cox, M., Ellsworth, D., & Haines, R. (1999). Visually exploring gigabyte data sets in real time. *Communications of the ACM*, 42(8), 82–90. doi: 10.1145/310930.310977

¹⁷ Kenwright, D., Banks, D., Bryson, S., Haines, R., van Liere, R., & Uselton, S. (1999). Automation or interaction: what's best for big data? *Proceedings Visualization 99* (Cat. No.99CB37067). doi: 10.1109/visual.1999.809940

¹⁸ Diebold, F. X. (2000). "Big Data" Dynamic Factor Models for Macroeconomic Measurement and Forecasting: A Discussion of the Papers by Lucrezia Reichlin and by Mark W. Watson. *Advances in Economics and Econometrics: Theory and Applications: Eighth World Congress Vol III*, 115–122. doi: 10.1017/ccol0521818745.004

¹⁹ Laney, D. (6 febbraio 2001). 3D Data Management: Controlling Data Volume, Velocity, and Variety. *Application Delivery Strategies*. Disponibile su <https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>

²⁰ Nella nota Doug Laney fa riferimento a tre caratteristiche principali quali Volume, Velocità e Variabilità, che successivamente verranno racchiuse nell'accezione di "3V" nel campo della big data analytics.

²¹ Bryant R. E., Katz R. H., & Lazowska E. D. (2008). *Big-Data Computing: Creating revolutionary breakthroughs in commerce, science, and society: A white paper prepared for the Computing Community Consortium committee of the Computing Research Association*. Disponibile su <http://cra.org/ccc/resources/ccc-led-whitepapers/>

anni duemila assume un ruolo da protagonista nelle ricerche scientifiche e sociali in moltissimi ambiti di applicazione.

1.1.2 Le 5 V

Sebbene non ci sia una definizione univoca all'interno della quale si racchiude il concetto di big data, gli studiosi sono concordi nell'accettare un modello formato dai 3 elementi introdotti da Laney più 2 aggiunti di recente e chiamato delle "5V". Ogni V racchiude un aspetto essenziale che un dataset analizzato deve avere per essere definito "big": Volume, Velocità, Varietà, Veracità, Valore (Baumann, P., & Riedel, M., 2019).

Il **Volume** concerne la quantità di informazioni ottenute o acquistate da un'impresa e le tecnologie non tradizionali utilizzati per raccogliere. Queste verranno poi analizzate per ricavare indicazioni rilevanti: giornalmente le aziende sono inondate da un flusso di dati in costante crescita, attraverso cui si accumulano facilmente terabyte o anche petabyte di informazioni. Vivendo in un'epoca di sovraccarico informativo (ogni minuto si effettuano circa 3 milioni e mezzo di ricerche su Google, su Facebook si mettono circa 3 milioni nuovi "like" e vengono scambiati nel mondo circa 16 milioni messaggi), il volume cresce esponenzialmente ed è difficile identificare un valore limite al di sotto del quale non si parla più di big data. Basti pensare che secondo analisti internazionali la produzione di dati nel 2009 era 44 volte minore rispetto a quella stimata quest'anno. Nonostante ora per convenzione si faccia riferimento ai big data quando si è in possesso di elevate moli di informazioni o di informazioni in forte crescita (>50% annua), i parametri restano abbastanza incerti e soggettivi considerando anche i trend futuri. È la "V" a cui si faceva originariamente riferimento quando si parlava nei primi articoli di big data poiché all'inizio l'attenzione principale era posta sulle dimensioni per l'archiviazione di informazioni.

Con **Velocità** si intende il tempo impiegato dagli strumenti tecnologici per processare i dati e/o la rapidità con cui questi vengono generati o ottenuti. Alcune attività richiedono risposte immediate, ecco perché massimizzare e velocizzare l'efficienza di analisi è essenziale (De Mauro, A., 2019). Le conseguenze della legge di Moore²² sono il principale veicolo che ha permesso questi progressi: aumento della velocità di calcolo dei microprocessori, crescita di velocità negli scambi di dati tra i dispositivi, ampliamento della capacità di archiviazione delle memorie esterne ed interne, passaggio dalla connessione analogica alla fibra ottica. Poiché la capacità di calcolo, trasferimento e analisi delle informazioni sono soggette a miglioramenti secondo progressioni esponenziali (la densità dei transistor²³ cresce) grazie al processo tecnologico, l'implementazione di questa caratteristica permetterà ai processori di operare sempre più celermente.

L'aspetto legato alla **Varietà** verrà analizzato più approfonditamente in uno dei paragrafi successivi relativo alle tipologie di dati. In generale questa caratteristica fa riferimento alla vasta gamma di informazioni che si trovano in un dataset anche perché sono generate da fonti diverse e registrate in formati differenti. L'eterogeneità e la poca organizzazione che compongono un insieme di dati grezzi da una parte hanno potenzialmente un grande valore

²² Gordon Moore nel 1965 sviluppò una legge ancora oggi valida Secondo la quale "la densità di transistor montati su chip tende a raddoppiare ogni 18 mesi".

²³ I transistor costituiscono la base per lo sviluppo dell'elettronica moderna e sono dispositivi a semiconduttore largamente usati anche nell'elettronica digitale. Le principali funzioni affidate all'interno di un circuito elettronico sono l'amplificazione di un segnale in entrata e il funzionamento da interruttore (switcher).

ma dall'altra necessitano di valide interpretazioni e rielaborazioni per fornire indicazioni corrette e utili agli analisti.

La quarta V sta per **Veracità** (nota anche come **Veridicità**) e indica il grado di affidabilità e quindi attendibilità dei dati raccolti o acquistati. Si deve porre particolare attenzione a questo aspetto per via dei cambiamenti continui a cui fonti, tecnologie e sistemi sono sottoposti nel corso degli anni che possono alterare le modalità di raccoglimento dei dati. Gli scienziati sono soliti dire *“Bad data is worse than no data”* per indicare che non tutti i dati che si possiedono hanno una validità assoluta e talvolta l'analisi delle informazioni sbagliate può condurre a conclusioni fuorvianti e non adatte alle finalità stabilite inizialmente. I pilastri fondamentali delle informazioni come qualità e integrità sono perciò imprescindibili per produrre analisi utili e veritiere.

Il **Valore**, aggiunto in seguito al modello base delle “3V”, rappresenta il fine ultimo nella definizione dei big data e anche della loro analisi. Questa, se effettuata secondo logiche coerenti, rappresenta una fonte inestimabile di valore. Si generano infatti conoscenze spesso valide soggettivamente ma essenziali vista la capacità dei dati di creare valore aggiunto attraverso le analisi predittive, i processi e le strategie innovative. Ciò accade perché le imprese che sfruttano correttamente queste elaborazioni riescono a indirizzare la gestione delle attività verso i bisogni del mercato. In conclusione, quindi, per definire un insieme di informazioni come big data non è necessario soltanto avere un dataset valido e selezionato ma bisogna aggiungere strumenti di analisi mirata e soprattutto produttivi. Senza la Big Data Analytics infatti i dati non assumerebbero il valore odierno ma rappresenterebbero soltanto una serie di informazioni sterili e non controllabili dalle sole capacità umane.

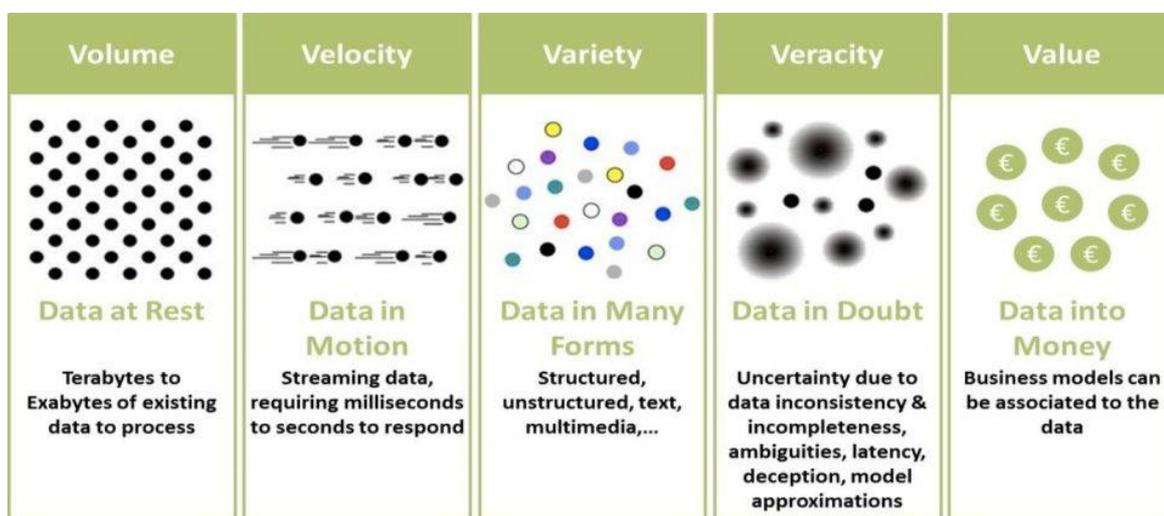


Figura 2: Le 5V dei Big Data

1.2 Qual è il loro valore?

I big data hanno un valore intrinseco ma questo è inutile finché non viene scoperto quel determinato valore. L'aspetto su cui ci si tende a soffermare è legato alla valenza che non tanto i big data quanto la loro "Analytics" ricopre sull'implementazione di processi produttivi, di strategie efficaci di *decision-making* e in generale di gestione dell'attività dell'azienda. Questo tipo di informazioni supporta soprattutto i top e middle manager nell'intraprendere scelte avvedute e nel ridurre la probabilità di errore: non ci si affida solo all'esperienza e all'intuizione dei dirigenti ma ad elaborazioni oggettive che possono anche andare oltre la razionalità limitata umana, se correttamente analizzate.

Il valore dei dati si sviluppa ulteriormente, secondo Cukier e Mayer-Schönberger, sia da un utilizzo primario per ricavare informazioni e stilare previsioni, sia dai possibili utilizzi secondari ottenibili attraverso dati ricombinati. In sostanza, il valore secondario è basato sul riutilizzo futuro: è possibile infatti approfondire l'ambito di una ricerca combinando più dataset. Si otterranno così informazioni estese oppure più generalizzate, a seconda della tipologia di dati uniti. È tuttavia necessario considerare che i dati, con il passare del tempo, diventano obsoleti più o meno velocemente, a seconda dei mutamenti nei trend e nel comportamento umano oppure nell'innovazione tecnologica dei software che li producono. Questa perdita di valore si riversa direttamente nella bontà dei dataset stessi e pertanto è richiesto che gli analisti si tengano aggiornati di continuo utilizzando modelli informatici sofisticati abbastanza da permettere una pulizia dei dati superflui e/o vecchi.

Non bisogna però dimenticare che oggi i big data per le imprese stanno diventando anche asset che valgono miliardi di dollari, il cui prezzo sarà destinato a salire negli anni. Basti pensare ad alcune delle più grandi aziende tecnologiche del mondo (IBM, Google, Microsoft...) dove gran parte del valore offerto proviene dall'acquisto dei loro dati che vengono prontamente analizzati per i più disparati scopi dagli acquirenti, in modo tale da generare un vantaggio competitivo nei confronti dei concorrenti. Stando al Report 2020 sull'industria Big Data Analytics²⁴ il mercato globale, che fu valutato a 37,34 miliardi di dollari nel 2018, dovrebbe raggiungere 105,08 miliardi di dollari entro il 2027 con un CAGR²⁵ del 12,3% per tutto il periodo di previsione dal 2019 al 2027. Solo in Italia il valore di questo mercato nel 2019 ha registrato una crescita del 23% rispetto all'anno precedente (1,5 miliardi di dollari), attestandosi su 1,85 miliardi di dollari (Osservatorio PoliMi, 2019).

1.3 Tipologie di dati

Quando si parla di eterogeneità di dati non si riesce a capire appieno la varietà delle informazioni che possono generarli; i Big Data comprendono tutto, dalle transazioni in dollari ai tweet, dalle sequenze genetiche alle ricerche effettuate sul web, dagli audio ai video. Pertanto, si richiede che tutte queste informazioni siano integrate per l'analisi e la gestione dei dati. Ciò è più difficile di quanto sembri.

²⁴ Research and Markets (Ed.). (2 marzo 2020). Big Data Analytics Industry Report 2020 - Rapidly Increasing Volume & Complexity of Data, Cloud-Computing Traffic, and Adoption of IoT & AI are Driving Growth. Disponibile su <https://www.globenewswire.com/news-release/2020/03/02/1993369/0/en/Big-Data-Analytics-Industry-Report-2020-Rapidly-Increasing-Volume-Complexity-of-Data-Cloud-Computing-Traffic-and-Adoption-of-IoT-AI-are-Driving-Growth.html>

²⁵ Compounded Annual Growth Rate o tasso annuo di crescita composto, rappresenta la crescita percentuale media di una grandezza in un lasso di tempo.

TIPO DI INFORMAZIONI TRATTATE IN MODO ANALITICO/QUANTITATIVO



Figura 3: percentuale sul totale dei dati processati nella BDA

Nel raccogliere i dati per poi essere utilizzati nei vari algoritmi si possono osservare tre tipologie di informazioni: dati strutturati, dati destrutturati e dati semi-strutturati (Marr, B., 2019)

I **dati strutturati** sono metainformazioni²⁶ inserite manualmente utilizzando uno schema predefinito e sono generalmente organizzati e conservati in un formato tabulare. Si immagini una tabella in cui ogni cella contiene un valore discreto: lo schema rispecchia il modo in cui sono organizzati i dati, la riga di intestazione della tabella utilizzata per descrivere il valore e il formato di ogni colonna. La tabella determina poi i vincoli richiesti per rendere i dati consistenti e calcolabili. Ci sono due fonti di dati strutturati:

- *Machine generated*: come si intuisce dal nome vi si racchiudono tutti i dati generati automaticamente da dispositivi tecnologici quali weblogs, GPS, sensori, informazioni di utilizzo acquisite da server, sistemi finanziari e perfino dispositivi medici. Per esempio un algoritmo riesce a distinguere un articolo da una pagina web grazie ai dati strutturati che comunicano al motore di ricerca le informazioni necessarie per distinguere i documenti.
- *Human generated*: comprende tutti i dati che noi inseriamo sui computer come informazioni personali, film guardati o siti web visitati. Possono essere usati dalle aziende per capire meglio il comportamento degli utenti e, come succede per Youtube o Netflix, raccomandare suggerire contenuti affini ai loro gusti.

Ancora, combinando i due tipi di dati strutturati un motore di ricerca come Google impiega i metadati di un sito o una pagina e alcune informazioni da noi inserite per fornire agli utenti i cosiddetti *rich snippets*²⁷.

I **dati destrutturati** sono formati tendenzialmente da file di ogni dimensione o formato (video, audio, testi, immagini, post sui social media, blog etc.) e, non essendo organizzati, sono sottoposti

²⁶ *Metainformazioni* o *metadati* sono termini in uso nel linguaggio informatico per definire un insieme di informazioni sui dati, spesso considerati anche come “dati su dati”. Sono necessari affinché un documento informatico venga correttamente formato, gestito e conservato nel tempo.

²⁷ Contenuti arricchiti da informazioni e grafica: oltre a veicolare dettagli aggiuntivi colpiscono l'attenzione dell'utente perché hanno una maggiore visibilità sui risultati di ricerca.

ad un'analisi molto più complessa rispetto alla precedente tipologia. Circa l'80% dei dati che si hanno è da ritenere non strutturato, sono comunque dotati di una struttura interna ma non è predefinita attraverso modelli. Le aziende che lavorano con questo tipo di dati hanno bisogno di sviluppare algoritmi all'altezza della complessità delle informazioni per ricavarne indicazioni efficaci. Una volta selezionate le informazioni rilevanti, tralasciate quelli irrilevanti e scelto la giusta metodologia di analisi, il valore generato consentirà di intervenire concretamente sui servizi forniti o sui processi interni aziendali. Ogni qual volta un'azienda raccoglie feedback dai clienti, raccoglie dati destrutturati; le risposte sotto forma di testo nei sondaggi sono dati non strutturati. Anche qui è valida la distinzione tra:

- *Machine generated*: Includono video di sicurezza, immagini satellitari, acquisizioni di dati dei radar e altro ancora.
- *Human generated*: Sono considerati tali tutti i contenuti video, audio, immagini, testi generati da noi sui social media, sui siti web, sui dispositivi mobili e sull'email. Gli esempi sono le foto caricate su Instagram oppure i commenti scritti sotto un post di Facebook.

Infine, un prodotto ibrido tra dati strutturati e dati destrutturati è costituito dai **dati semi-strutturati**, che possiedono caratteristiche proprie delle due tipologie già presentate. Mancano quindi di una struttura fissa per la catalogazione secondo certi schemi ma hanno alcune proprietà organizzative che rendono l'analisi più semplice rispetto ai dati non strutturati. I vantaggi di questa classe di file sono la flessibilità e la facilità di interazione tra risorse eterogenee. Allo stesso tempo però le debolezze consistono nella complessità di interpretare le relazioni tra i dati e nella difficoltà di archiviazione per via di schemi poco rigidi. In particolare, nel formato di file Xml i vincoli strutturali sono assenti però ci sono, contemporaneamente, sistemi che organizzano le informazioni inserite e ne orientano la logica aumentando l'interoperabilità.

1.4 Modalità di analisi dei Big Data

Mentre inizialmente le attenzioni dei ricercatori si sono concentrate sul volume e sulla crescita nella dimensione dei dati, negli ultimi anni si è posto sempre di più l'accento su modalità di analisi abbastanza efficaci da produrre notevoli miglioramenti nel sistema. La Big Data Analytics è, come abbiamo visto prima, la modalità attraverso cui si realizza il processo di analisi dei big data. Esistono però diverse tipologie di analisi e non bisogna confondere sistemi di Business Intelligence con quelli di Big Data Analytics (Carbone, M.R., 2020).

La differenza principale sta nel fatto che la Business Intelligence utilizza metodi statistici basati su analisi descrittive, sistemi più tradizionali che però elaborano dati con alta densità di informazione. I modelli realizzati sono semplicistici rispetto alla BDA e come risultato si ottengono indicazioni e rappresentazioni che hanno una minore abilità di adattamento alle mutevoli esigenze del mercato. La tipologia di gestione dei dati utilizzata è la *descriptive analytics* (analisi descrittiva), ci dà informazioni sulla situazione corrente ma anche passata delle funzioni, business unit o informazioni esterne esaminate. È infatti largamente diffusa tra gli analisti delle aziende perché interagire e lavorare con questo genere di sistemi è immediato e abbastanza facile e non richiede competenze particolarmente complesse da sviluppare.

Procedendo invece con la definizione di Big Data Analytics è necessario sapere che questa utilizza la statistica inferenziale e alcuni concetti di identificazione dei sistemi non lineari per connettere enormi insiemi di dati tra loro. Si può assumere che questa sia composta da due elementi principali:

Big Data e Advanced Analytics (European Banking Authority, 2020). Quest'ultimo processo coinvolge tecniche di analisi predittive e prescrittive, finalizzate a raccomandare e comprendere azioni basate sull'analisi delle informazioni (Big Data).

Le tecniche di *Advanced Analytics* sono classificate secondo quattro modalità: *diagnostic analytics*, *predictive analytics*, *prescriptive analytics*, *autonomous and adaptive analytics* (European Banking Authority, 2020).

1.4.1 Strumenti per l'analisi

Prima ancora di procedere con la spiegazione dei singoli tipi di analisi è necessario comprendere le metodologie e gli strumenti che i data scientist²⁸ utilizzano per elaborare le informazioni. In questo lavoro verranno analizzati i metodi analitici più diffusi quali Data Mining, Machine Learning (ML) e Artificial Intelligence (AI) (Statistical Analysis System - SAS).

L'**Intelligenza Artificiale (AI)** ha come obiettivo lo studio e la realizzazione di agenti (macchine) capaci di svolgere mansioni simili a quelle degli esseri umani, comprendere il loro ambiente talvolta mutevole e massimizzare le probabilità di successo attraverso le azioni. Altre due caratteristiche simili a quelle della mente dell'uomo che questa tecnologia mira ad ottenere sono le capacità di *problem solving* ed apprendimento. Per questo motivo è una risorsa molto preziosa per le aziende. Il suo scopo è quello di migliorare in modo significativo le abilità e le attività degli esseri umani: dalle auto con guida autonoma agli oggetti in grado di eseguire analisi autonome e prendere decisioni senza l'intervento umano (videocamere, termometri...). Nell'attuale progettazione dei sistemi basati sull'AI, strumenti quali il *deep learning* (apprendimento profondo, parte anche del Machine Learning) e sul *natural language processing* (NLP - elaborazione del linguaggio naturale) sono la base per un efficiente raggiungimento delle finalità. Utilizzandoli, i computer possono imparare a svolgere compiti specifici processando i big data e riconoscendo in essi alcuni modelli.

Il deep learning è una parte del ML ed è strettamente legato alle funzioni e alla struttura del cervello perché progetta algoritmi chiamati reti neurali: questo non mira solo ad apprendere in autonomia ma anche su più livelli di profondità, esattamente come accade con la mente umana. Le applicazioni di più successo del deep learning sono i giochi automatici quali scacchi e Go (imparando a giocare automaticamente e a battere campioni umani), la traduzione simultanea e la classificazione di oggetti all'interno di una fotografia.

Il natural language processing (NLP), invece, si concentra molto sull'area tematica della linguistica e ha come obiettivo l'analisi e la comprensione del linguaggio naturale (quello degli uomini) che sia scritto o parlato. Lo strumento permette di elaborare sia la sintattica, e quindi creare associazioni, relazioni e classificazioni tra le categorie morfologiche delle frasi, sia la semantica, interpretando il linguaggio per comprendere il significato in base al contesto e le varie categorie del contenuto. Un esempio sono gli assistenti vocali sviluppati da grandi compagnie quali Siri (Apple) ed Alexa (Amazon) che svolgono numerose funzioni grazie al NLP.

La maggior parte delle informazioni presenti nei dataset non sono direttamente evidenti e allo stesso tempo è difficile che analisi condotte prettamente dagli umani vengano concluse

²⁸ La figura del data scientist ha il compito di preparare i dati per l'analisi, sviluppare strategie per l'analisi, esplorare, analizzare e visualizzare i dati, creare modelli con i dati nonché di implementare i modelli nelle loro applicazioni.

in tempi brevi. Per **Data Mining** si intende l'individuazione di informazioni implicite, e quindi non risapute a priori, tramite estrapolazione mirata da grandi banche dati chiamate *data warehouse*²⁹. Il processo può svolgersi utilizzando anche multiple *data warehouse* e incrociandone le informazioni, si ottiene così una prospettiva più precisa oppure più generale a seconda degli scopi. Le tecniche applicate alle operazioni di estrazione dei dati sono varie (tradizionali e non, come il Machine Learning) e in gran parte automatizzate, consistendo in specifici software e algoritmi progettati e realizzati per una specifica finalità che raggiungono velocità di analisi impensabili prima d'ora. Negli ultimi anni, in particolare, si utilizzano modelli di reti neurali (meccanismi di apprendimento che si evolvono con l'integrazione di nuove informazioni), analisi delle associazioni (relazioni causali che permettono di individuare anomalie o pattern ricorrenti), *clustering* (raggruppamento dei dati in classi omogenee prima totalmente sconosciute) e *time series* (regressioni a più variabili che si concentrano sull'aspetto predittivo) per svolgere attività di data mining.

Il **Machine Learning (ML)** è una branca dell'Intelligenza Artificiale e si occupa di creare algoritmi che apprendono o migliorano la performance in base ai dati che analizzano. I già citati algoritmi servono per risolvere problemi che non hanno strategie definite in modo tale che i computer impari un'attività e, attraverso in processo iterativo, migliori la sua esecuzione di volta in volta (prodotti suggerimenti da Amazon e contenuti suggeriti da Netflix, intercettazione di frodi...). I due modelli più diffusi di ML sono il *supervised learning* e l'*unsupervised learning*.

Nel primo caso gli algoritmi vengono allenati ad affinare una specifica attività utilizzando schemi già classificati, infatti chi li inserisce conosce già gli output dati gli input. L'algoritmo di apprendimento impara abbinando input ad output corrispondenti così come precedentemente stabilito, compara i risultati per trovare eventuali errori e poi modifica il modello di conseguenza. Il *supervised learning* si usa principalmente per attribuire il giusto valore a dati non ancora classificati e quindi anticipare eventi futuri basati sui dati storici; classificazione, regressione (sfruttare dipendenze lineari o non lineari per anticipare l'andamento di una variabile sulla base di altre variabili) e *gradient boosting* (algoritmo secondo il quale si passa da un insieme di *weak learner*³⁰ a uno *strong learner*³¹ soprattutto mediante alberi decisionali³², il suo scopo è quello di minimizzare una generica funzione di costo) sono alcuni modelli molto utili. Una diretta applicazione consiste nell'individuare possibili clienti in di una compagnia assicurativa che, più facilmente rispetto agli altri, potrebbero chiedere un risarcimento. Il secondo metodo è adatto per informazioni che ancora non sono state classificate da nessuno, il sistema non ha quindi la possibilità di comparare la sua "risposta" con quella giusta. L'algoritmo che compone un apprendimento non supervisionato è adatto per analizzare i dati transazionali³³ ed ha il compito di riconoscere ciò che gli viene mostrato, il suo scopo è esplorare i dati per trovare alcune strutture intrinseche, se necessario anche valori anomali. Tre delle tecniche più usate sono mappature *nearest-neighbor* (riconoscimento di pattern per la classificazione di dati basandosi sulle proprietà dei dati vicini a quello considerato), decomposizione ai valori singolari, algoritmi di *clustering k-means* (volto a trovare in ogni dataset un numero fisso di cluster). Per esempio, le aziende che utilizzano l'apprendimento non supervisionato

²⁹ Il *data warehouse* è un sistema di data management che raccoglie e conserva grandi quantità di dati, anche storici, che hanno più origini.

³⁰ Modelli deboli, generano risposte poco migliori rispetto a quelle che si avrebbero in modo casuale.

³¹ Modello complesso e più attendibile rispetto al weak learner, si avvicina ad un modello ideale.

³² Strumento che funge da supporto alle decisioni ed ha una struttura ad albero.

³³ Anche detti On Line Transaction Processing (OLTP), i dati transazionali vanno a supporto delle attività di gestione e operative. Registrano infatti tutte le informazioni scambiate nelle funzioni tra e nelle funzioni aziendali e tra l'azienda i propri clienti, mostrandole all'analista.

possono rivolgere campagne di marketing mirate per consumatori con caratteristiche simili, raggruppati seguendo elementi distintivi sempre tramite l'*unsupervised learning*.

Uno strumento aggiuntivo di analisi è il **cloud computing**, costituito da servizi di archiviazione o di elaborazione di dati che offre il provider al cliente finale. Tutto ciò si svolge tranquillamente in rete in maniera virtuale (connessione internet e web in primis), senza l'utilizzo delle tradizionali componenti fisiche come gli hardware e anche i software. Le risorse e le applicazioni sono quindi gestite online; ciò comporta un grande beneficio soprattutto per le aziende che non puntano più ad utilizzare reti private, perché molto più costose del cloud in termini di sviluppo, innovazione, componenti e risorse umane.

Le applicazioni presenti sul cloud vengono appunto aggiornate in automatico dai fornitori del servizio e sono immediatamente operative. Questa tipologia di computing usufruisce della connessione internet per consentire il salvataggio dei dati, lo svolgimento dei calcoli e il corretto funzionamento dei software in maniera molto più efficiente e soprattutto centralizzata. Difatti le tecnologie cloud stanno avendo un enorme successo per la loro capacità di accedere a risorse computazionali mediante i numerosissimi dispositivi, anche portatili, e la rete, espandendo la possibilità di connessione mobile e da qualsiasi parte del mondo in cui gli utenti si trovano.

Esistono due tipologie di piattaforme cloud: *private* e *public*.

Il primo consiste nella fornitura del servizio da parte di terzi alle aziende e individui in generale: in questo caso vi sono evidenti vantaggi economici perché è una soluzione appetibile per le imprese di qualsiasi dimensione. Tutti i costi di gestione del servizio sono affidati al provider e chi vuole utilizzarlo deve solamente pagare una sorta di abbonamento. Il private cloud Consiste invece in un sistema informatico interno, realizzato dall'azienda e utilizzabile solo internamente. In questo caso l'ambiente privato cloud è totalmente autogestito, nonostante le sue caratteristiche d'uso siano simili a quelle del pubblico. La soluzione prevede poi una maggiore adattabilità dei sistemi alle esigenze proprie della società e tendenzialmente un minor rischio relativo alla sicurezza dei dati.

Il cloud offre infine servizi in tre aree differenti: SaaS (*Software as a Service*), PaaS (*Platform as a Service*) e IaaS (*Infrastructure as a Service*).

Il SaaS è un modello di cloud progettato per distribuire un software on demand come servizio. Questo consiste nell'ospitare un qualsiasi software nel cloud, utilizzandolo come fosse installato direttamente sul dispositivo al livello locale. All'utente viene data la possibilità di utilizzare direttamente l'applicazione senza preoccuparsi di configurarla e mantenerla. Nonostante noi non lo realizziamo, facciamo uso dei SaaS molto più spesso di quanto si pensi: Google Docs, Microsoft 365 e Google Sheets sono tra gli esempi più famosi di Software as a Service.

Per quanto riguarda il PaaS, questo funge da servizio intermedio tra SaaS e IaaS, essendo un'implementazione software che si fonda sulla IaaS. Non a caso suo scopo è ottimizzare le operazioni che preparano la condizione di sviluppo di un software, fornendo gli strumenti necessari per questa parte di cloud. È di conseguenza molto utile per gli sviluppatori di software perché favorisce anche la collaborazione tra i vari utenti.

In ultimo, l'IaaS Offre una virtualizzazione al 100 % dell'hardware di un tradizionale server (RAM, CPU, spazio di archiviazione...). Una volta che l'utente ha scelto le caratteristiche che più si addicono all'uso che deve farne nel cloud, egli può usufruire del servizio anche collegandosi con dispositivi con bassa potenza di calcolo ma ottenere lo stesso prestazioni molto soddisfacenti.

1.4.2 Tipologie di *Analytics*

La ***Diagnostic Analytics*** (analisi diagnostica) è una forma sofisticata di analisi dei dati orientata al passato: non ha l'obiettivo di capire cosa è successo (analisi descrittiva) ma di scoprire le cause che hanno portato al verificarsi di determinati eventi e comportamenti. Per fare ciò, gli algoritmi utilizzano dati di proprietà e sfruttano le informazioni esterne per aiutare le aziende a trovare una soluzione rapida evitando di ripetere gli stessi errori. Prendendo come base le analisi descrittive si procede ad una loro integrazione con tecniche avanzate quali Data Discovery, Data Mining e correlazioni per identificare le anomalie. Per esempio, attraverso il Data Discovery si osservano pattern e trend nascosti e relazioni causali che sono difficilmente individuabili altrimenti; per aiutare questo processo spesso gli esperti stilano modelli d'impatto visivo e quindi intuitivi (tabelle pivot, mappe di calore...).

Nella ***Predictive Analytics*** (analisi predittiva) si ha un orientamento verso il futuro che mira a supportare il business, specialmente nella parte gestionale, sulla previsione di cosa potrebbe accadere analizzando i dati rivolti al passato. Gli schemi trovati nei dati storici e transazionali sia strutturati che destrutturati vengono utilizzati per identificare rischi e opportunità futuri. Ciò comporta l'uso di avanzate tecniche legate alla statistica inferenziale come i modelli di regressione e di *forecasting*, Machine Learning e Data Mining Offrono un prezioso contributo in merito. L'analisi predittiva permette alle aziende di svolgere un ruolo proattivo e anticipatore nelle azioni e nei risultati perché, se le previsioni sono accurate, i manager potranno prendere decisioni più efficaci e guidare più velocemente l'azienda verso il raggiungimento degli obiettivi a lungo termine.

Prescriptive Analytics (analisi prescrittiva) è un termine che include la combinazione di tecniche analitiche sia rivolte all'indietro sia al futuro per suggerire la soluzione ottimale basata sui dati disponibili in un preciso momento. Questa non solo prevede che cosa accadrà ma, integrando informazioni del passato, spiega anche perché accadrà e fornisce utili raccomandazioni in merito alle azioni da intraprendere. L'Intelligenza Artificiale è lo strumento migliore per effettuare questa analisi dato che è caratterizzata da una certa flessibilità per modellare diversi risultati aziendali basati su rischi e scenari futuri, in modo tale da ottimizzare l'impatto che ciò avrà sull'azienda. Per via della complessità nell'amministrazione delle analisi prescrittive e l'alto livello di competenze necessarie per svilupparle la maggior parte delle aziende non le utilizza ancora, tuttavia sono un ottimo strumento per le finalità sopra indicate.

L'***Autonomous and Adaptive Analytics*** (analisi autonoma e adattiva) è la tecnica più complessa e utilizza modelli di analisi predittiva che apprendono automaticamente e aggiungono risultati aggiornandosi in tempo reale tramite il Machine Learning. Ciò include la capacità di auto-generare nuovi algoritmi con approfondimenti suggeriti per compiti futuri, in base a correlazioni e modelli nei dati che il sistema ha identificato e su volumi crescenti di Big Data.

Capitolo 2

“Big Data, Big Dupe”: alcune critiche

Se da una parte possiamo definire i big data e la loro *analytics* il nuovo oro per le numerose possibilità offerte dalla loro analisi, non è tutto oro ciò che luccica. Accanto ai grandi benefici e alle potenzialità che questo fenomeno offre ci sono numerosi aspetti critici da menzionare per rendere più completo il quadro di analisi. Dalle problematiche nella definizione a quelle legate alla privacy, i campi che necessitano dei chiarimenti sono numerosi e ora come ora costituiscono alcune barriere nella completa accettazione dei benefici potenzialmente riconoscibili a questa tipologia di dati. Tante infatti sono le voci che criticano i big data e ritengono che il fenomeno sia sopravvalutato e talvolta anche pericoloso nella sua larga diffusione. Ciò che più preoccupa è la mancanza di determinati principi volti a prevenire l’uso improprio delle informazioni e a conferire una certa tutela agli interessati. Se non si strutturano una serie di regolamentazioni valide e se non si diffonde la giusta cultura dei big data, come sottolinea il presidente dell’AGCOM³⁴ Angelo Cardani nel 2019, una diretta conseguenza potrebbe essere la possibile “alterazione dell’ecosistema informativo planetario”.

Come osservabile nella **Figura 4**, è possibile raggruppare le “sfide” che la Big Data Analytics si prefigge di risolvere nel futuro in tre categorie strettamente legate al ciclo di vita dei dati (Sivarajah, U., et al., 2017).

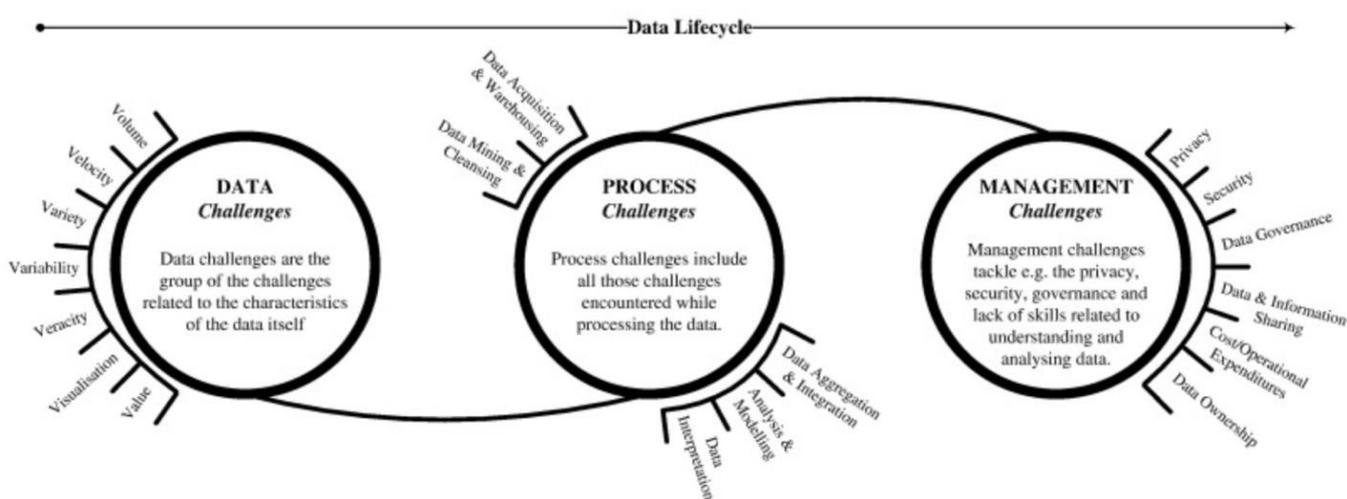


Figura 4: le varie fasi del ciclo di vita della BDA e le rispettive sfide

Nella prima fase (**Data Challenges**) si osservano le criticità che fanno riferimento alle caratteristiche intrinseche dei big data (e.g. le “5V”, la qualità, le fonti), inclusi i dubbi sulla loro definizione.

³⁴ Acronimo di Autorità per le Garanzie nelle Comunicazioni; questo ente ricopre principalmente due ruoli: tutelare i consumi nelle libertà fondamentali degli individui ed assicurarsi che all’interno del mercato vi sia una corretta competizione tra le aziende.

La fase centrale (**Process Challenges**) del ciclo di vita riguarda i processi utilizzati dalla BDA i cui problemi riguardano come prendere i dati, integrarli, trasformarli, selezionare il giusto modello per la loro analisi e riportare i risultati.

Nell'ultima parte (**Management Challenges**), invece, si affrontano le difficoltà connesse alla gestione di queste informazioni: la sicurezza, la privacy, gli aspetti etici e la governance sono punti fondamentali di questa sezione.

In generale, ogni soluzione offerta per valutare ed analizzare i big data dovrebbe rispettare anche alcuni principi fondamentali dell'etica quali la chiarezza, la prevenzione dei danni, l'autonomia umana e la correttezza (European Banking Authority, 2020).

In questo capitolo verranno affrontate alcune di queste criticità, in particolare quelle che costituiscono i rischi maggiori per lo sviluppo e il miglioramento della Big Data Analytics.

2.1 Difficoltà nella definizione

Come fatto presente nel capitolo precedente ci sono molte incertezze nel dare una definizione completa, univoca e formale del termine big data. Non a caso, Stephen Few (2018) nel suo libro afferma che *“ci sono quasi tante definizioni di big data quante sono le opinioni delle persone”*.

Talvolta, ed è il caso di uno dei libri più famosi sull'argomento, una spiegazione del concetto non è nemmeno ritenuta necessaria. Gli autori si limitano solo a dire che *“scienze come l'astronomia e la genomica, che per prime hanno sperimentato l'esplosione negli anni 2000, coniarono il termine big data. Il concetto sta ora emigrando a tutte le aree della conoscenza umana”* (Mayer-Schönberger, V., Cukier, K., 2013). Così facendo si crea un alone di incertezza e di opinabilità su ciò che genericamente può essere definito come big data. Ciò ha un impatto negativo sulla concezione stessa del fenomeno e su tutti coloro che vi si confrontano quotidianamente mediante l'archiviazione, l'analisi e la gestione di queste informazioni.

In più, non molti aspetti nelle definizioni di cui si è parlato nel secondo paragrafo del primo capitolo differiscono dal concetto di dati che già esisteva in passato. Un ulteriore punto di debolezza è costituito dall'ambiguità implicita nel classificare un qualsiasi dataset come big data, non si hanno infatti elementi certi che stabiliscano una linea di demarcazione tra le due tipologie di dati (Few. S., 2018).

Si presenteranno quindi i problemi associati ai numerosi fattori che, attraverso le varie interpretazioni, caratterizzano in maniera più frequente le definizioni dei big data (Few. S., 2018).

La prima categoria da analizzare è il volume: si parla sempre di dataset molto vasti ma nessuno ha mai stabilito una soglia al di là della quale i semplici dati diventano big data. Spesse volte si fa riferimento alla grandezza attraverso la difficoltà che gli analisti hanno a processare insieme tutte le informazioni usando software e tecniche tradizionali. Anche qui, però, non si nota un chiaro riferimento alla connotazione “tradizionale” di questi strumenti visto il continuo progresso nella data science e la velocità con cui cambiano le metodologie di analisi. Restringere tale accezione aggiungendo la richiesta di numerosi sforzi in termini di tempo, creatività e denaro da soddisfare per trovare soluzioni innovative e funzionali (strutturare nuovi algoritmi, ideare metodi di archiviazione) non pare risolutivo, perché lascia ancora ampie interpretazioni.

Un'altra descrizione diffusa è associare il volume ad informazioni che non si adattano facilmente ad un database relazionale³⁵ standard. In questo caso sono ancora stabiliti limiti arbitrari, molti database si originano dalla grandezza dell'hardware in cui sono archiviati. Per esempio, seguendo questa definizione di volume non si potrebbe ritenere accettabile connotare come big data le informazioni presenti su più fogli Excel e contenute in un qualsiasi computer; è così chiaro come questa affermazione teoricamente possa ricadere nella descrizione precedente ma allo stesso tempo sia troppo semplicistica e richieda vincoli più specifici.

In ultimo, alcuni studiosi hanno interpretato la definizione di big data basata sul volume tramite informazioni contenute su piattaforme di archiviazione che “raccolgono e immagazzinano tutto”, inteso come memorizzazione di dati a tutti i livelli, anche quelli che sembrano inutili al momento potrebbero rivelarsi valide in futuro in qualche modo. Una simile visione però comporta enormi costi supportati dalle aziende per l'archiviazione illimitata di dati e molti esperti affermano che buona parte delle informazioni generate nel mondo sono superflue e continueranno ad esserlo negli anni a venire.

Alcune spiegazioni enfatizzano l'aspetto dell'eterogeneità (che coincide con la “V” di varietà), sottolineando come i big data siano tali soprattutto per la loro generazione da disparate fonti e la conseguente presenza sotto varie tipologie.

Nemmeno considerando questa descrizione si stabilisce una soglia: qual è il numero minimo di fonti di dati necessari per qualificarli come abitata? Quali sono le tipologie accettabili? C'è un numero minimo anche di formati in cui dati possono presentarsi? Le numerose definizioni non chiariscono né menzionano mai questi aspetti.

Anche la complessità dei dati dei big data è stata ed è tuttora oggetto di discussione perché implica opinioni diverse e autoreferenziali. D'altronde questo fattore è sempre stato presente nell'analisi dei dati in generale. Come molti aspetti delle spiegazioni precedenti anche questo manca di limitazioni poiché non è chiaro ciò che si intende quando si parla di dati con informazioni complesse. Parlare di complesso nelle tecnologie che riescono a sviluppare le piccole e medie imprese nella BDA è completamente diverso rispetto alla complessità di informazioni e strumenti a disposizione e progettati dalle grandi imprese e dalle multinazionali. La soggettività che questa prospettiva presenta va assolutamente rivista e delineata secondo criteri validi a tutti i livelli.

Strettamente legata alla complessità c'è un'altra connotazione, oggetto di incertezza nello studio dei big data: spesso si afferma che non sono i dati in sé a caratterizzare i big data ma le metodologie utilizzate per darvi un senso. Il problema qui si riscontra dal momento in cui si introducono le metodologie di analisi avanzata delle informazioni a fungere da confine tra big data e semplici dati. Ma quali sono esattamente gli strumenti di cui si parla? Di solito però la risposta varia a seconda degli analisti e ancora una volta l'ostacolo verso una definizione univoca è l'ambiguità. È possibile affermare che tendenzialmente si tratta di metodi capaci di cambiare in modo sostanziale l'uso dei dati rispetto al passato ma, così facendo, non si giunge comunque ad una spiegazione oggettiva.

Ultime ma non meno importanti sono le problematiche relative all'uso improprio che si fa del termine da parte di tutti coloro che ne parlano per enfatizzare le loro argomentazioni. Questo contribuisce alla diffusione di idee e caratteristiche sbagliate sui big data e all'aumento della confusione che già di per sé crea l'incertezza sulla definizione. Un'affermazione che rappresenta appieno il concetto fu fatta da Mike Caravetta, Data Scientist e manager della Ford Motor Company, secondo cui i big data sono *“Un racconto – sia esso attraverso informazioni grafiche o altri supporti visuali che lo spiegano in un modo che aiuta gli altri a comprendere tra vari settori”*.

³⁵ Anche detto RDBMS (Relational Database Management System), il database relazionale si basa sull'archiviazione delle informazioni acquisite in un insieme di tabelle di dati bidimensionali. Queste tabelle sono suddivise in celle e catalogate in righe e colonne. Nonostante siano memorizzati su archivi (e quindi tabelle) diversi grazie alle relazioni, da cui prende il nome questo database, i dati possono essere elaborati come fossero conservati in un unico archivio.

La precedente asserzione dimostra come al giorno d'oggi il termine big data sia talmente tanto diffuso che ci si riferisce ad esso per indicare qualsiasi argomento in voga e relativo all'informazione tecnologica. Questa conclusione è diretta conseguenza del fatto che esistono molteplici spiegazioni tra di loro confuse: possono essere facilmente manipolate per intendere ciò che si vuole.

Per finire, si possono riassumere le debolezze nella definizione del termine big data in due punti:

1. Non è possibile per ora distinguere chiaramente i big data dai semplici dati;
2. C'è mancanza di una convergenza, anche minima, sui fattori essenziali che caratterizzano il concetto.

A supporto di questa tesi, una ricerca (Favaretto, M. et al., 2020) ha riscontrato una generale incertezza o difficoltà tra gli studiosi riguardo l'uso del termine che potrebbe anche derivare dalla tendenza di riconoscere i Big Data come fenomeno culturale in continua trasformazione ed evoluzione.

In seguito all'argomentazione presentata in questo paragrafo è evidente il fatto che si deve al più presto esplicitare in maniera oggettiva, uniforme e soprattutto chiara il concetto, per evitare eventuali discordanze nel modo in cui ognuno interpreta e percepisce il fenomeno.

2.2 Trattamento dei dati personali

Ogni volta che un individuo si avvale di un motore di ricerca per i più disparati scopi, lascia una serie di tracce digitali che i proprietari e i gestori del browser raccolgono, archiviano e analizzano. Questi ultimi avranno a disposizione informazioni sugli argomenti che gli utenti trovano più interessanti, sui siti più utilizzati, su quando e come li usano e potranno elaborarli per offrire servizi e suggerimenti personalizzati, rendendo i contenuti più appetibili per i fruitori e, di conseguenza, aumentando i loro ricavi sia pubblicitari sia per un uso più frequente del loro servizio.

Da qui, un altro impiego dei dati per i gestori dei motori di ricerca che tutti noi generiamo è la vendita agli inserzionisti, che per una delle due parti genera monetizzazione e quindi entrate cospicue, per l'altra invece è molto vantaggiosa perché fonte di analisi finalizzata a capire i comportamenti dei consumatori. È da specificare che il valore delle informazioni aumenta con l'aumentare dei dati personali inclusi nel dataset.

Con l'espressione "dato personale" si indica ogni elemento che rende identificabile indirettamente o identifica direttamente una persona fisica (Faggioli, G., 2018). È possibile però dividere in quattro categorie i dati personali: i *provided data*, volontariamente e consapevolmente forniti dagli utenti; gli *observed data* la cui raccolta avviene in automatico attraverso strumenti come i cookie³⁶; i

³⁶ I cookie sono informazioni inserite nel browser quando un utente naviga su un sito web o accede ad un social network tramite vari dispositivi (pc, smartphone o tablet). Data la varietà di informazioni contenute in essi, i cookie hanno una triplice funzione. I cookie tecnici vengono impiegati per memorizzare dati specifici sugli utenti al fine di accedere ad una pagina web, infatti eseguono autenticazioni informatiche e monitorano le sessioni. Quelli di tipo *analytics* sono usati dai siti web per sviluppare statistiche sull'andamento del servizio e sul suo utilizzo in generale, si raccolgono infatti informazioni sul numero di utenti e come essi visitano il sito, sempre in forma aggregata. Gli ultimi cookie sono di profilazione e, come si può capire dal nome, hanno la finalità di profilare gli utenti mentre navigano in rete, capire i loro movimenti e le abitudini di consumo o di consultazione dei siti, principalmente per personalizzare i tipi di pubblicità da inviare.

derived data che vengono prodotti da terze parti tramite le due precedenti tipologie; gli *inferred data*, ricavati attraverso la statistica inferenziale (correlazioni/relazioni causali tra dataset) ed usati per profilare e classificare le persone (Faggioli, G., 2018).

L'elaborazione di questi dati, pur essendo molto utile sia per le aziende che per i consumatori, costituisce fonte di preoccupazione perché c'è la possibilità di risalire ai singoli individui e violare alcuni diritti degli utenti. Infatti nei dataset si trovano molto spesso informazioni quali nome, data di nascita e indirizzo IP che sono chiaramente ricollegabili a persone fisiche. Spesso le modalità che le aziende usano per la registrazione di tali informazioni non sono chiare e raramente le persone sono consapevoli dei dati sensibili che diffondono operando quotidianamente. Se qualche azienda dovesse avere a disposizione uno strumento per collegare tutti i dati relativi ad un individuo (informazioni legate gli acquisti online, opinioni e preferenze politiche espresse sui social network, informazioni sullo stato di salute, metodi di pagamento online...) ma provenienti da fonti diverse, si avrebbe una sua profilazione completa con ripercussioni negative sull'etica e la legalità della BDA.

A questo punto nasce un trade-off per gli utenti: da una parte riceveranno servizi personalizzati, su misura per i loro interessi, dall'altra dovranno rinunciare a gran parte della loro privacy.

Come le aziende intendono far fronte a questo problema? Ci sono dei regolamenti che disciplinano l'uso dei big data? Quali aziende sono autorizzate ad acquistare le informazioni personali e fino a che punto possono farlo? Come si trasforma il concetto di privacy?

Queste criticità sono tra le più importanti e discusse dell'intera letteratura sui big data, si ha notizia continuamente di scandali che coinvolgono il trattamento di dati personali da parte di aziende probabilmente non autorizzate o che hanno avuto accesso alle informazioni sensibili in maniera impropria.

I principali dubbi sorti tra big data e privacy sono la scarsa trasparenza nel trattamento dei dati e l'insufficienza di una semplice anonimizzazione delle informazioni (Faggioli, G., 2018).

Per quanto riguarda il primo aspetto, una volta che i dati inseriti dagli utenti vengono archiviati nei software dell'azienda che li ha raccolti, i big data vengono persi di vista da coloro che li hanno immessi nel sistema. Questi potranno essere utilizzati dalle compagnie in qualsiasi modo purché rientri nel range di azioni elencate una volta accettata l'informativa dagli utenti, costituendo un alto rischio di uso improprio. Le persone devono essere a conoscenza dei tipi di dati raccolti da ogni dispositivo o sito web, di come questi si usano e delle informazioni condivise con le terze parti, sia direttamente che indirettamente (completa *disclosure*). Delicati sono i principi normativi a livello internazionale per rendere conformi ai regolamenti l'informativa e il modulo di raccolta dei consensi. La struttura non deve essere ambigua o generica nella descrizione delle finalità nel trattamento dei dati ma non deve essere nemmeno troppo intricata per evitare difficoltà di comprensione nel modo in cui le informazioni sono integrate e processate. Il non rispetto di queste regole provoca l'automatica nullità del consenso.

In secondo luogo, nonostante i dati raccolti possano essere formalmente anonimi, quando si combinano più banche dati il risultato può essere un'identificazione di un profilo anche attraverso informazioni singolarmente anonime. Ciò accade perché gli algoritmi che analizzano contemporaneamente più dataset trovano connessioni reciproche e generano talvolta nuovi dati personali. In molti casi quindi la sola anonimizzazione non è sufficiente per garantire una certa privacy agli utenti.

Ulteriori criticità di cui si deve tener conto in materia di trattamento dei dati personali sono il controllo e la notifica per gli utenti (Isaak J., Hanna, M.J., 2018).

Fino ad ora il controllo che gli utenti possono esercitare sui dati inseriti nelle piattaforme è pressoché assente. Ad ognuno di noi dovrebbe essere permesso di cancellare le informazioni di identificazione personale dai dispositivi, dai siti e dai servizi cloud. In più, si devono rispettare appieno le opzioni degli utenti di non tracciare i movimenti e di non conservare i dati critici tra una sessione e l'altra; è pertanto necessario che gli interessati esplicitino l'autorizzazione nella conservazione delle varie componenti dei dati nelle informative che presenta l'azienda. Il consenso per raccogliere i dati sui singoli individui online non deve essere poi esteso in automatico per la raccolta di informazioni sui contatti a questi collegati. Ultimo ma non meno importante, se c'è un gruppo di utenti che ha bisogno di particolari tutele e attenzioni è quello dei minori: si fa poca differenza nell'individuazione di questa categoria ma è di fondamentale rilevanza proteggerli nel diffondere o analizzare i loro dati sensibili, almeno fino al raggiungimento di un'età di consenso obbligatoria legalmente.

Analizzando infine l'aspetto della notifica bisogna menzionare il dovere da parte delle organizzazioni che raccolgono memorizzano i dati di comunicare tempestivamente e chiaramente agli utenti qualsiasi perdita o abuso nelle informazioni private. Inoltre, gli utenti hanno il diritto di venire a conoscenza degli enti responsabili della inosservanza della loro privacy dal momento in cui l'azienda oggetto di violazione ne viene a conoscenza.

2.2.1 Cambridge Analytica

Lo scandalo che ha coinvolto il social network Facebook e la società di consulenza Cambridge Analytica è un chiaro esempio di come i precedenti regolamenti a protezione della privacy fossero inefficaci per via delle novità che sono state introdotte con l'uso dei big data. Da qui è stato posto sotto i riflettori il dibattito che si interessa dell'impatto sociale della tecnologia e dei conseguenti e potenziali rischi per il benessere collettivo. Risulta quindi innegabile che la successione di fatti e le indagini che hanno riguardato e tuttora riguardano Cambridge Analytica hanno fortemente messo in discussione le leggi sulla privacy negli Stati Uniti ma anche nell'Unione Europea e hanno evidenziato come queste fossero gravemente carenti. In questo contesto si esaminerà come Cambridge Analytica è stata in grado di trarre grandi vantaggi dalla sua alleanza con Facebook per accedere alle informazioni private degli utenti, aiutando le istituzioni a definire i principi per una politica sulla privacy dei dati migliore e più approfondita.

Nei primi mesi del 2018 si scoprì che Facebook, illimitatamente e senza autorizzazione, concesse l'accesso alla società inglese Cambridge Analytica alle PII³⁷ (*Personally Identifiable Information*) di oltre 87 milioni di iscritti in tutto il mondo, ignari per anni della pratica. Attraverso questo caso si intuisce chiaramente come le istituzioni sia americane sia internazionali non siano state capaci di prevedere e fronteggiare l'impatto della tecnologia, e in particolare del trattamento dei dati raccolti digitalmente, sulla struttura della società e soprattutto sui diritti e doveri del cittadino (Isaak J., Hanna, M.J., 2018).

³⁷ Informazioni di identificazione personale - per essere considerate tali, le informazioni sono direttamente associate a una persona specifica. I dati non associati al rispettivo proprietario o forniti anonimamente prima che si raccolgano non sono considerate come PII.

Facebook è stato oggetto fino al 2019 di un'indagine approfondita da parte della Federal Trade Commission (FTC)³⁸, l'accusa era la violazione di un accordo tra i due enti del 2011³⁹.

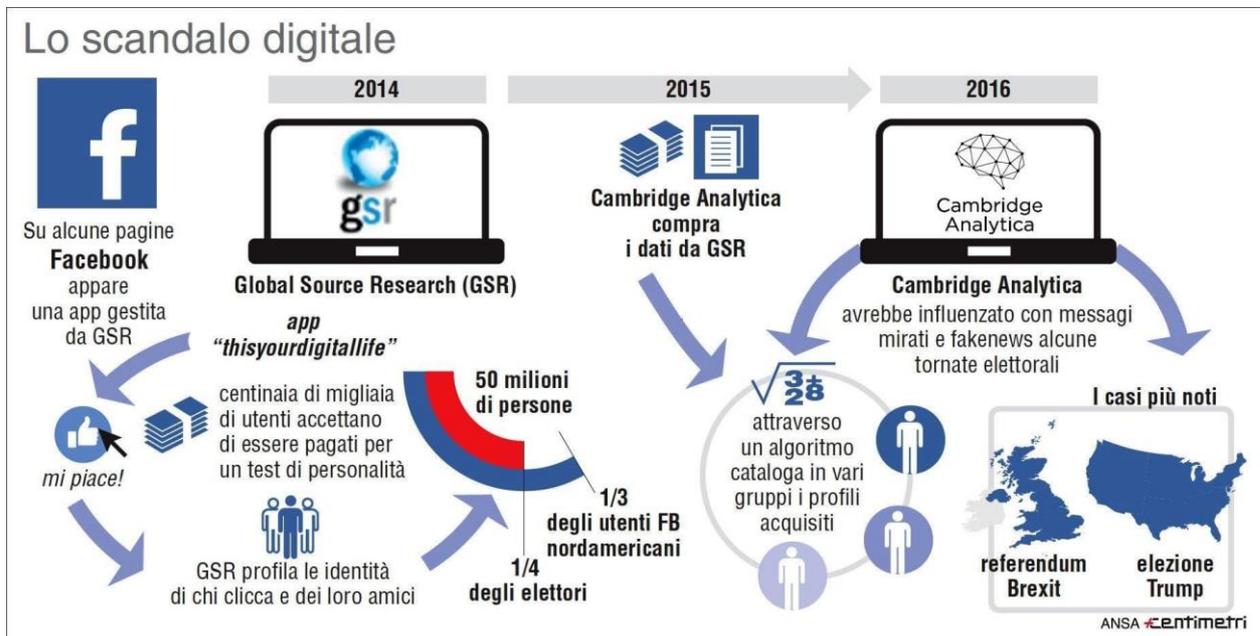


Figura 5: Rappresentazione schematica delle vicende legate a Cambridge Analytic

Il centro di psicomatria dell'università di Cambridge analizzato nel 2013 le informazioni ricavate da un test della personalità promosso su Facebook ed effettuato da alcuni volontari. Il test "OCEAN" (Openness - apertura, Conscientiousness - coscienza, Extraversion - estroversione, Agreeableness - piacevolezza, and Neuroticism - nevrosi) aveva l'obiettivo di valutare il profilo psicologico di coloro che ne erano stati sottoposti e relazionarlo ai dati delle attività su Facebook (per esempio like e condivisioni).

La ricerca ha portato allo studio di 350.000 partecipanti negli USA decretando una forte correlazione tra il profilo a 5 fattori di personalità e le azioni svolte sul social network. In poche parole, i ricercatori hanno scoperto che si potrebbe profilare accuratamente la personalità di un individuo anche solo analizzando queste metriche, tralasciando quindi uno strumento formale e più diretto quale il test OCEAN (Isaak J., Hanna, M.J., 2018). In questa fase iniziale, però, non sembra che Facebook abbia esposto i suoi utenti né tantomeno i loro contatti a qualsiasi violazione della privacy. Infatti, l'università si è rifiutata di condividere i dati raccolti e le relative modalità di analisi con quella che sarebbe poi diventata Cambridge Analytica.

Una volta resa nota la scoperta, altre aziende hanno voluto approfondire e perfezionare questo studio. La Global Science Research (GSR) ha avviato, in collaborazione con

³⁸ La Federal Trade Commission è un'agenzia indipendente degli Stati Uniti e fu istituita nel 1914. L'ente ha l'obiettivo di garantire l'alta competitività dei mercati tramite le leggi antitrust, mirando anche alla protezione dei consumatori. La protezione dei consumatori è intesa dall'FTC come tutela da pratiche commerciali predatorie o fuorvianti. Attraverso il rispetto delle leggi antitrust, l'agenzia governativa previene ed elimina le pratiche commerciali anticoncorrenziali, incluso anche il monopolio coercitivo.

³⁹ Federal Trade Commission. (29 novembre 2011). Facebook Settles FTC Charges That It Deceived Consumers By Failing To Keep Privacy Promises. Disponibile su <https://www.ftc.gov/news-events/press-releases/2011/11/facebook-settles-ftc-charges-it-deceived-consumers-failing-keep>

Cambridge Analytica, un progetto volto a identificare le variabili richieste per elaborare i profili OCEAN mediante un test di personalità diffuso in due piattaforme: Amazon Mechanical Turk, un sito web di *crowdsourcing*⁴⁰ e Qualtrics, che svolge sondaggi. Il questionario, una volta che gli utenti hanno accettato la richiesta, ha concesso a GSR di accedere al loro profilo Facebook. Si è innescata così una reazione a catena poiché questo accesso è stato impropriamente traslato alle informazioni degli amici degli utenti per via di un'API⁴¹ aperta di Facebook (che terminerà poi nel 2015). Cambridge Analytica è riuscita così ad accedere a queste informazioni personali che Facebook deteneva. Bisogna precisare inoltre che entrare in possesso dei dati specifici degli utenti non era assolutamente funzionale nel raggiungimento dello scopo primario, ovvero ricavare la metodologia corretta per performare le analisi comportamentali partendo principalmente dai dati sui social media.

Un'altra importante precisazione va fatta dal momento in cui non avere avuto un profilo su Facebook non ha implicato il fatto di non essere stato coinvolto nella violazione delle informazioni: la serie di dati raccolti non si è limitata solo al social network ma si è estesa ad altri aspetti di preferenza personale. Tutti i siti delle organizzazioni che presentano il logo di Facebook vi sono collegati, di conseguenza la società di consulenza inglese ha tracciato le azioni non solo degli iscritti ma anche degli utenti che hanno semplicemente navigato in un particolare sito internet. Un esempio diretto è l'utilizzo dei *web beacon*⁴² che sono molto spesso connessi ai cookie, permettendo una veloce vendita a terzi degli accessi degli utenti perché applicati su tutti i siti web.

Cambridge Analytica, una volta ottenute le numerose informazioni, ha avuto l'idea di unirle con nuovi dati generati da altre piattaforme sugli acquisti online, sui risultati di voto, sui browser aperti per creare "oltre 5.000 punti dati su 230 milioni di adulti negli Stati Uniti" (Isaak J., Hanna, M.J., 2018). Combinando l'analisi OCEAN con tutti i dati raccolti, sia pubblici sia privati, la società inglese ha elaborato una strategia di *micro-targeting*⁴³, secondo la quale ogni singolo consumatore o elettore ha ricevuto messaggi o offerte personalizzate, in grado di influenzare i comportamenti nelle decisioni di acquisto o anche nelle scelte politiche al livello di voti. Unendo le notizie reali ai contenuti non vincolati di internet e mostrati, gli individui, navigando su internet, si sono confrontati con messaggi di rinforzo senza nemmeno realizzare che le informazioni a loro mostrate fossero uniche nel loro genere perché fortemente personalizzate e volte ad influenzarli nelle scelte. Ancora più grave è il fatto che non abbiano ricevuto alcun avvertimento per far notare agli utenti che si trattasse di chiari messaggi di campagne politiche. È esattamente ciò che si sospetta sia

⁴⁰ Crowdsourcing è il termine utilizzato per indicare lo sviluppo, la progettazione o la realizzazione da parte della collettività di un progetto. Le persone sono in genere esterne all'azienda o alla persona che lo idea e gli utenti aderiscono su base volontaria o per mezzo di un invito. Economicamente si può definire un modello che esternalizza parte delle attività di un'impresa.

⁴¹ Acronimo di *Application Programming Interface* (interfaccia di programmazione delle applicazioni), le API hanno lo scopo di realizzare e integrare software applicativi grazie a set di definizioni e protocolli. L'impiego di queste convenzioni facilita l'interazione tra diversi programmi che altrimenti risulterebbero più distanti tra loro e molto meno funzionali. Consentono infatti ai software o servizi di comunicare con altri software o servizi tramite la struttura del loro linguaggio di programmazione. Le API, in caso di gestione degli strumenti esistenti oppure creazione di nuove tecnologie, agevolano la progettazione insieme all'utilizzo e all'amministrazione e, oltre ad essere flessibili, assicurano un certo grado di innovazione.

⁴² Un web beacon è una piccola immagine che rileva l'attività web di un individuo e di un gruppo (pagine viste, numero di utenti che hanno visualizzato...). Si trova spesso inserito nelle e-mail o sui siti web perché è progettato per tracciare le informazioni e come queste si trasformano nel tempo.

⁴³ Il *micro-targeting* è un meccanismo particolarmente avanzato, ma anche potenzialmente invasivo, che viene molto utilizzato nella funzione marketing, tramite il quale un'azienda, e nel caso in esame un'azienda di consulenza, è in grado di individuare quegli elettori che, per un qualsiasi motivo, sono più facili da persuadere.

successo durante le elezioni presidenziali del 2016 negli Stati Uniti; sono stati associati il profilo di personalità a cinque fattori con moltissimi messaggi mirati nel “Project Alamo”, di cui il presidente Trump ha usufruito per la sua campagna elettorale. Ma che legame c’era tra Cambridge Analytica e la campagna elettorale del presidente Trump? La risposta è semplice: la società di consulenza ha contribuito a strutturare la campagna del futuro presidente degli Stati Uniti, utilizzando le informazioni provenienti dalla violazione per fornire indicazioni utili sulle strategie che hanno poi portato al successo. L’allora CEO dell’azienda inglese Alexander Nix ha riferito come l’obiettivo principale fosse individuare gli elettori potenzialmente invogliati nel votare il loro cliente oppure allontanare gli stessi dal votare per l’avversario. Per raggiungere questi scopi sono stati utilizzati messaggi creati appositamente per la campagna di Trump e messaggi che hanno sfruttato notizie disponibili sul web; queste ultime si pensa che contenessero anche comunicazioni finanziate da campagne di origine russa per influenzare le elezioni statunitensi.

L’implementazione dei processi adottati e relativi alla Big Data Analytics (*micro-targeting*) è stata abbastanza facile perché in questo caso Cambridge Analytica aveva prontamente a disposizione una vasta gamma di informazioni personali, inoltre i costi sostenuti per applicare certi metodi sono stati abbastanza bassi. Puntare alla messaggistica mirata per esercitare influenze sul comportamento degli elettori significa non rispettare, o comunque aggirare i regolamenti già esistenti sul consenso informato, sulla divulgazione e probabilmente sull’intervento di un paese straniero. I voti così guadagnati o fatti perdere hanno avuto un certo effetto sui risultati delle elezioni. La strategia pubblicitaria ha da una parte massimizzato l’impatto sugli utenti e dall’altra ha derivato pattern non osservabili su ampiezza macro, questo è successo grazie alle pronte reazioni avute (trovare i clickbait⁴⁴ funzionanti), monitorando in tempo reale gli esiti dei messaggi sul target di elettori. Sicuramente non è l’unica spiegazione da dare per i risultati delle elezioni presidenziali nel 2016, ma il contributo che ha apportato è stato utile se non critico.

Un altro avvenimento che ha coinvolto in prima persona la società Cambridge Analytica è stato il caso della Brexit in Gran Bretagna. In questa occasione ad agire in maniera illegale è stata una società collegata di Cambridge Analytica (AggregateIq), che ha svolto il lavoro di propaganda a favore dell’uscita della Gran Bretagna dall’Unione Europea. L’illecito, oltre alla solita violazione della privacy per alcuni profili su Facebook, sta nel fatto che la campagna pro-Brexit ha superato il budget massimo di donazioni usate poi per finanziare le ricerche e gli strumenti utilizzati da AggregateIq. La strategia pubblicitaria utilizzata è la stessa che la società di consulenza ha implementato negli USA. Gli sviluppi a cui i messaggi mirati hanno portato sono stati forse più incisivi sulle decisioni degli elettori e quindi sul risultato del referendum svoltosi il 23 giugno 2016.

La società Cambridge Analytica, a seguito degli scandali che l’hanno coinvolta, ha dichiarato bancarotta il 2 maggio 2018 e ha dovuto quindi chiudere. Come conclusione delle indagini svolte su Facebook, la FTC ha deciso di multare l’azienda americana per una cifra pari a 5 miliardi dollari. Mark Zuckerberg si è poi impegnato a riformare la policy di Facebook, modificandola e migliorandola in modo tale da prevenire simili eventi in futuro.

L’analisi condotta in questa sezione suggerisce un definitivo cambiamento delle politiche aziendali e delle regolamentazioni governative per proteggere gli individui, siano questi consumatori, elettori o semplicemente utenti, e garantire la tutela dei dati personali. Le

⁴⁴ “Clickbait” è il termine usato per indicare un contenuto web la cui principale funzione è attirare il maggior numero possibile utenti per generare maggiori rendite pubblicitarie online.

nuove iniziative devono essere volte anche a informare i cittadini sulle attività che vengono svolte con i loro dati per consentire loro una maggiore partecipazione nel processo informativo.

2.2.2 General Data Protection Regulation (Regolamento Europeo n° 679/2016)

Alcuni scandali come quello di Cambridge Analytica hanno contribuito alla crescita di consapevolezza da parte delle autorità internazionali nell'emanare regolamenti aggiornati sul trattamento dei dati personali.

Un diretto esempio è la nuova normativa europea in materia di protezione dei dati che è diventata pienamente operativa a partire dal 25 maggio 2018 in tutti i 28 paesi dell'UE. Il Regolamento Generale sulla Protezione dei Dati (GDPR), era stato però pubblicato nella Gazzetta Ufficiale europea già il 4 maggio 2016. Questo, entrando a far parte del sistema giuridico e attuandosi allo stesso modo in ogni Stato membro, si è prefissato l'obiettivo di armonizzare normative diverse che erano presenti solo a livello nazionale.

Si procede quindi all'abrogazione della direttiva 95/46/CE⁴⁵ (Direttiva Madre), la più importante regolamentazione sulla privacy a livello comunitario dal 1995. Una volta stabilito dal trattato di Lisbona che la protezione dei dati personali sarebbe diventata diritto fondamentale dei cittadini, garantita quindi allo stesso modo in tutto il territorio, si ebbe la necessità di strutturare una regolamentazione funzionale al continuo progresso digitale e più rigorosa (Saetta, B., 2018). La precedente direttiva, infatti, risaliva agli anni in cui internet stava muovendo i primi passi e i big data non erano rilevanti né godevano della stessa attenzione di oggi. Si era pertanto creata una frammentazione nel modo in cui ogni stato aveva strutturato le normative in termini di privacy, aumentando i rischi per la protezione dei dati online: tali criticità avrebbero potuto alterare la concorrenza a livello europeo per via di un'impropria circolazione dei dati tra i vari Paesi. È per queste ragioni che la commissione europea ha voluto riformare la disciplina dell'UE materia di protezione dei dati. Da un'idea basata sul trattamento dei dati subordinato al consenso, e quindi a una visione proprietaria delle informazioni, si passa al concetto di controllo, in cui l'interessato deve essere a conoscenza delle modalità d'uso dei propri dati e il gestore consapevole della tutela necessaria per la società e per il singolo individuo dai possibili rischi nella loro violazione (Saetta, B., 2018).

Il GDPR considera come diritto fondamentale delle persone fisiche il diritto alla protezione dei dati, attribuendovi il concetto di autodeterminazione informativa che è un elemento cardine di una democrazia e status necessario per l'evoluzione della figura del cittadino. Il nuovo approccio del legislatore si trasforma da *formale e reattivo* in *concreto e proattivo* (Saetta, B., 2018): l'azienda che tratta i dati non si limita più solo ad avere il consenso da parte degli utenti ma viene anche responsabilizzata nel rendere conto delle operazioni che svolge con i dati personali e nell'assicurarsi che vi sia un effettivo rispetto della normativa. Il principio dell'*accountability* a livello aziendale infatti permea l'intero regolamento poiché le imprese hanno il compito di strutturare la gestione e l'organizzazione anche in base alle tipologie di trattamento dei dati e alla loro protezione (AGCM, 2019). Infine, le aziende titolari del trattamento devono sì conformare quest'ultimo sulla base dei principi

⁴⁵ Direttiva 95/46/CE del Parlamento europeo e del Consiglio, del 24 ottobre 1995, relativa alla tutela delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati. Disponibile su https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=uriserv:OJ.L_.1995.281.01.0031.01.ITA

del GDPR, ma è necessario anche che introducano misure di sicurezza, per eliminare il rischio tipico connesso alla loro attività, anticipandolo e valutandolo.

La nuova normativa prevede inoltre (Saetta, B., 2018):

- Consolidamento delle garanzie e dei diritti azionabili dall'utente al fine di controllare le informazioni date ed esercitare l'autodeterminazione informativa;
- Verificata la violazione della norma da parte di un'azienda, si provvederà ad imporre una sanzione amministrativa di importo fino al 4% del suo fatturato globale.
- Centralità della *governance* e del controllo su quanto le aziende rispettano e sono conformi alla normativa. Questo avviene grazie ad una efficace cooperazione tra le Autorità di controllo nazionali, che vengono valorizzate, e il Comitato europeo per la protezione dei dati;
- Introduzione, come da indicazioni della Corte di giustizia europea, del diritto alla cancellazione. I dati verranno rimossi su richiesta dell'interessato quando manca l'interesse pubblico alla notizia;
- Trasferimento dei dati personali sui servizi online grazie al diritto alla portabilità dei dati;
- In caso di violazione grave dei dati da parte di terzi le aziende avranno l'obbligo di notifica;

Tali riforme però non sono complete; tre sono le debolezze riscontrate nell'analisi del GDPR. In primo luogo, le nuove direttive si basano sostanzialmente sulla prima legge europea per la tutela della privacy (direttiva 95/46/CE), e anche vero che si riconoscono cambiamenti verso la giusta direzione ma i modelli di business contemporanei hanno già mostrato ripetutamente inadeguatezza della vecchia normativa.

Una seconda debolezza è la possibile re-identificazione degli utenti mediante l'uso di correlazioni nell'intelligenza artificiale, nonostante i dati siano anonimizzati.

La terza ed ultima riguarda il chiaro perseguimento degli interessi propri dell'azienda quando invece il raccoglimento dei dati dovrebbe andare a vantaggio anche dei consumatori, evidenziando un esteso controllo di tale mercato.

Sebbene attraverso l'emanazione di questo regolamento si siano fatti dei passi avanti in tutta Europa, mancano ancora alcuni aspetti da tutelare perché se si pensa all'elaborazione, analisi e ancor prima all'organizzazione dei dati, ci sono casi inaspettati in cui si dovrebbe richiamare la libertà di decisione sia dei singoli che della società. In merito a ciò il regolamento non si esprime pienamente, avendo mantenuto alcune scorie dell'impostazione giuridica della Direttiva Madre, in cui la tutela era portata al singolo nel trattamento dei dati e nella finalità d'uso degli stessi (AGCM, 2019).

2.3 Illusione o realtà?

Saranno ora presentate due aspre critiche sul concetto dei big data e sul fatto che questa potrebbe essere una mera illusione che ha portato le persone a parlare molto positivamente di un concetto che in realtà non è mai esistito, né tanto meno può essere comparato all'oro.

Stephen Few nel suo libro afferma che, anche se i big data avessero un significato chiaro, che al momento non hanno, non importerebbe perché nessuna delle affermazioni sui big data è vera. Sono sempre stati un'illusione. La popolarità talvolta alimenta altra popolarità anche quando la concretezza è completamente assente. Ad oggi poche persone si sono poste la domanda sulla

legittimità dei big data, ciò avviene non perché questa è riconosciuta ma semplicemente perché è un assunto generale. Il continuo parlare di qualcosa, in questo caso i big data, cresce fino ad essere una mera assunzione della sua esistenza senza però avere prove certe.

Quindi, se i big data sono un'illusione, perché così tante persone ne parlano come se fosse reale? La risposta a questa domanda si trova nel fatto che per Few il fenomeno è una campagna di marketing straordinariamente di successo. L'elemento che più ha contribuito al successo di questa campagna di marketing è stato porre la parola "big" davanti alla parola "data". È nella natura umana desiderare tutto ciò che viene espresso in termini di dimensioni mega: nei dati come nel cibo questa formula ha spesso esito positivo (Few, S., 2018).

Si tratta solo di un modo nuovo di guadagnare per i promotori e per i provider del servizio. Le affermazioni sul fenomeno sono state progettate al fine di distribuire grandi ricavi agli ideatori e a tutte le organizzazioni che traggono beneficio nel supportarli, inclusi i consulenti e gli analisti che lavorano nelle aziende. Non a caso queste figure non hanno mai messo in dubbio il concetto e nemmeno si interessano delle implicazioni etiche che nascerebbero nel proporre una tale illusione. Consideriamo, per esempio, la definizione di big data che fa leva sull'assunzione che possiamo e dobbiamo archiviare tutti i dati che sono prodotti nel mondo, nel caso in cui si dovessero rivelare utili in un certo momento del futuro. Se così fosse, perché le organizzazioni in possesso della quantità maggiore di dati non sono le più efficienti? A dire il vero le organizzazioni con più dati si trovano in grande difficoltà.

È infatti altamente improbabile che gli usi secondari di alcuni tipi di big data, per adesso mai applicati, creeranno abbastanza valore da bilanciare il costo di archiviazione. Tali costi non sono intesi soltanto nel modo diretto (risorse impiegate nell'acquisto di hardware), ma in essi vengono compresi anche il tempo "spreco" e lo sforzo mentale impiegato per ideare un simile sistema di archiviazione dei dati. E allora perché continuare a tenerli nonostante la loro poca utilità? La risposta è semplice: le stesse persone che promuovono i dati beneficiano a livello economico dall'archiviazione di enormi quantità di informazioni. Seguendo il ragionamento di Few è facilmente dimostrato che i big data sono soltanto un'illusione.

La critica pubblicata sul Financial Times (Hale, T., 2019) rincara poi la dose: i big data non sono il nuovo oro, né il nuovo petrolio e quindi nemmeno il nuovo legame connesso all'inflazione o il nuovo nulla. La scrittura viene prodotta in modo interpretativo, così come la generazione dei dati, e questi sono anche infinitamente divisibili, al contrario dell'oro. Si possono elencare tutti i dati possibili ed immaginabili su un individuo.

Quando parliamo di big data ci troviamo quindi di fronte a una serie di idee peculiari che si ispirano profondamente ad usi fuorvianti del termine. Tali idee presuppongono che i dati ricoprano la funzione di una di merce, e persino le confutazioni delle precedenti posizioni comportano differenze intrinseche tra i dati, che tra l'altro possono essere riutilizzati, e il petrolio, che invece non può (Hale, T., 2019). Spesso si parla di monetizzazione dei dati, affermando che è preziosa; i clienti dovrebbero essere compensati per la loro fornitura alle aziende ma ciò non avviene.

Se i dati degli utenti hanno un valore diverso da zero, è possibile che si verifichi un'iperinflazione quasi immediata, dato che tutti potremmo continuare a stampare-produrre dati su noi stessi all'infinito (in effetti, potremmo stampare dati su una singola transazione infinite volte) (Hale, T., 2019). Devono pertanto esserci vincoli alla nozione di big data. Ma, come affermava anche Few, secondo Hale ci sono valori nascosti per gli inserzionisti e i venditori in tutti i tipi di pratiche comportamentali e rituali. Non vi è alcun limite ovvio né a questi valori, né alle combinazioni imprevedibili in cui emergono. E, in effetti, l'intero punto è che emergono continuamente causando un circolo vizioso.

Capitolo 3

Ambiti di applicazione nella gestione aziendale

Quando si parla di concetti molto in voga ultimamente quali e Imprese 4.0 e Industria 4.0 si fa un implicito riferimento alle varie applicazioni che la Big Data Analytics prevede in ambito aziendale. Tali applicazioni sono state sperimentate in numerosissime funzioni aziendali, a tutti i livelli e nelle più disparate industrie. Chiaramente si nota una polarizzazione di tali fenomeni sui dati e contemporaneamente su ideali che implicano una visione d'insieme sul futuro ampia e alcune volte più precisa rispetto ai big data (Bellini, M. 2020). Da questi principi parte una volontà per le aziende nel ricercare innovazioni *data-driven*, il cui fulcro sono due concetti chiave che sintetizzano il significato della trasformazione digitale avvenuta negli ultimi tempi: conoscenza e precisione (Bellini, M. 2020). La prima parola spinge le aziende, attraverso l'impiego dei dati, ad esplorare ambiti applicativi precedentemente impensabili; la seconda consente di approfondire sia le conoscenze pregresse sia quelle più recenti e aumenta l'accuratezza delle azioni o delle decisioni sottostanti.

Malgrado i grandi cambiamenti apportati nelle strutture e nelle competenze aziendali dal lavoro dei dati, l'analisi che si effettua è spesso limitata poiché circoscritta e contingente, permettendo di raggiungere solo determinati obiettivi. Lo step successivo è sfruttare al massimo delle potenzialità la ricchezza e il valore che sono contenute nei dataset: questo si raggiunge solamente se ci sarà una visione d'insieme di tutte le fonti e dell'impatto che esse potranno avere sul miglioramento dell'azienda in toto. L'Italia ne è un esempio lampante; secondo una ricerca dell'Osservatorio Big Data & Business Analytics del Politecnico di Milano il mercato della BDA in Italia è *“dinamico e sempre più maturo, con imprese che mostrano un livello avanzato di utilizzo delle tecnologie, sperimentazioni complesse e competenze di Data Science, affiancate da altre che, pur in ritardo, si stanno attivando aumentando gli investimenti e puntando su progetti di integrazione dei dati”*. È improbabile pensare che alcune imprese non si adegueranno mai al progresso tecnologico in corso, perché competere con aziende in grado di prendere decisioni in tempo reale, grazie all'analisi delle informazioni, risulterebbe improbabile o quantomeno troppo difficile e darebbe vita ad una competizione ad armi impari. Chiaramente il divario più significativo si registra tra le PMI e le grandi imprese in quanto gli investimenti effettuati e le competenze richieste erano minori nel primo caso. I dati dimostrano come, anche per via della composizione del tessuto imprenditoriale italiano, la strada da fare è ancora lunga ma già ora si iniziano a vedere i primi significativi passi in avanti. Infatti l'investimento in progetti di Analytics per le PMI è il 62% del totale contro il 93% delle grandi imprese (PoliMi - Osservatorio Big Data & Business Analytics, 2019).

A livello internazionale, invece, la situazione varia un po' perché le altre economie sviluppate non hanno tendenzialmente una presenza di Piccole e Medie Imprese forte come in Italia. Essendo presenti numerose multinazionali e grandi imprese, l'uso dell'Advanced Analytics è più diffuso e l'avanzamento verso il modello *big data-driven* procede ad un ritmo leggermente maggiore.

È innegabile a questo punto un cambiamento più che positivo portato dai big data perché il gap tra le piccole imprese e le grandi società. Nonostante la diversità dei budget di cui dispongono, le aziende più piccole hanno ugualmente accesso a zettabyte di informazioni. Pertanto, in futuro ci sarà un livello di competizione nell'intero mercato più alto e ciò genererà maggiori vantaggi per i clienti. Il motivo principale è l'introduzione di molti software che processano i big data e sono gratuiti, come Google Analytics o Azure HDInsight (prodotta da Microsoft) e Skytree. Le PMI non dovranno necessariamente sostenere le spese della licenza per un software ma potranno direttamente investire

questo denaro nello sviluppo di conoscenze e competenze adatte ad usare in maniera corretta quegli strumenti.

Attraverso questa riflessione si dimostra come il progresso tecnologico dei dati stia cambiando e continuerà a modificare la struttura economica prima delle aziende e poi del Paese in cui queste operano.

Molti sono i dubbi e le opinioni contrastanti degli studiosi della materia, per la maggior parte riguardano la tutela della privacy, la crescente disponibilità e la più facile accessibilità ad informazioni di vario tipo. Allo stesso tempo sembra doveroso far presente che sotto il punto di vista commerciale ed economico si nota un'occasione unica fornita dalla Big Data Analytics, è per questo importante che le aziende colgano tale opportunità per operare nel modo più efficace possibile e il più velocemente possibile.

Nel corso del capitolo saranno approfondite le modalità in cui la BDA opera correntemente in alcuni ambiti aziendali, unendo ad essi alcuni esempi applicativi nella realtà per comprendere meglio l'importanza che ha assunto, sta assumendo e continuerà ad assumere questo fenomeno.

3.1 Management

“You can't manage what you don't measure” è una frase famosa, datata ma ancora attuale perché perfettamente applicabile al concetto di gestione dell'azienda attraverso le analisi dei big data.

Continuare a mantenere il controllo senza troppi riferimenti accuratamente misurabili in un mondo completamente sommerso da software, informazioni digitali, dispositivi tecnologici e rivoluzioni informatiche è altamente complesso. La razionalità limitata di noi esseri umani ci impedisce di considerare allo stesso tempo tutte le variabili necessarie per effettuare delle scelte chiave quando si tratta di gestione aziendale. È per questo che i dirigenti delle imprese hanno bisogno di strumenti che li supportino e li aiutino a percorrere le giuste strade per far crescere le rispettive società e contemporaneamente raggiungere un maggiore benessere sociale: algoritmi di Machine Learning, AI e Advanced Analytics sono solo alcuni esempi.

Il management è il secondo tra i campi che maggiormente hanno beneficiato della rivoluzione big data e nel futuro tale fenomeno è destinato a svilupparsi ancora di più. Mentre nel passato la cultura decisionale dei top manager era fondamentalmente basata sull'intuito e sull'esperienza, oggi le loro congetture possono essere confermate facendo riferimento alla cultura della Big Data Analytics. Nei processi gestionali viene quindi introdotto una sorta di metodo scientifico, che fino ad ora in questo ambito era stato raramente utilizzato per mancanza di componenti essenziali. In primo luogo i manager, a tutti i livelli, iniziano a formulare ipotesi sulla base della loro intuizione riguardo determinati argomenti; la seconda fase consiste nello sperimentare le ipotesi precedentemente individuate; il terzo step implica un'analisi quantitativa e non più soltanto qualitativa dei risultati; la quarta ed ultima fase prevede l'eventuale conferma delle ipotesi e una successiva scelta della decisione finale.

La tendenza che si sta sviluppando ultimamente, per poi essere sfruttata appieno nel futuro, è la creazione dei modelli di business *big data-driven (DDBM)*: l'idea di partenza in questi business model non è più basarsi prettamente sulle intuizioni manageriali ma fare riferimento all'analisi dei big data per costruire da zero o implementare il vantaggio competitivo che tanto cerca ogni azienda per distaccarsi dai competitor (Brownlow, J. et al., 2015).

Il processo è iniziato pochissimi anni fa e non sarà di breve durata poiché bisogna lasciare il tempo necessario alle aziende di sviluppare le competenze e le tecnologie adatte a supportare un progresso di tale importanza.

3.1.1 Processo di *decision-making*

Il processo di *decision-making*, alla base delle logiche d'azione del management, è stato rivoluzionato dall'avvento della Big Data Analytics. Innanzitutto, sembra doveroso specificare che una società generalmente prende tre tipi di decisione, presentati nel *decision-making model* (LathamDrive, n.d.):

Decisioni strategiche: Sono quelle prese dal board of directors di una impresa; consistono in piani o indicazioni che hanno un impatto materiale o nel lungo periodo, includono informazioni sulla strategia corporate di un'azienda. Tra queste possiamo trovare l'approvazione della distribuzione degli utili, la vendita di una controllata, l'adozione di piani a supporto gli impiegati, l'approvazione del budget annuale...

Decisioni tattiche: Vengono prese dall'alta dirigenza, tra cui ci sono CEO, CFO e COO e articolano in maniera più dettagliata le indicazioni della strategia generale, di solito impattando sul medio termine dell'impresa; questi dirigenti dovranno poi rispondere delle decisioni che hanno preso al board of directors. tra le decisioni tattiche si includono la firma dei contratti di non divulgazione, ampi acquisti secondo il budget preventivamente approvato, struttura e grandezza della forza lavoro, strategie di vendita e di marketing...

Decisioni operazionali: Coinvolgono manager di medio e basso livello oltre che supervisor, includono decisioni che solitamente si prendono giorno dopo giorno e per questo hanno un risvolto a breve termine sulla compagnia. Il tipo di decisioni è molto specifico ed è legato ad un basso livello di incertezza e sono l'applicazione delle strategie aziendali a medio termine. alcuni esempi sono assegnare lavori specifici ai dipendenti, stabilire per la formazione su particolari compiti, organizzazione dei turni dei dipendenti, acquisto di materiale d'ufficio.

In un ambiente volatile in cui sulle aziende ricadono aspettative altissime nel soddisfacimento dei bisogni del cliente, nella gestione dei costi, nell'affrontare al meglio la competizione, il processo di *decision-making* (in termini di decisioni operazionali, tattiche e strategiche) deve essere molto rapido e basato sui dati disponibili in quel preciso momento (Jeble. et al, 2018). Le decisioni alla base di moltissimi processi di gestione aziendale verranno trattati nei prossimi paragrafi mentre questo è più concentrato sulle implicazioni che la BDA porta nel "puro" processo di *decision-making*.

Ci sono vari studi che mirano a capire meglio le emozioni, le opinioni e le attitudini dei consumatori sui siti di social networking (SNSs). I manager aziendali per prendere decisioni tattiche e strategiche (strategia di marketing e vendite di un bene) vogliono sapere, oltre ai pareri dei clienti sui loro prodotti, ciò che pensano i clienti della concorrenza sui relativi prodotti. Il data mining sui social media è finalizzata a ricavare analisi comparative sulle informazioni di vendita di un'altra azienda e dei competitor ma anche sui pareri dei consumatori (intelligenza competitiva) (Jeble. et al, 2018).

Con l'impiego delle informazioni sulle curve di domanda e offerta nella BDA molte compagnie visionaria hanno implementato capacità di prendere una decisione in tempo

reale. Queste analisi che effettuano includono competenze che non possono essere trovate da nessun'altra parte nel business model tradizionale. Le strategie di gestione della domanda sono frutto anche di analisi storiche su dati relativi all'andamento della curva di domanda.

Da qui si può intendere che tra i maggiori obiettivi della data science c'è il miglioramento dell'abilità di manager di prendere più accurate decisioni aziendali (Jeble. et al, 2018).

3.1.2 Supply Chain Management

Il Supply Chain Management (SCM) ha largamente applicato moltissime tecnologie che sono direttamente connesse alla BDA come sensori, IoT, GPS, codici a barre, RFID⁴⁶ per connettere, coordinare e anche far integrare i vari elementi che compongono la catena di approvvigionamento di un'azienda. Pare quindi evidente come la supply chain sia stata completamente rivoluzionata dalla BDA e grazie a questo abbia compiuto molti progressi. Un altro beneficio che l'analisi dei big data comporta è la capacità di adattarsi prontamente nel contesto di un mercato che cambia velocemente ed è in continua evoluzione. Sebbene sia empiricamente dimostrato che ci sono molteplici vantaggi nell'adottare questo tipo di approccio, una recente ricerca ha evidenziato che solo il 17% delle aziende ha introdotto con successo la BDA in una o più componenti della supply chain (Nguyen, T., et al., 2018). Ciò avviene poiché da una parte alcune aziende si trovano in difficoltà nel far leva sui big data per ragioni di costo e dall'altra le restanti conseguono risultati non soddisfacenti per via di strategie frammentate e strutturate in maniera confusionaria.

Secondo Nguyen et al. (2018) ci sono una serie di tecniche che tendono ad essere usate di più quando le società combinano il supply chain management con la Big Data Analytics. Gli strumenti di modellizzazione più usati sono i network neurali nelle funzioni di *forecasting*, gli SVM⁴⁷ quando si parla di classificazione e l'euristica⁴⁸ in caso di ottimizzazione dei processi. Tra le tecniche di simulazione e *clustering* troviamo invece gli algoritmi k-means, spesso presente nelle fasi iniziali per raggruppare i dati secondo logiche più omogenee. Infine, nei modelli predittivi le procedure più utilizzate sono gli alberi decisionali e tecniche avanzate di data mining.

⁴⁶ È l'acronimo di Radio Frequency Identification; questa tecnologia di identificazione automatica si ottiene grazie alla propagazione nell'aria delle onde elettro-magnetiche e consente di rilevare automaticamente (hand free) e a distanza oggetti, animali o persone sia in movimento che statici.

⁴⁷ Support vector machines (SVM) - sono un metodo molto diffuso di ML per riconoscere alcuni pattern, vengono usati anche per le analisi di regressione.

⁴⁸ Tecnica di progettazione degli algoritmi per risolvere problemi in maniera più rapida rispetto ai metodi classici o per trovare una soluzione, se pur approssimava, quando gli stessi metodi tradizionali non riescono a farlo.

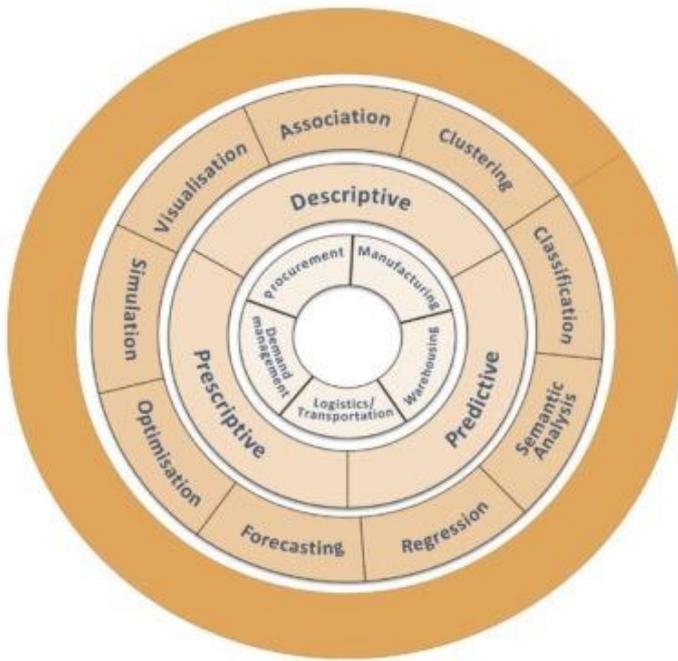


Figura 6: struttura della classificazione

Uno studio (Nguyen, T., et al., 2018) ha dimostrato come alcune aree tra quelle che vanno a formare l'intera supply chain hanno mediamente incorporato i big data meglio di altre, sviluppando più tecniche ed avendo risultati migliori. Come è deducibile, esistono anche applicazioni della BDA che combinano le diverse fasi della SC come fossero reti interconnesse a più livelli. Queste vengono integrate tra di loro ma sono le più rare da strutturare, perciò le aziende ne fanno uso meno spesso. I progressi che si possono effettuare, sperimentando tale approccio, sono però notevolmente superiori, vista la netta riduzione dei costi sopportati dalle imprese. In tale ambito vengono progettati in larga parte algoritmi composti da modelli di analisi miste (*predictive, prescriptive e descriptive*).

Verranno ora presentati in ordine decrescente di utilizzo tutte le sezioni del SCM con le relative applicazioni più riuscite.

Fra tutte le aree della SC la parte della **logistica** e dei **trasporti** è quella in cui la Big Data Analytics è usata di più come supporto nel processo di *decision-making*. Si punta quindi a sviluppare algoritmi e efficienti soprattutto di gestione delle scorte per la logistica in transito e sistemi intelligenti di trasporto (ITS). Nel primo caso si studia come monitorare e controllare lo stato del prodotto usando alcuni sensori, nel secondo invece le applicazioni si concentrano sul monitoraggio delle operazioni sul traffico in tempo reale e su una gestione proattiva della sicurezza. Vengono generati giornalmente così tanti dati (dimensioni, destinazione, origine, peso...) che si ha un grande potenziale per raggiungere la massima efficienza operativa e allo stesso tempo non trascurare l'esperienza del cliente. Per questo la *prescriptive analytics* è la più usata nel raggiungere i precedenti obiettivi: all'interno di essa meccanismi di ottimizzazione dei percorsi e pianificazione della logistica contribuiscono ad una volontà generale di ottimizzare che è fondamentale nel SCM. Inoltre, spesso nascono alleanze strategiche tra fornitori del settore logistico, imprese stesse che si occupano di logistica e clienti con lo scopo di produrre enormi dataset ed archivarli utilizzando strumenti come i *Warehouse Management Solutions (WMS)* e i *Transport Management System (TMS)*. Le informazioni sono poi condivise e permettono l'accesso a tutti i membri per svolgere analisi più ampie e complete.

Un'altra componente implementata è la **produzione**, strettamente legata al controllo e alla pianificazione della catena produttiva, oltre che alla ricerca e sviluppo (R&S). Anche qui la metodologia preferita è un'analisi prescrittiva, grazie all'adozione di sistemi come il *Cyber Physical System (CPS)*, adottato pure nell'area trasporti e logistica. Il CBS, Elemento chiave dell'industria 4.0, è composto da una serie di differenti tecnologie che generano un sistema

intelligente e capace di interagire con sistemi diversi e fisicamente distanti. Non a caso in tale sistema c'è una riproduzione degli oggetti fisici nell'ambito digitale e in più sono aggiunti elementi con capacità di memorizzazione, calcolo e comunicazione.

Il modello di classificazione dei dataset è molto diffuso nella produzione perché permette di effettuare previsioni con un particolare livello di accuratezza. Le congetture vertono in particolar modo sulla manutenzione e sull'eventuale diagnosi dei problemi delle apparecchiature e sulla sezione di pianificazione e controllo, con potenziale supporto della produzione. Questa capacità permette ai produttori di identificare colli di bottiglia e scovare le componenti dei processi che hanno una performance non soddisfacente, in modo tale da migliorare la produzione e sfruttare al massimo le risorse a disposizione (Darvazeh, S. et al., 2020).

Nella sezione della **gestione della domanda**, invece, modellare la curva di domanda futura e rilevare la struttura della curva di domanda corrente sono tra le più utilizzate iniziative della BDA. I due aspetti hanno un orientamento rispettivamente a lungo e a breve termine. Ma comunque danno l'opportunità alla azienda di specializzarsi su prodotti personalizzati (Darvazeh, S. et al., 2020). Meno attenzione invece è posta sulla previsione della domanda futura, che comunque fornisce indicazioni sui possibili gusti futuri dei consumatori. In questo ultimo caso è logico affermare che la tipologia di analisi più utilizzata è di tipo predittivo, determinare l'andamento della domanda negli anni a venire si può fare tramite algoritmi di *forecasting*. La *semantic analysis*, che fa parte del NLP (Natural Language Processing), viene adoperata per il profilo di rilevazione della domanda e consiste nel verificare la significatività dei dati sotto forma di testo.

Già spostandosi sui SaaS (*Software-as-a-Service*⁴⁹) per immagazzinare le informazioni, la natura cloud dei big data fa ridurre notevolmente alle imprese i costi degli hardware e della loro manutenzione (Darvazeh, S. et al., 2020). Lo **stoccaggio** (*warehousing*) poi vede come applicazioni più riuscite la gestione dell'inventario e il miglioramento dell'assegnazione dello spazio di archiviazione, mentre l'uso della BDA a supporto del processo di raccolta ordini è meno diffuso. La tecnica più usata è un'analisi di tipo prescrittivo ma anche quella descrittiva è inserita in termini di algoritmi di associazione, finalizzata a scoprire relazioni ricorrenti tra le varie componenti dell'inventario per minimizzare i relativi costi di stoccaggio.

In ultimo, il processo di **procurement** è la parte che fino ad ora trova meno applicazione nell'analisi ma comunque indirizza la sua attenzione verso la selezione dei fornitori, il miglioramento del reperimento e la gestione del rischio delle risorse. Anche qui gli analisti prediligono la *predictive analytics*. Di nuovo lo strumento della classificazione torna utile per essere implementato: ARM e alberi decisionali in combinazione sono una risorsa preziosa in tale sezione. Il meccanismo di *Association Rule Mining* aiuta a trovare in automatico correlazioni nascoste nei dataset ma abbastanza ricorrenti. Per esempio, stabilito che nei dataset sono presenti dati sia interni sia esterni, identificare quali sono i maggiori rischi nella supply chain aiuta a gestire il rischio appena questo si presenta: ciò significa avere già pronto un piano d'azione in caso di emergenza (Darvazeh, S. et al., 2020).

⁴⁹ SaaS è un recente modello innovativo progettato per distribuire un software. Questo consiste nell'ospitare un qualsiasi software nel cloud, utilizzandolo come fosse un software installato direttamente sul dispositivo al livello locale. Le componenti essenziali però sono un abbonamento e avere una connessione internet. Nonostante non ce ne accorgiamo facciamo uso dei SaaS molto più spesso di quanto si pensi: Google Docs, Microsoft 365 e Google Sheets sono tra gli esempi più famosi di Software as a Service.

Data la struttura ad “ecosistema stratificato” della Supply Chain, i vantaggi del cloud computing quali scalabilità, interruzioni ridotte ed economicità conferiscono un prezioso valore aggiunto nel sistema che consiste nella coesione. Se prendiamo in esempio i dati meteorologici oppure quelli registrati nel GPS, come vedremo nel caso applicativo, si nota un’ottimizzazione nei percorsi per la consegna della merce. In particolare, si tratta di una combinazione tra strumenti dell’IoT e AI che riducono le tempistiche di inattività degli asset e determinati costi operativi (McDaniel, S., 2019).

Se le aziende non trasferissero i dati nel cloud, si ritroverebbero ad operare tramite silos di dati. Ma questa scelta sarebbe controproducente per competere nel mercato poiché le aziende non usufruirebbero della natura *plug-and-play* degli elementi contenuti nel cloud, che facilitano di molto le operazioni di gestione aziendale della SC (McDaniel, S., 2019). Tali meccanismi possono verificarsi attraverso le funzionalità del Machine Learning e ed alcune componenti cloud personalizzate. Non a caso, il valore di mercato delle soluzioni SCM basate su cloud dovrebbe superare gli 11 miliardi di dollari entro il 2023 (McDaniel, S., 2019).

Il caso di Ups

Miliardi di informazioni di vario genere vengono raccolte ogni settimana lungo tutto il network di UPS.

United Parcel Service, fondata nel 1907 a Seattle e meglio conosciuta tramite il suo acronimo UPS, è la compagnia leader mondiale nel trasporto specializzato e nei servizi di logistica, operando in più di 220 Paesi.

La società è correntemente in possesso di un vantaggio competitivo importante che permette di far crescere ulteriormente il loro business grazie all’uso di tecnologie avanzate come la BDA, grazie anche all’Advanced Technology Group che ha fondato (Roby, K., 2020).

Ciò è possibile perché UPS ha a disposizione risorse a livello globale e un network integrato di preziosi asset fisici, tecnologici e umani. Per far sì che il network operi efficientemente e velocemente, avendo un occhio di riguardo anche sul suo impatto ambientale, la compagnia americana ottimizza, progetta, implementa e talvolta acquista gli strumenti tecnologici necessari per continui miglioramenti nei sistemi di logistica e trasporto.

UPS utilizza i sensori e GPS che sono collegati ai suoi database per registrare le condizioni, gli spostamenti, le performance dei veicoli che trasportano la merce; registra perfino le azioni dei guidatori e la loro condotta in materia di sicurezza. Importante è anche la dotazione di apparecchi che monitorano lo stato di consegna della merce ed i servizi offerti ai propri clienti, oltre a ricercare una certa minimizzazione dei costi legati al trasporto. Solo i 100.000 *smart trailer device* producono circa 60 milioni di messaggi alla settimana, informazioni preziose per ridurre i tempi delle operazioni e i costi ad esse legati (Roby, K., 2020). Combinando queste ultime ed il sistema Legacy aziendale si provvede poi a registrare gli aggiornamenti e a scegliere anche percorsi alternativi ma più efficaci all’ultimo minuto. In generale, tipologie di dati più utili per svolgere queste funzioni sono velocità, distanze, numero di fermate, aspetti di sicurezza, monitoraggio dei vicoli, quantità di carburante speso al chilometro...

L’implementazione di questa trasformazione digitale operata dalla società tramite AI, ML e IoT è chiaramente visibile già nei suoi hub: alcuni rimorchiatori senza conducente muovono la merce da una parte all’altra dell’edificio, assicurandosi che i prodotti da spedire

vengono, tramite i *Network Planning Tools* (NPTs), posizionati nelle giuste postazioni per ottimizzare i percorsi durante la spedizione o la consegna (Roby, K., 2020). Proprio i NPTs sono la migliore fonte di big data per UPS poiché danno una panoramica in termini di picchi e riduzioni dei volumi della rete. *“Che si tratti di un treno, di un rimorchio, che si stia spostando in un aereo, sappiamo cosa si sta muovendo in tutta la rete. Ci basiamo su questi dati per operativa per determinare come ottimizzare al meglio l'intelligenza artificiale e gli algoritmi di ricerca operativa”* afferma Rob Papetti, la mente dietro tali dispositivi.

Tra le più importanti innovazioni che UPS ha implementato c'è uno strumento chiamato ORION (On-Road Integrated Optimization and Navigation), la tecnologia proprietaria che oggi impiega il Machine Learning, alcuni algoritmi avanzati e l'Intelligenza Artificiale. Da quando ORION è stato attivato UPS è riuscita a risparmiare circa 10 milioni di litri di carburante e 160 milioni di km ogni anno, per non parlare della maggiore velocità nei processi di spedizione.

La prima versione fu lanciata nel 2012 e si serviva di algoritmi di Machine Learning che processavano circa 66.000 strade negli Stati Uniti, in Europa e in Canada per offrire agli autisti dei veicoli UPS soluzioni più efficienti nel prelevare e consegnare i prodotti.

Nel 2016 la compagnia americana apportò una modifica aggiungendo l'opzione UPSNav. La modifica fornisce ai conducenti indicazioni passo dopo passo per effettuare nella maniera più efficace consegne e ritiri, soprattutto in luoghi specifici che non sono visibili dalle strade come per esempio le banchine di carico.

Ad inizio 2020 il sistema è stato implementato inserendovi un meccanismo di Dynamic Optimization che ricalcola in tempo reale i percorsi di ogni singolo pacco da spedire in base al cambiamento degli ordini, del ritiro nella merce e dello stato del traffico (Peterson, K. – UPS, 2020). In aggiunta, si porrà maggiore attenzione su nuovi prodotti/servizi che richiedono maggiore reattività, affidabilità ma anche accuratezza e le esigenze dei clienti che cambiano di ora in ora potranno essere soddisfatte nel miglior modo possibile. Per concludere, le ultime implementazioni, oltre a ridurre più miglia e consumi, permetteranno anche di diminuire le emissioni di carbonio.

Puntando uno sguardo sul futuro, il CIO di UPS Juan Perez ha affermato che è molto entusiasta dal potenziale non espresso dei big data. Così aggiunge che negli anni a venire *“le Supply Chain dovranno essere più agili e reagire ancora più rapidamente, e saranno i dati a renderlo possibile”*.

3.2 Marketing

Tutte le volte che vediamo annunci online per prodotti che abbiamo cercato precedentemente o sono collegati ai nostri hobby e interessi, si tratta di BDA a lavoro. Più le società sanno cosa fanno le persone e come lo fanno, meglio possono anticipare e guidare le loro esperienze e comportamenti online.

Prima dell'avvento dei big data il marketing era basato principalmente su una forma di comunicazione a senso unico. I feedback si ottenevano passato un po' di tempo e tutto il meccanismo di messa a punto delle strategie per ottenere una migliore risposta dal mercato era molto lento. Si sentiva quindi la necessità di aprire tale divisione ad una comunicazione a doppia entrata. La funzione marketing è forse stata tra le prime nell'aver sperimentato le applicazioni della Big Data Analytics; il marketing ha sempre dovuto confrontarsi con i dati raccolti mediante sondaggi, ricerche di mercato, livello di soddisfazione dei clienti e altri fattori, ma ora più che mai ha a disposizione una enorme quantità di informazioni per lavorare ancora più efficacemente

rispetto a prima. Comparando le prime con i dati dei clienti raccolti oggi quali interazioni sui social media, percentuali di clic, dati di acquisto online, dati di geolocalizzazione, comportamenti nella navigazione sul web, utilizzo dei dispositivi mobili e altro ancora in termini comparativi, non c'è confronto.

Jeff Bezos e il suo colosso dell'e-commerce Amazon vengono ritenuti i pionieri dell'introduzione di metodologie di elaborazione per i big data nel marketing. Una prima e rivoluzionaria intuizione del fondatore e CEO dell'azienda americana fu "Cosa succederebbe se l'azienda potesse raccomandare libri specifici ai clienti in base alle loro preferenze di acquisto individuali?"⁵⁰. Tale strategia di marketing volta ad aumentare i ricavi fu inizialmente molto difficile da implementare poiché il sistema doveva operare e aggiornarsi in tempo reale. La svolta si ebbe con la progettazione dell'algoritmo *item-based collaborative filtering*: questo permette di basare i suggerimenti di Amazon sulla somiglianza tra prodotti acquistati dai clienti, suddividendoli in categorie (Atlantic Technologies, 2020). Tuttora l'algoritmo viene utilizzato dall'azienda e si è dimostrato piuttosto di successo tanto che, sul totale delle vendite, un terzo proviene da questi sistemi personalizzati e di *recommendation*, e disponendo di più informazioni riesce ad ottenere risultati ancora più accurati.

Da questa prima iniziativa si è poi sviluppato un fenomeno che ha prima coinvolto le aziende più importanti al mondo e poi influenzato a cascata altre imprese, diffondendo un trend che ad oggi è tra i più riusciti nel connubio tra BDA e funzioni aziendali.

Entrare in possesso delle informazioni giuste però comporta soltanto un grande potenziale, non implica automaticamente aver progettato una strategia di marketing efficace. Si tratta infatti unicamente di dati grezzi che senza un'altra componente hanno un'utilità relativa: si pensi ai big data come materia prima di qualità che senza l'ingegno di uno chef capace rimarrebbero tali. Se è vero che i big data hanno fatto conoscere nuovi canali per ottenere informazioni dai mercati e dai clienti, ci sono altri fattori ugualmente importanti per completare il quadro. Come abbiamo visto nell'esempio precedente sono le intuizioni derivate dai big data, le azioni che si intraprendono, le tecniche analitiche che fanno la differenza.

Le aziende osservano un impatto considerevole in tre aree chiave una volta integrate le informazioni con una strategia di gestione del marketing (SAS).

Una prima modalità di applicazione della Big Data Analytics nel campo del marketing è migliorare il *customer engagement*: si dà quindi l'opportunità di estendere le conoscenze di base su ogni singolo cliente. Tra gli obiettivi primari del marketing sin dall'inizio c'è identificare le caratteristiche degli individui che acquistano un prodotto e, una volta soddisfatti i giusti bisogni, di un futuro cliente. Mentre nel passato ciò era molto più difficile e lungo da studiare, oggi le tempistiche si sono accorciate notevolmente grazie all'Advanced Analytics e alla disponibilità di zettabyte di informazioni. Le aziende hanno accesso in tempo reale e riescono a processare moltissime informazioni per capire innanzitutto chi sono e dove sono situati i clienti e poi i gusti, le preferenze, le modalità e le tempistiche in cui desiderano essere coinvolti nelle iniziative dell'azienda. Uno strumento molto utile per aumentare il *customer engagement* è la *market basket analysis* (analisi del paniere di mercato) attraverso cui si sperimentano tecniche di data mining e *unsupervised learning* volte a trovare i pattern ricorrenti tra i beni/servizi venduti e ricavarvi alcune correlazioni.

Il tutto è finalizzato a prevedere future scelte negli acquisti e a proporle direttamente ai consumatori (Bobriakov, I., 2019), dando così vita ad una nuova tendenza settoriale chiamata *predictive marketing*. Un'ulteriore conseguenza portata da tale iniziativa è un più preciso processo di targeting:

⁵⁰ Citazione tratta da Mayer-Schönberger Viktor, & Cukier, K. (2013). Big data: the essential guide to work, life and learning in the age of insight. London, UK: John Murray.

la massimizzazione del rapporto tra il cliente e l'azienda e dunque un più alto valore aggiunto che percepiscono i consumatori sono due dirette conseguenze.

La *customer retention* e la **fidelizzazione del cliente** includono degli strumenti utili della Predictive Analysis per capire i fattori che esercitano un'influenza sul meccanismo della fidelizzazione. Inizialmente l'azienda ha bisogno di capire gli elementi necessari per attrarre nuovi clienti e far sì che questi si sentano soddisfatti dalla *customer experience*.

I programmi fedeltà fungono da espedienti essenziali nella raccolta dati perché grazie a loro si registrano moltissime informazioni sulle abitudini di acquisto, sulla distribuzione e preferenze dei clienti più leali verso i prodotti/servizi delle imprese. La combinazione tra i big data e l'analisi permette di conoscere ciò che spinge i clienti a continuare a fidarsi di quel brand, a farli tornare ancora e ancora ad acquistare determinati prodotti. Si riesce quindi anche a modellare, sulla base delle informazioni ricevute ed in seguito elaborate, l'offerta da proporre a questo tipo di consumatori procedendo verso una personalizzazione a tutti i livelli.

In ultimo, gli algoritmi di ML possono integrarsi ai classici modelli di cross-sell utilizzando più fonti di dati per suggerire ai clienti di acquistare beni/servizi complementari ed aumentare le vendite a breve termine. Nel lungo periodo, però, se i prodotti offriranno un valore aggiunto e collegato alle preferenze dell'individuo, si aumenterà notevolmente la brand loyalty.

Includere l'analisi dei dati ha dei risvolti anche sull'**ottimizzazione** e sul **miglioramento delle prestazioni** del marketing. L'uscita di una campagna pubblicitaria, la promozione di un nuovo prodotto oppure le altre iniziative che la sezione marketing struttura per un'azienda generano opinioni anche contrastanti. Saperne a distanza di poco tempo l'andamento grazie ai feedback forniti dai clienti significa avere la possibilità di elaborarli tramite algoritmi, studiare gli aspetti da migliorare, così come puntare sugli elementi di successo. Le opinioni sono presentate sotto forma di dati non strutturati: i social media e i blog con i relativi post, commenti, tag, video e messaggi riportano all'azienda un insieme di esperienze vissute nella prospettiva dei clienti (Bobriakov, I., 2019).

Attraverso questi processi è più facile e agevole pianificare gli step da seguire nella progettazione delle strategie, stimare i risultati di una campagna di marketing nonché effettuare test e misurazioni. Ci si muove quindi verso un'ottimizzazione dei costi e dei contenuti offerti poiché gli analisti già sanno su cosa puntare e su quali canali di distribuzione diffondere i messaggi.

Un quarto aspetto da menzionare è il processo che porta alla **segmentazione dei clienti** (sia nel marketing B2B sia nel B2C), implementato e velocizzato tramite metodologie di *clustering*. In particolare, la micro-segmentazione è una tecnica avanzata e pare un fenomeno sempre più usato nel marketing (Bobriakov, I., 2019). La ragione per cui ha acquistato una certa importanza è che permette di riunire gli individui in categorie molto più accurate, specialmente se ci si concentra sugli aspetti comportamentali.

Gli algoritmi di clustering svolgono la stessa funzione della segmentazione: raggruppare gli oggetti, in questo caso i consumatori, secondo tratti comuni. Secondo Jagota (2019) il processo avviene usando fattori singoli o multipli di classificazione che a loro volta si generano da dati strutturati o non strutturati. Il primo caso con fattori strutturati prevede una costruzione più semplice degli algoritmi poiché si vuole segmentare il mercato solo tramite un elemento (località, fascia di reddito, età, grandezza dell'azienda...). Utilizzando un dato non strutturato la struttura dell'analisi diventa più complicata ma è lo stesso risolvibile definendo una misura di somiglianza (per esempio la *jaccard similarity*⁵¹) e infine procedendo alla clusterizzazione. Ultima e più complessa tra tutte è la segmentazione multifattoriale che combina due o più classi di dati, sia strutturati sia non strutturati, per formare un cluster. Qui diventa di fondamentale importanza il tipo di misura di somiglianza

⁵¹ In ambito dell'insiemistica numero di elementi che hanno in comune diviso per il numero di elementi nella loro unione. Lo stesso si può applicare

utilizzata degli analisti poiché avrà un grande impatto sulla struttura dei gruppi, che può cambiare di molto a seconda delle aziende prese in considerazione.

Di base, il più grande vantaggio che l'applicazione della BDA porta al marketing è una più approfondita conoscenza dei consumatori.

3.2.1 Customer Relationship Management

Nella società odierna non è accettabile non conoscere i propri consumatori. Coloro che vengono a contatto con le aziende attraverso i più disparati canali vogliono un'assistenza appropriata. Amministrare i rapporti con i clienti e quindi voler migliorare il proprio Customer Relationship Management (CRM) implica che l'azienda debba interagire efficacemente con i clienti, sia attuali sia futuri. Ciò significa lavorare affinché tutti i punti di contatto con il cliente siano gestiti al meglio ed è un prerequisito per fornire un buon servizio clienti. Introducendo le tecniche della Big Data Analytics è adesso possibile prendere in considerazione anche informazioni non strutturate che non vengono date direttamente dal cliente ma sono ricavabili da fonti terze, in modo tale da trasformare questa sezione in un punto di forza dell'azienda (van Rijmenam, M., 2019). In passato infatti il CRM sfruttava quasi solamente dati strutturati, inseriti dagli interessati e la mancanza di informazioni a 360 gradi impediva il soddisfacimento degli obiettivi prefissati. Oggi la combinazione tra BDA e CRM va oltre e coinvolge l'aspetto del servire il cliente. Ma cosa significa esattamente servire il cliente tramite i big data?

Secondo van Rijmenam (2019) ci sono quattro aspetti da considerare quando si tratta di assistenza al cliente che portano il tradizionale Customer Relationship Management a un livello superiore: gestione, interazione, analisi, conoscenza.

Oltre alla gestione del cliente tramite dati strutturati quali informazioni sui contatti, indirizzi e ultime interazioni, c'è la parte che prevede un approccio a rovescio (dall'interno verso l'esterno): l'azienda gestisce il consumatore registrando alcune informazioni basiche e mandando comunicazioni ad orari e su canali prestabiliti. Nonostante la poca innovazione nei processi, questo ambito è comunque necessario per comprendere le caratteristiche primarie di un cliente.

Interagire con i clienti significa operare secondo logiche bilaterali, in cui i primi contattano la società e questa risponde più velocemente possibile. Significa anche coinvolgere tutte le componenti aziendali e tutti i canali a loro collegati come community e social network per creare nuovi dati strutturati come e-mail, commenti e post.

Si passa poi all'analisi dei dati immagazzinati tramite l'interazione e la gestione; questo è un processo in cui protagonisti sono gli analisti ma l'impresa svolge sempre la funzione di guida e indirizza gli analisti verso macro-obiettivi. Non a caso gli analisti agiscono inoltrando regolarmente report alla componente marketing oppure se sono richiesti approfondimenti. Con l'uso dei big data a queste mansioni si aggiunge alle richieste una particolare attenzione verso la proattività e la risposta in tempo reale.

La risultante è una conoscenza approfondita di tutti i clienti e come le loro abitudini cambiano con le variazioni nel mercato, grazie all'analisi dei dati misti e alla loro elaborazione (gestione+interazione+analisi). Anche qui nascono suggerimenti e raccomandazioni predittive volte a fornire prodotti giusti a prezzi giusti e tramite i canali giusti per ottenere un tasso di conversione più alto.

Prima si è parlato semplicemente di analisi delle informazioni ma non sono state specificate le diverse tipologie di analisi (van Rijmenam, M., 2019) performate per fornire un migliore servizio ai clienti.

Svolgere analisi di influenza permette di individuare tra tutti i consumatori che possono esercitare una notevole influenza sugli altri. Il vantaggio maggiore in tale prospettiva è saperne approfittarne per conoscere la struttura delle influenze aziendali.

Le analisi di modello vengono effettuate su uno o più dataset, meglio se composti da dati misti, e combinano le varie informazioni per trovare modelli applicabili ai comportamenti dei clienti. Un esempio è la modellizzazione causale in cui gli algoritmi hanno lo scopo di capire se e come un fattore influenza un altro.

Dare le migliori raccomandazioni ai clienti (analisi delle raccomandazioni) aumenterà il tasso di conversione; questi due fattori sono quindi direttamente proporzionali.

Nella *sentiment analysis* i ricercatori adoperano alcune tecniche per scoprire il pensiero dei clienti sui prodotti e, anche attraverso il *Net Promoting Score* (NPS), migliorare il servizio clienti concentrandosi sull'affrontare i problemi prima che vi sia un'eccessiva diffusione tra i consumatori.

L'analisi di marketing mira all'implementazione delle interazioni aziendali e con i clienti per diffondere le giuste decisioni e i corretti messaggi del marketing.

Infine, per costruire un Customer Relationship Management funzionante ed efficace è fondamentale la completa partecipazione dell'organizzazione nel connubio tra big data e CRM.

Il caso “Nike”

Il caso che verrà presentato in questa sezione è applicabile alle aree sia di marketing sia di Customer Relationship Management. Nike ha sviluppato una strategia che include tutti e due gli aspetti.

Tra le più note aziende di abbigliamento, calzature ed accessori sportivi c'è la multinazionale statunitense Nike, fondata nel 1964 con sede a Beaverton, Oregon. Per decenni l'azienda ha agito seguendo strategie di vendita dal commercio all'ingrosso ma ultimamente ha deciso di differenziare le strategie di vendita operando secondo logiche di vendita Direct-to-Consumer: Nike Direct, l'iniziativa è stata fondata proprio per questo scopo e quindi ne è l'esempio. Accanto alle tradizionali strategie di vendita che continuano ad essere rilevanti per l'azienda americana, la più recente visione di marketing tramite Nike Direct è largamente guidata dalle tecnologie di analisi dei dati e questa iniziativa ha contribuito per 10 milioni di dollari sul totale delle vendite del 2018 che nel 2020 sono destinate a crescere del 60% circa (Biswas, S., 2020).

Il grande focus nell'approfondimento dei rapporti con i propri clienti deriva dal fatto di voler offrire prodotti personalizzati per migliorare l'esperienza unica con cui il consumatore si confronta quando acquista un prodotto Nike. Per raggiungere l'obiettivo la compagnia processa l'enorme quantità di dati che raccoglie giornalmente in merito a gusti, preferenze e abitudini dei clienti grazie alle numerose applicazioni di training, shopping online e lifestyle che gestisce.

Sfruttando l'analisi dei dati e le relative tecnologie, Nike genera un valore aggiunto per i propri consumatori tramite due strade (Biswas, S., 2020):

- *Migliore consumer experience* - Ricavare informazioni importanti sugli utenti per prevedere il loro comportamento permetterà di offrire prodotti e promozioni su misura.
- *Aumento dei servizi per i clienti* - I dati che le app registrano sono elaborati tramite l'Advanced Analytics per poi essere trasformati in mezzi per rafforzare il rapporto tra l'azienda e i clienti. si potrà quindi offrire una varietà di servizi personalizzati come allenamenti ideati da esperti e suggerimenti nell'abbigliamento.

Nike ha lavorato sulla *customer retention* e l'acquisizione di nuovi clienti targettizzandoli e prevedendo il momento giusto per agire secondo le precedenti modalità ed avere una loro risposta più reattiva.

Per far funzionare al massimo questa strategia il volume e la qualità dei dati che genera sono essenziali. L'affidabilità dei modelli di *predictive analytics* e la *sentiment analysis* progettati in azienda, oltre che dalla loro corretta struttura dipende dalla bontà delle informazioni di base inserite nell'analisi. Operare secondo la logica Direct-to-Consumer significa articolare le attività secondo diverse parti. Poiché tutte le aree convergono verso unico obiettivo che è quello di servire i clienti personalmente su scala globale, Nike ha pensato bene di svilupparlo e fondarlo per la maggior parte sulla BDA (Biswas, S., 2020).

1. *Negozi brandizzati Nike*: Oltre all'aumento di questo tipo di store, la compagnia americana si sta concentrando sull'aumentarvi il livello tecnologico. Ciò li aiuta sia ad avere rapide informazioni e approfondimenti sul tipo di domanda dei consumatori sia a controllare le impressioni sul brand.
2. *Cancellazione del contratto con Amazon*: Questa mossa strategica mira ad avvicinare i contatti con i clienti e a registrare più dati che prima erano archiviati da Amazon.
3. *Acquisizione di data analytics companies*: Incorporare imprese o startup che hanno già brevettato tecnologie sull'analisi dei big data ha significato non partire completamente da zero, ma introdurre subito conoscenze avanzate che hanno permesso a Nike di avere un vantaggio competitivo su altre aziende dello stesso settore, che però hanno iniziato a sviluppare soltanto internamente determinate conoscenze.

Nel 2018 infatti l'azienda ha proceduto con la prima acquisizione di Zodiac, una startup che analizzava i dati dei consumatori anche tramite algoritmi di suggerimento. Le novità che Zodiac ha portato sono importanti: raggruppando le caratteristiche fondamentali degli utenti sui dispositivi connessi alle tecnologie Nike si rendono note le loro abitudini di acquisto. Se un individuo di solito compra scarpe di questo brand due volte l'anno e sono già passati 12 mesi dall'ultimo acquisto, l'azienda lo contatterà e lo inviterà a riprendere il tradizionale ciclo di acquisto (Barseghian, A., 2019).

Nike Inc. ha poi rilevato Invertex per rafforzare le conoscenze e le competenze sull'intelligenza artificiale.

L'acquisizione più recente è avvenuta nel 2019 con una piattaforma basata sul cloud e chiamata Celect. Quest'ultima, integrata con il sito web e l'applicazione Nike, usa le informazioni per anticipare l'andamento della domanda locale e capire quando e in che modo i clienti acquisteranno e quali prodotti.

4. *Ricerca di data scientist e machine learning engineer*: C'è un'altissima richiesta da parte dell'azienda di Data Scientists ed esperti nel mondo digitale per implementare le conoscenze su BDA e AI, così da progettare rapidamente tecnologie all'avanguardia e migliorare nell'attuazione della loro strategia.

È fondamentale a questo punto che Nike, essendo un *first mover* perché per prima ha portato l'analisi dei dati nel settore *sports apparel*, continui ad investire in questo senso per mantenere il vantaggio competitivo acquisito, anche a seguito dell'arrivo di *new entrants* quali Under Armour e Adidas che potrebbero metterlo in pericolo (Barseghian, A., 2019).

Capitolo 4

La trasformazione del settore dei trasporti aerei mediante la Big Data Analytics: alcuni case studies

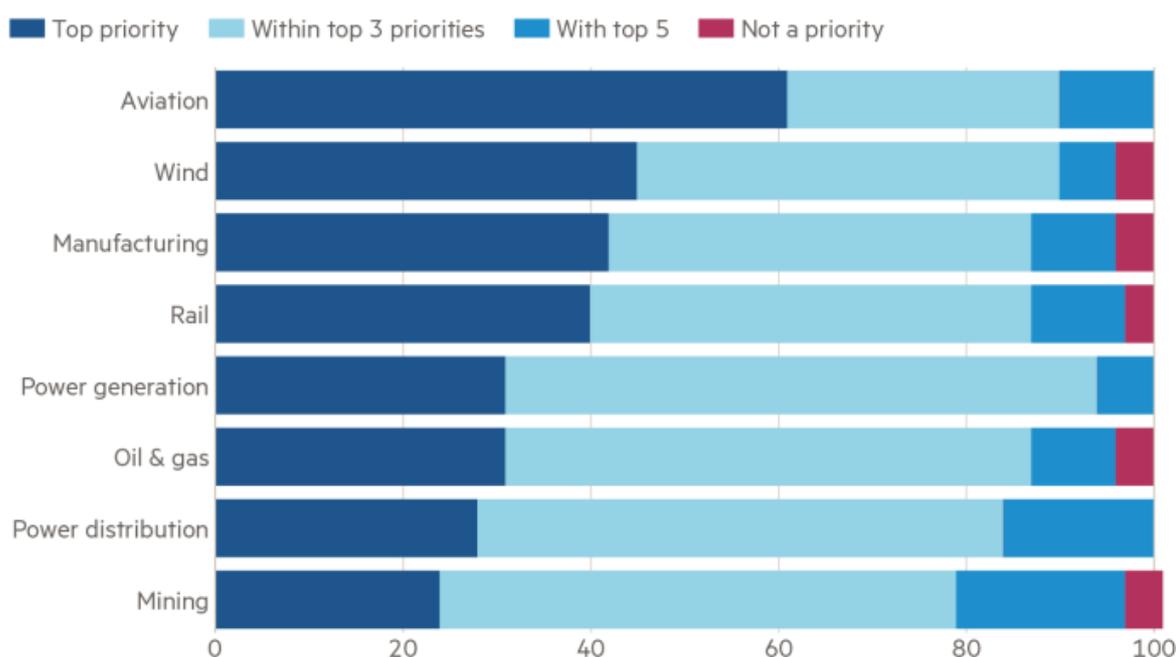
“Buongiorno signor Rossi, mi congratulo con lei per aver raggiunto il traguardo di 300.000 miglia aeree con la nostra compagnia. Il pranzo verrà servito tra due ore, potrà gustare il filet mignon cucinato proprio come piace a lei. Vorrei scusarmi ancora a nome di tutta la compagnia per il leggero ritardo del suo volo con destinazione Miami durante la scorsa settimana e informarla che oggi arriveremo in orario. Se poi controlla il suo bagaglio da stiva potrà notare che sta per essere caricato sul velivolo. Le abbiamo inoltre preparato anche il suo solito drink di benvenuto per iniziare nel migliore dei modi questo volo con noi.”

Potrebbero tranquillamente essere in un futuro non molto lontano le prime parole di un qualsiasi assistente di volo rivolte ad un cliente in prima classe alla luce dei progressi fatti nella Big Data Analytics da moltissime compagnie che operano nel settore dei trasporti aerei.

Come si può osservare dall'immagine (**Figura 7**) l'attenzione da porre sui big data nelle compagnie aeree è ritenuta la maggiore priorità per più del 60% degli intervistati, percentuale notevole soprattutto se si compara con le opinioni dei colleghi che operano in altri settori produttivi. Questo dato fa capire quanto le compagnie aeree stiano puntando sulla BDA e la grande fiducia che vi si ripone per la relativa crescita settoriale.

Big data analytics is important for airlines

Respondents who said big data was important relative to other priorities in their company (%)



Source: Accenture
© FT

Figura 7: % di intervistati che valuta la priorità che hanno i big data nella loro compagnia

Fino a circa vent'anni fa l'aereo era nient'altro che un mezzo di trasporto, comparabile ad una macchina oppure ad un autobus. Ora è tutto diverso. Oltre a provvedere allo spostamento di persone e merci adesso l'aereo trasporta con sé un insieme di valori aggiunti sia per il cliente sia per la compagnia che vi opera; in termini pratici e più semplicistici i velivoli trasportano enormi quantità di dati utili, per lo più privi di un'organizzazione.

Secondo gli studi effettuati da un gruppo di ricercatori dell'università indiana Amrita School of Engineering, in genere un volo transatlantico di una durata media genera circa 1 terabyte di informazioni (Waldmeir, P., & Hodgson, C., 2018). In pratica in solo in esso sono racchiusi i dati del servizio clienti, della biglietteria informazioni, delle previsioni meteo, delle comunicazioni aeroportuali ma si possono anche ottenere utili informazioni al di fuori del velivolo. I consumatori oggi cercano un'esperienza coinvolgente, personalizzata e senza intoppi e tutti i progetti di investimento sulle nuove tecnologie prima e di trasformazione digitale poi sono stati ideati per soddisfare questi bisogni. Attraverso tipologie di dati come quelle citate precedentemente si possono trarre indicazioni importanti per proporre servizi aggiuntivi e una customer experience personalizzata. È giusto precisare che questi scopi si raggiungono nel caso in cui i dati vengano correttamente organizzati e processati.

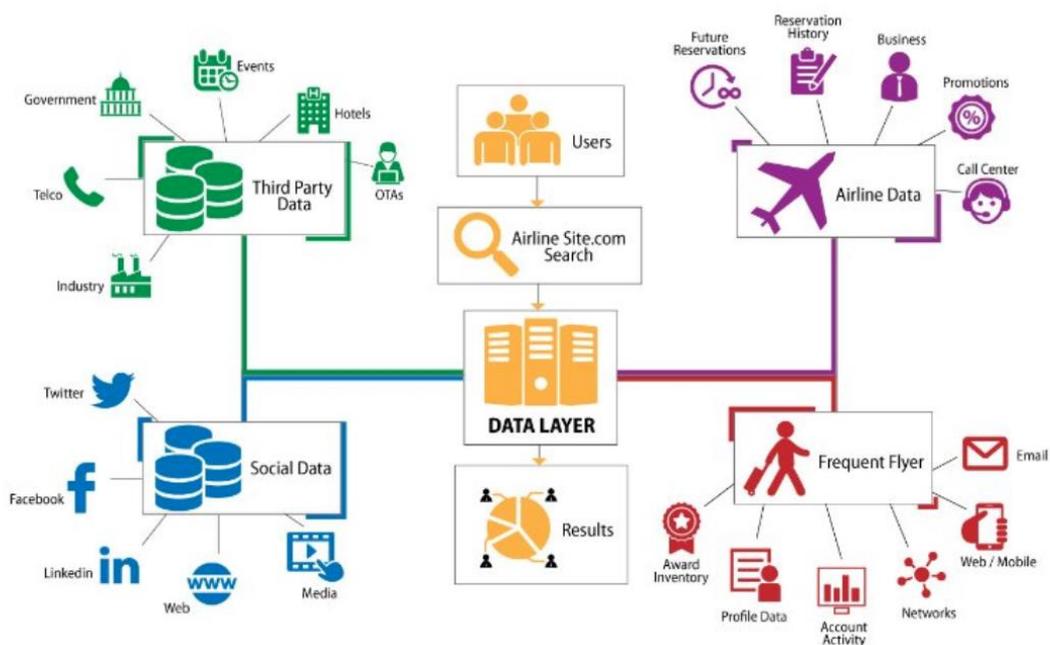


Figura 8: le varie fonti da cui le compagnie aeree possono ricavare i big data

Nel periodo precedente all'introduzione della BDA il settore dei trasporti aerei si poteva considerare nella fase di maturità; una curva di domanda abbastanza stabile, tendenza di tutte le compagnie a ridurre i costi sono le più evidenti caratteristiche di quella fase del ciclo di vita del prodotto. Con l'avvento dell'Advanced Analytics è come se l'industria avesse iniziato un nuovo ciclo perché le nuove tecnologie hanno portato ad una *disruption*. Non a caso i margini di profitto medi si stavano e si stanno tuttora riducendo, seppure in misura minore, e una completa applicazione dell'analisi dei dati sembrerebbe essere per le compagnie aeree un antidoto alla pressione dell'intensa competizione nel mercato e degli stretti margini (Waldmeir, P., & Hodgson, C., 2018).

Mentre molte compagnie sono state restie ad accogliere la digitalizzazione, alcuni tra i più importanti player del settore sono stati i primi ad introdurre i big data e ancora oggi cercano di implementare le loro strategie digitali e continuare a progettare il loro innovativo.

In questo capitolo saranno affrontate le maggiori applicazioni degli strumenti dell'Advanced Analytics al settore dei trasporti aerei. Si potrà notare come l'intera industria è stata completamente rivoluzionata sotto il punto di vista strategico e finanziario dall'avvento di questo fenomeno tramite le iniziative strategiche delle varie compagnie aeree.

4.1 Vueling Airlines e la disponibilità del servizio

Vueling Airlines ha sede a Barcellona ed è una compagnia low-cost quotata nella borsa di Madrid. Fu fondata nel 2004 e oggi la sua flotta conta 123 aerei, 155 destinazioni situate in Europa, in Medio Oriente e in Nord Africa.

Inizialmente la compagnia non faceva esattamente un uso ottimale dei dati poiché non possedeva le conoscenze adatte a risolvere i problemi che si poneva tramite la BDA. Il management voleva però comprendere meglio e in tempo reale delle informazioni per la disponibilità del servizio e l'andamento della domanda.

Una volta ideate varie startup a Barcellona in collaborazione con varie università, la situazione si ribaltò completamente. Vueling e le piccole aziende incominciarono pertanto ad analizzare e cercare di controllare fattori che influenzano pesantemente l'andamento del volo come il cattivo tempo, lo spazio aereo congestionato e gli scioperi. Queste erano tutte componenti che, se arginate con il metodo di Intelligenza Artificiale, avrebbero permesso alla compagnia aerea di avere orari di volo più robusti ed aiutare l'equipaggio con il processo di *decision-making* per migliorare l'esperienza del cliente. I risultati? Vueling è riuscita a trovare modi per ridurre l'impatto negativo dei fattori esterni, ha rafforzato il servizio clienti e molte fasi della *customer experience*; ciò è dimostrato da una crescita dei ricavi superiore al 4 % (Feliu, C., 2019).

I 10 milioni di dati generati ogni giorno nelle transazioni vengono impiegati anche nel design degli orari di volo per inserire nuove destinazioni che hanno un'alta domanda e quindi incontrare i desideri dei clienti (Feliu, C., 2019). Determinate città del continente africano ne sono la dimostrazione data l'inaugurazione dei nuovi percorsi e un aumento di più della metà di posti disponibili tra quelli già esistenti sui velivoli.

4.2 Delta Airlines nel monitoraggio dei bagagli

Delta Airlines ha sede in Georgia e si annovera tra le più grandi compagnie aeree negli Stati Uniti. Fondata nel 1924, oggi possiede una flotta di circa 800 vettori e opera tra l'intero continente americano, l'Europa, l'Asia, l'Africa e il Medio Oriente.

Delta ha quindi cercato di sviluppare un approccio alla big data Analysis rispettando un fondamentale principio insito nel suo business model quale il rapporto con i clienti. Non a caso

l'approccio tecnologico è volto a rafforzare il *customer retention*, come si evince chiaramente dal report annuale del 2017⁵².

La compagnia americana vuole quindi creare valore per propri passeggeri sia prima, sia durante, che dopo il volo per stringere un legame personalizzato e duraturo tra il brand ai clienti. E quale idea migliore se non concentrarsi sul potenziamento di un servizio che spesso è inefficiente in molte compagnie aeree come il monitoraggio dei bagagli? I maggiori investimenti nelle nuove strumentazioni digitali vertono su (Feliu, C., 2019):

- *Macchinari self-service per il controllo dei bagagli*: il primo aeroporto ad aver introdotto tali macchinari è stato quello di Minneapolis nel 2017. Ciò rende migliore e più veloce l'esperienza in aeroporto dei viaggiatori; alcuni macchinari sono stati anche dotati di riconoscimento facciale.
- *App di tracciamento dei bagagli*: una delle preoccupazioni più grandi dei passeggeri prima, durante e dopo il volo è sapere l'ubicazione dei loro bagagli e assicurarsi che questi non vengano persi nel viaggio. L'app creata da Delta mira a risolvere esattamente questo problema: gli utenti potranno vedere in tempo reale le stesse informazioni che lo staff usa per monitorare la posizione dei vari bagagli, si dimostra quindi un sapiente uso dei dati che sono già stati raccolti. In più, la compagnia ha effettuato un investimento di 50 milioni di dollari per sostituire alla scansione manuale del codice a barre i dispositivi RFID.

I frutti di questi investimenti si sono già visti, considerando il download da parte di più di 11 milioni di persone dell'app di tracciamento dei bagagli che ha implicato maggiore fedeltà dei clienti verso il brand. In aggiunta, le informazioni raccolte dal nuovo sistema RFID ha comportato un'identificazione di pattern chiave per scoprire i maggiori errori nella gestione dei bagagli, implementando soluzioni efficaci. Infine, il programma di mantenimento improntato sulla *predictive analytics* ha ridotto notevolmente le cancellazioni, i ritardi e le interruzioni di servizio. La società ha pubblicamente affermato che questi strumenti hanno drasticamente migliorato i punteggi dati dai clienti per la *customer satisfaction*, spingendo la compagnia verso ulteriore crescita.

4.3 EasyJet e le nuove strategie di prezzo

EasyJet PLC ha sede a Londra ed è una compagnia low cost quotata in borsa (London Stock Exchange). Fondata nel 1995, la sua flotta è composta da oltre 250 aerei e negli anni è arrivata ad offrire circa 820 percorsi in più di 30 paesi europei.

"I vincitori nel settore saranno quelli che massimizzeranno i big data e noi intendiamo fare questo". L'affermazione del CEO di EasyJet Johan Lundgren rispecchia appieno quella che è la volontà dell'azienda di far leva sulle conoscenze e competenze necessarie sviluppando nel 2018 un gruppo di lavoro composto da big data scientist e guidato da Luca Zuccoli, primo Chief Data Officer della compagnia (Feliu, C., 2019).

EasyJet ha utilizzato l'analisi dei dati per risolvere problemi concentrati in due aree chiave: ottimizzazione e impostazione dei prezzi.

Tra i progetti meglio riusciti c'è un algoritmo che permette al prezzo di un singolo posto di adattarsi in tempo reale alle variazioni nella domanda. Allo stesso tempo l'algoritmo analizza i dati storici

⁵² 2017 Annual Report of Delta Airlines Inc. (Rep.). (2017). Delta Airlines. Disponibile su http://www.annualreports.com/HostedData/AnnualReportArchive/d/NYSE_DAL_2017.pdf

in possesso di EasyJet per predire fino ad un anno di distanza alcuni trend nella curva di domanda. Tale strumento ha permesso di aumentare il profitto medio per posto di quasi il 20% in quattro anni.

Il team si è poi concentrato sulla gestione del magazzino, strutturando un modello predittivo per ridurre gli sprechi alimentari così da ottimizzare l'offerta di cibo sui velivoli. Grazie alla partnership quinquennale sviluppata con Airbus per accedere alla piattaforma di dati Skywise, la compagnia inglese ha voluto sfruttare pienamente i dati disponibili per il miglioramento del SCM. L'algoritmo e la partnership hanno ridotto cancellazioni e ritardi ma anche le tempistiche di intervento nell'evenienza di un problema tecnico. Ciò è fattibile anticipando le problematiche prima che accadano secondo tecniche predittive e sostituendo le parti danneggiate ancor prima che queste possano presentare difficoltà.

Nel 2018 la considerazione del brand EasyJet ha scavalcato British Airways nel Regno Unito, raggiungendo il primo posto e uno tra i fattori che più hanno contribuito a questo successo è stato il sapiente uso dell'intelligenza artificiale e dei big data.

4.4 SouthWest Airlines, ottimizzazione dei consumi di carburante e sicurezza sugli aerei

South West Airlines ha sede a Dallas, Texas ed è la più grande compagnia low-cost nel mondo oltre ad essere stata una pioniera nel business model low-cost di una compagnia aerea. Fondata nel 1871, opera utilizzando una flotta di 700 velivoli e serve circa 100 destinazioni di 9 paesi diversi compresi gli Stati Uniti.

Testando nel 2016 una piattaforma analitica, SouthWest Airlines ha effettuato un'analisi sui driver dei costi e dei ricavi. Come risultato è saltato all'occhio subito il fatto che l'azienda spendesse moltissimo denaro sul carburante (tra i 4 e i 6 miliardi dollari ogni anno) (Feliu, C., 2019). Di conseguenza la gestione del carburante era un fattore critico per la crescita nella compagnia e anche un suo piccolo miglioramento avrebbe voluto impattare positivamente sui costi annuali. Per gli analisti e la dirigenza dell'impresa in generale è diventato un problema da risolvere con la massima rapidità.

Da qui nasce la partnership con General Electric e la sua Flight Efficiency Analytics⁵³: questa soluzione è riuscita a scovare correlazioni tramite i dati per supportare strategie e decisioni operative più consapevoli. Il problema del carburante è stato quindi affrontato con GE e insieme hanno realizzato un modello di ottimizzazione dei consumi che ha avuto un enorme impatto sui ricavi aziendali. Nonostante la riservatezza mantenuta the Southwest Airlines sul risvolto quantitativo dell'iniziativa, è stato pubblicamente affermato che l'abbattimento di questi costi è stato sostanziale.

Un altro aspetto su cui si è focalizzata la compagnia americana è stata la gestione del contatto con i clienti. Anche qui l'azienda era alla ricerca di un'alleanza strategica o di una partnership che possedesse le conoscenze adatte per sviluppare questo tipo di tecnologia. Ha quindi deciso di

⁵³ MarketScreener. (30 settembre 2015). General Electric: Southwest Airlines Selects GE's Flight Efficiency Analytics: MarketScreener. Disponibile su <https://www.marketscreener.com/GENERAL-ELECTRIC-COMPANY-4823/news/General-Electric-Southwest-Airlines-Selects-GE-s-Flight-Efficiency-Analytics-21121640/>

collaborare con l'impresa Aspect, che era già da prima un fornitore dell'azienda per la Automatic Call Distribution (ACD⁵⁴).

Anche qui la scelta si è rivelata vincente perché ha permesso a SouthWest Airlines di transitare senza problemi verso meccanismi *data-driven* di *decision-making*. Ad oggi può fare affidamento su diversi strumenti che permettono di offrire un servizio clienti quasi impeccabile. Un esempio è l'algoritmo di *speech analytics* che riesce a captare le sfumature in ogni interazione registrata nell'interazione tra il servizio clienti ed i clienti stessi per risolvere spiacevoli inconvenienti dei consumatori e agire proattivamente.

In ultimo, la terza collaborazione promossa da SouthWest Airlines alla NASA volta a migliorare i livelli di sicurezza durante il volo. Da una parte la compagnia aerea voleva conoscere più approfonditamente determinate tecnologie e ricavare il massimo valore dai dati, dall'altra l'agenzia spaziale statunitense voleva trovare applicazioni ai processi specializzati di data mining nelle tecnologie di tutti i giorni. Il sistema automatizzato, nato dalla collaborazione, include il Machine Learning e attraverso questo si cercano inconsistenze tali da creare possibili criticità nella sicurezza dei velivoli. Il direttore della sezione sicurezza in volo si esprime così in merito alla cooperazione *“La NASA ci ha fornito uno strumento che ci ha aiutato a guardare i nostri dati in un modo a cui non avevamo mai pensato prima”*.

4.5 La funzione *“Know me”* di British Airways

British Airways ha sede a Londra ed è la compagnia di bandiera del Regno Unito ed opera globalmente. È stata fondata nel 1971 e usufruisce dei 294 aerei a disposizione per trasportare i passeggeri in circa 200 destinazioni di tutto il mondo.

La compagnia ha riconosciuto in colossi come Uber, Amazon e Google delle grandi minacce per il futuro dell'industria dei trasporti aerei. Secondo le loro aspettative tali imprese saranno capaci di competere soprattutto a livello di prezzi e di concezione futura di ogni business con le attuali compagnie aeree. Per tale motivo nel 2018 British Airways si è impegnata ad investire 4,5 miliardi di sterline per operare una trasformazione digitale su tutte le offerte ai clienti⁵⁵.

Gran parte di questi fondi sarà utilizzato per potenziare ulteriormente attraverso la BDA sezioni come il Customer Relationship Management e migliorare sotto l'aspetto della fidelizzazione.

Poiché precedentemente c'era un gap di informazioni tra la compagnia dei clienti, il programma *“Know me”* di British Airways, lanciato nel 2013 e poi implementato notevolmente negli anni a venire, utilizza l'Advanced Analytics per offrire ai passeggeri una piacevole esperienza di viaggio e offerte targettizzate, rilevanti e personalizzate. Tale meccanismo è stato possibile grazie all'interazione dei clienti con il brand e viceversa.

L'azienda è partita analizzando i dati per comprendere al meglio la propria *customer base* e ha capito che inizialmente era composta in gran parte da professionisti a corto di tempo che hanno bisogno di servizi precisi e veloci. I relativi dati forniti dal programma *“Know me”* vengono processati e poi trasmessi agli assistenti di volo tramite i tablet, in cui potranno osservare informazioni relative ai clienti come l'itinerario di viaggio e la cronologia dei reclami a cui la compagnia ha accesso. *“Stiamo essenzialmente cercando di ricreare la sensazione di essere*

⁵⁴ Consiste in un sistema di comunicazione privata che riesce a convogliare le chiamate in ingresso verso sistemi di Interactive Voice Response (IVR - risposta automatica) oppure verso un insieme di operatori.

⁵⁵ Palladino, D. (8 novembre 2017). British Airways, 4,5 miliardi di investimenti per rispondere alle low cost. Disponibile su https://www.webitmag.it/british-airways-45-miliardi-di-investimenti-per-rispondere-alle-low-cost_132345/

ricosciuto che ottieni nel tuo ristorante preferito quando lì sei il benvenuto, ma nel nostro caso sarà dato da migliaia di dipendenti per milioni di clienti. Questo è solo l'inizio” ha affermato nel 2012 Joe Boswell, a capo della *customer analytics*.

Dopo una serie di migliorie apportate al sistema, oggi British Airways permette ogni giorno alla sua crew di servire al meglio i propri clienti, anticipando i problemi e risolvendoli prontamente in modo tale da stupire positivamente i passeggeri. È diventato quindi semplice provvedere ad una micro-segmentazione dei clienti, permettendo comunicazioni personalizzate così come offerte e servizi. È così che i clienti premium, i viaggiatori frequenti e coloro che usufruiscono di British Airways per la prima volta avranno trattamenti differenziati basati sia sulle precedenti esperienze sia sulle loro preferenze e/o bisogni. Se dalle comunicazioni che la crew riceve un cliente ha riscontrato problemi o ritardi di ogni tipo, i gli assistenti di volo sapranno dare loro la giusta attenzione per riparare al precedente errore e migliorare la *customer experience*.

Un altro punto di forza su cui è basata la strategia di servire i clienti al meglio è l'estrema agilità tenuta in considerazione nel modellare gli strumenti di analisi, che si rispecchia in un rapido adattamento dei servizi al cambiamento delle esigenze del cliente (Feliu, C., 2019).

In generale, si può affermare che l'iniziativa è stata un successo in termini di ricavi e fidelizzazione da parte dei clienti, anche osservando i feedback di questi ultimi. I passeggeri sono ben felici di essere coccolati da British Airways tanto da riconoscere anche il fatto che le informazioni rilasciate siano utilizzate per il bene del cliente e non solo per uno scopo puramente economico. A rompere un po' questo equilibrio è intervenuto nel 2018 il furto di informazioni sui consumatori su circa 500 mila carte di pagamento, a seguito del quale un anno dopo sono stati condannati a pagare 204 milioni euro di multa, una diretta conseguenza delle norme del GDPR⁵⁶.

4.6 Raccomandazioni future sul settore trasporti aerei

A seguito dell'analisi delle modalità di applicazione in diversi ambiti della Big Data Analytics nel settore dei trasporti aerei è importante notare ancora i grandi progressi effettuati da questa industria sin da quando è stata manifestata la volontà di processare i big data. Maggiore qualità del servizio clienti, previsioni più accurate, aumento della fidelizzazione dei clienti, cost management e ottimizzazione dei processi nella Supply Chain sono solo alcune delle tante applicazioni di successo che stanno caratterizzando molte società di trasporto aereo.

Ovviamente, se adesso le compagnie aeree si concentrano su uno o più ambiti in cui agire tramite la BDA, in un futuro non troppo lontano sarà solo un dato di fatto e non più una novità e le aziende dovranno concretamente pensare ad applicarla a 360°. Questa è la tendenza e se non la si comprende ed attua, continuare ad operare in un settore così altamente digitalizzato e tecnologizzato senza i giusti strumenti sarà al limite dell'impossibile.

La vera sfida è sviluppare oggi una solida strategia basata sui dati a livello corporate per il domani. Nel prossimo decennio la più grande fonte di vantaggio competitivo non approfondirà *se* si useranno i big data e *in che ambito*, ma sul *come* le imprese struttureranno i loro modelli analitici e, in particolare, come si articolerà la *customer experience* che fungerà da differenziatore tra i vari brand.

⁵⁶ Calzetta, G. (9 luglio 2019). Multa da 204 milioni di euro a British Airways per il furto dei dati dei suoi clienti. Disponibile su <https://www.ilsole24ore.com/art/multa-204-milioni-euro-british-airways-il-furto-dati-suoi-clienti-AC8ESSX>

Come hanno già fatto alcune compagnie aeree citati precedentemente un altro punto chiave da tenere in considerazione sia ora che in futuro è la scelta del personale con le competenze adatte e la formazione di un team altamente qualificato e robusto. I data scientist sono quindi fondamentali risorse e bisogna fare affidamento sul loro sapere per risolvere problemi tramite l'analisi delle informazioni. Anche le partnership con aziende che hanno una preparazione specifica e possono giovare nel trasferimento di *skill* e *know-how* quando la compagnia ha bisogno di rapide integrazioni alle conoscenze pregresse.

Ultima ma non meno importante è l'idea di sviluppare una strategia dei big data rispettando quelli che sono i valori aziendali, poiché il processo di unione tra i valori dell'impresa e la nuova strategia sarà incorporato più facilmente e implementato più velocemente a livello di cultura aziendale, che si ricorda deve essere sempre improntata verso l'innovazione per avere successo (Feliu, C., 2019).

Capitolo 5

Quali sviluppi per i Big Data in futuro?

Che tipo di prodotti e servizi digitali comprenderemo nel futuro? Quali apprezzeremo di più? Che tipo di strumenti tecnologici legati alla Big Data Analysis ci permetteranno di fare la differenza sul mercato? Come possiamo sfruttare al meglio la tecnologia per progettare i modelli dell'avvenire? Quanto incideranno la creatività, la ricerca umana, il valore aggiunto sui potenti strumenti di calcolo ideati?

Nell'era digitale in cui viviamo, in moltissimi angoli del mondo, i ricercatori hanno l'obiettivo di trovare nuove tecnologie o comunque qualcosa di funzionale ed innovativo per supportare le tecnologie già esistenti. Di conseguenza, ogni giorno vengono registrate nuove tendenze o scoperte che cambiano il modo di vivere, operare o concepire determinati fenomeni.

Proprio per questo motivo alle domande che sono state poste prima non c'è una risposta univoca, ma piuttosto parziale perché ci si basa sull'anticipazione di trend futuri con le informazioni che si hanno a disposizione in un determinato momento, bisogna pertanto ricordare che si tratta pur sempre di previsioni.

Se già nell'economia digitale odierna i big data fanno da propulsore per effettuare varie analisi e arrivare a ricavarne indicazioni utili nella gestione del proprio business, nel futuro saranno ancora più protagonisti tali elementi, anche al di fuori della parte digitale dell'economia. Sarà quindi un processo inevitabile perché il valore che genera l'Advanced Analytics, se usata nel modo giusto, non è comparabile con il valore generato senza. Prima o poi quindi tutte le aziende, anche nel loro piccolo, dovranno adeguarsi a questi cambiamenti, che porteranno loro nuovi modi di fare business il nuovo applicazioni interessanti e per la crescita dentro e fuori il loro settore.

AI e ML continueranno a migliorare nell'impostazione dei relativi algoritmi e il modo in cui verranno usate varierà nel tempo permettendo crescita e cambiamenti nelle società. Il Machine Learning sta diventando sempre più sofisticato ogni anno che passa. E come afferma vicepresidente e General Manager di Intel, Wei Li, *“dobbiamo ancora vedere il suo pieno potenziale, oltre alle auto a guida autonoma, ai dispositivi di rilevamento delle frodi o alle analisi delle tendenze del commercio al dettaglio”*.

Un particolare focus si porrà poi sulla *prescriptive analytics*, oggi ancora troppo poco usata per via dei complicati metodi di progettazione, ma di fondamentale importanza per suggerire azioni di migliorative in grado di operare a tutti i livelli (processi, marketing, meccanismi *cost saving*, personalizzazione dei prodotti e dei servizi ...). È tramite quest'area dell'Advanced Analytics che si vedono infatti i risultati più stupefacenti se si pensa a una diminuzione dei costi oppure ad un aumento dei ricavi, per non parlare del raggiungimento di un ottimo rapporto con i clienti.

Avendo assistito negli anni precedenti all'esplosione della tecnologia cloud, un secondo campo di sviluppo primario è la progettazione di un'architettura digitale in grado di coordinare e governare una mole di dati in continua espansione a velocità record.

Microsoft Azure, Google Cloud Platform hanno cambiato le modalità di archiviazione elaborazione dei big data. In precedenza, per le aziende era importante la crescita fisica dei propri data center per elaborare sistemi con alta intensità di dati. Adesso, grazie a servizi come *pay-as-you-go*, la soluzione cloud garantisce una certa facilità d'uso, agilità e scalabilità per le aziende (Khvoynitskaya, S., 2020) che non sono più costrette a spendere molte delle loro risorse per l'ampliamento degli spazi di archiviazione negli edifici.

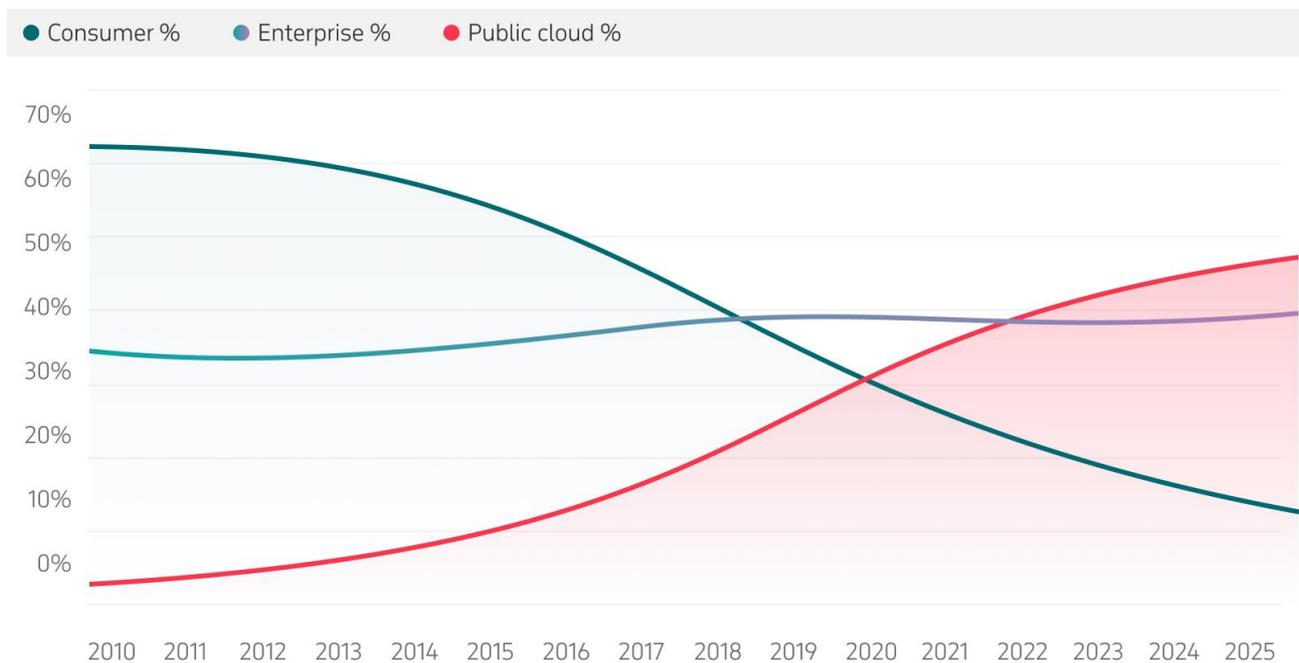
Questa rapidità di archiviazione e produzione dipende anche dal fatto che oggi strumenti in precedenza puramente presenti in forma fisica come transazioni, infrastrutture (IaaS) e inventari, trovano un loro spazio anche solo a livello virtuale e perciò si procederà verso una visione olistica dei meccanismi dell'azienda grazie all'efficace gestione dei big data (Pearlman, S., 2019).

Analisi di tipo conservativo sugli sviluppi dei big data mostrano anch'esse una costante e progressiva diminuzione delle infrastrutture archiviate fisicamente, a fronte di una rinnovata fiducia verso i sistemi di archiviazione virtuale (Pearlman, S., 2019). Ciò comporta però un'alta dipendenza (che tenderà a crescere in futuro) delle imprese di ogni settore dagli strumenti o dalle imprese partner, in grado di supportare e realizzare questi processi di immagazzinamento dei petabyte di dati generati quotidianamente.

Secondo Khvoynitskaya (2020) si stanno aggiungendo al cloud tradizionale altri due modelli di archiviazione, come è anche possibile notare nella **Figura 9**:

- *Ambienti ibridi* in cui si opera una scelta che usa logiche prioritarie nella decisione di mantenere parti dei dati nei server collocati in azienda e spostare il resto sulle piattaforme cloud;
- *Ambienti multi-cloud* in cui le imprese propendono per la totale archiviazione una combinazione tra più piattaforme cloud sia pubbliche sia private.

Where is the data stored?



Data source: Data Age 2025, sponsored by Seagate with data from IDC Global DataSphere, Nov 2018

Figura 9: andamento negli anni dell'archiviazione su varie piattaforme

Anche in base a ciò che abbiamo analizzato fino ad ora sembra chiaro quanto la data science, i big data in sé e pure la loro elaborazione diventeranno vitali prossimamente.

Oltre agli sviluppi individuati precedentemente si possono osservare alcune macro-tendenze che forniranno ulteriore carburante per l'evoluzione del fenomeno BDA. Verranno quindi presentate nei paragrafi seguenti.

5.1 I big data e l'Internet of Things (IoT)

L'Internet of Things (IoT) è tra le maggiori piattaforme, nonché enorme fonte di risorse, da cui si generano i big data, visti i miliardi di dispositivi che consumano oppure producono un grande volume di informazioni. Nell'IoT si assiste all'interconnessione tra sensori intelligenti e dispositivi integrati all'interno di una rete infrastrutturale molto dinamica, che permette nei campi di applicazione flessibilità, agilità, scalabilità e ubicità. Il bisogno di tenere sotto controllo, poi analizzare e infine agire su tali informazioni, genera delle problematiche relative alla sicurezza di alcune comunicazioni e calcoli, la privacy, alla verifica dei dati e data mining.

Infatti la scarsità di tecnologie associate agli strumenti dell'IoT si configura come una delle più grandi debolezze a cui vanno incontro le imprese. L'analisi in tempo reale è ciò che meglio caratterizza l'Internet of Things e spesso ancora non ci sono i mezzi per archiviare, processare e condividere subito i risultati, in modo tale da prendere decisioni efficaci immediatamente. Di conseguenza, nuove soluzioni e più strumenti devono essere messi a disposizione per unire completamente e definitivamente due fenomeni quali big data e IoT.

Secondo RCN Business (n.d.), nel futuro, con il crescere degli *smart devices* utilizzati nel mondo e la loro combinazione con big data e intelligenza artificiale, più dati verranno generati e maggiori opportunità ci saranno per le aziende di innovare attraverso la creatività e l'Advanced Analytics.

È stato stimato che il mercato IoT globale passerà dai 157 miliardi di dollari poi raggiunti nel 2016 a 457 miliardi entro il 2020 e che l'industria dei Big Data passerà da \$ 130,1 miliardi a più di \$ 203 miliardi nello stesso periodo (RCN Business., n.d.).

5.2 Artificial Intelligence per valutare la qualità dei dati

Di tutta la vastità di dati che viene prodotta e immagazzinata ogni giorno dalle varie organizzazioni, le informazioni non sono sempre e totalmente utilizzabili. Ciò accade perché entra in gioco una proprietà fondamentale che devono possedere i big data per essere processati: la qualità.

La qualità fa riferimento alla percezione di quanto i dati siano adeguati ad essere inseriti nei modelli algoritmici nel risolvere un problema. Bisognerebbe quindi procedere con una valutazione per rilevare la bontà dei dati, basandola su sei dimensioni (Raj, J. 2019):

- *Unicità*: se ogni inserimento dei dati è unico nel suo genere;
- *Coerenza*: un dataset si considera coerente quando tutte le tecnologie dell'impresa riproducono le stesse informazioni;
- *Precisione*: intesa come il grado di accuratezza del dato rispetto all'oggetto nel mondo reale oppure ad un determinato evento;
- *Validità*: i dati sono validi se coincidono con il formato, la tipologia e l'intervallo della relativa definizione;
- *Atemporalità*: si riferisce alla disponibilità dei dati quando richiesti;

- **Completezza:** i dati sono completi quando soddisfano le aspettative predeterminate. completezza in questo caso è sinonimo di *expected comprehensiveness*.

Stabilito che le organizzazioni sono consapevoli dell'importanza dei big data e del loro contributo sulla crescita aziendale, se una società è in possesso di dati di qualità scarsa, l'analisi, seppur progettata in maniera corretta, non darebbe i suoi frutti e anzi, porterebbe a comportamenti sbagliati dell'azienda. Si dice che le imprese credono che oltre il 27% dei ricavi totali siano sprecati causa imprecisione dei dati impiegati (Lindes, J., 2019).

I ricercatori stanno facendo lavorare insieme ML e AI per risolvere tale problema ed aprire alla possibilità futura di costruire algoritmi in grado di selezionare i dati di qualità e separarli dal dataset generale. La futura "depurazione" dei dati opererà, secondo Raj, J. (2019), verso quattro tendenze differenti.

Acquisizione automatica dei dati - Citando una ricerca effettuata da Gartner, ogni anno c'è il mancato guadagno di 14,2 milioni di dollari per via della cattiva qualità dei dati acquisiti. Si stanno sviluppando però degli strumenti che utilizzano l'AI attraverso la cattura intelligente, molto valida quando bisogna implementare il meccanismo di immissione dei dati. Grazie a tale processo le aziende si assicureranno che le informazioni chiave saranno registrate ed archiviate, senza riportare le problematiche precedenti e registrare perdite così ampie.

È importante poi sottolineare che l'automatizzazione è un elemento chiave di questi algoritmi di cattura, si evita quindi che le attività manuali di acquisizione delle informazioni possano alterare il dataset. Le risorse sono tutte trasferibili sulla parte analitica e gestionale della BDA e delle relative attività.

Identificazione di registrazioni duplicate - La cattiva qualità è dovuta non solo alle informazioni in eccesso o mancanti ma anche alla presenza di registrazioni duplicate, con possibile conseguenza di banche dati obsolete. Anche qui può venire in soccorso l'intelligenza artificiale, con l'obiettivo primario di eliminare i doppi di dati nei registri delle organizzazioni e mantenere un certo standard di qualità. Per le aziende si tratta di un compito arduo dato che bisogna progettare sistemi molto sofisticati per essere all'altezza della funzione precedentemente descritta.

Un esempio di successo nell'implementazione di tali sistemi intelligenti è nel Customer Relationship Management dell'azienda Salesforce. L'algoritmo, tramite le impostazioni predefinite, sfrutta la capacità di riconoscere chiavi duplicate nei contatti e negli account aziendali ed eliminarle per garantire l'assenza di informazioni sovrapponibili (Raj, J. 2019).

Rilevamento delle anomalie - Gli errori umani purtroppo influiscono sulla qualità e, di conseguenza, sull'utilità di un dataset. Anche qui l'intelligenza artificiale può avere un impatto positivo trovando rimuovendo questo tipo di errori. Se poi le aziende sviluppano anche algoritmi di apprendimento automatico sulle anomalie riscontrate, si avrà un ulteriore miglioramento nel livello delle informazioni.

Inclusione di dati di terze parti – Un'ultima frontiera su cui si sta sviluppando l'innovazione nella qualità dei dati grazie all'AI è l'aggiunta di informazioni utili al dataset. Mantenere e correggere l'integrità è fondamentale, ma si possono registrare dati più completi e di maggior valore se le organizzazioni inseriscono elementi ulteriori grazie ad informazioni registrate da terze parti. In questo caso la funzione dell'intelligenza artificiale è suggerire se e quali informazioni recuperare in una banca dati e trovare le eventuali connessioni tra loro.

Dal momento in cui l'azienda possiede un insieme di informazioni in un unico posto e di alta qualità, ha una più grande probabilità di prendere decisioni efficaci.

5.3 Data as a service (DaaS)

Gli strumenti di AI, ML, data mining e BI richiedono dati da un unico database relazionale (RDB) che fornisce alte prestazioni. Tuttavia, la maggior parte delle imprese che li utilizza dispone di più fonti, formati di archiviazione e dispositivi da cui provengono i dati. Per ovviare a questo problema i data scientist utilizzano e personalizzano i processi di Extract, Transform and Load (ETL) e tecnologie proprietarie per combinare e uniformare le informazioni tra i vari metodi di archiviazione e allo stesso tempo migliorare la facilità con cui si accede alle soluzioni di analisi (GlobalData Technology, 2020).

Tale approccio però è troppo dispendioso sia a livello di tempo sia di risorse impiegate per risolvere i problemi di incompatibilità tra sistemi. Gli analisti senza questo problema si potrebbero altrimenti concentrare sul miglioramento dei modelli di Advanced Analytics. Ulteriori debolezze che comportano i processi ETL sono la complessa gestione (governance) dei dati, gli alti costi di infrastruttura, la bassa agilità del sistema e la maggiore complessità dei dati (GlobalData Technology, 2020).

Gli analisti, decisi a risolvere definitivamente questo problema, hanno pensato bene di ideare le piattaforme DaaS - Data as a Service, un servizio presente nel cloud così come era stato fatto precedentemente con i software (SaaS), le piattaforme (PaaS) e le infrastrutture (IaaS). Attualmente Microsoft, Oracle e Salesforce offrono alle imprese soluzioni di DaaS.

Il DaaS garantisce alle aziende un accesso a tutti i dati selezionati e su richiesta, e un altro suo obiettivo è la consolidazione e l'organizzazione dei dati in pochi posti centralizzati per migliorare i meccanismi di trasparenza quando si condividono e usano dati provenienti da servizi e sistemi diversi.

DaaS è utile poi per la condivisione dei dati tra diverse organizzazioni, ottenendo collaborazioni durature e rafforzate e trasferendo le conoscenze da un'organizzazione all'altra (Hazelcast, 2020). Inoltre, i clienti dei proprietari dei sistemi Data as a Service risolvono gran parte dei problemi di archiviazione, analisi e interpretazione, poiché insieme alle altre tre componenti del cloud riescono a svolgere tutto senza necessariamente impiegare i loro server in questi processi.

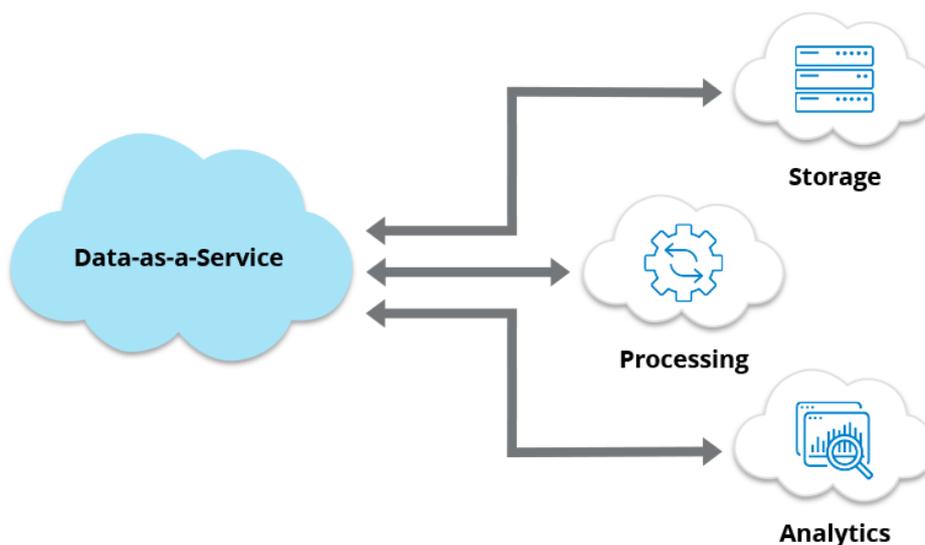


Figura 10: DaaS consente ai clienti di gestire nel cloud archiviazione, elaborazione e analisi

Alcuni dei vantaggi di tale tecnologia possono anche essere visti come debolezze (Hazelcast, 2020). Dal momento in cui l'intento è offrire un servizio potenzialmente utilizzabile da tutti, le aziende potrebbero essere sopraffatte dalle opzioni tecnologiche. Talvolta infatti c'è incertezza nel determinare quale sia approccio giusto per operare in un DaaS e perfino da dove iniziare.

Le soluzioni correntemente esistenti di servizi Web o SOC⁵⁷ potrebbero non soddisfare i requisiti dell'approvvigionamento e sicurezza di DaaS su Internet e cloud.

Essendo già un'innovazione recentissima e orientata al futuro, nei requisiti di miglioramento di tale strumento c'è sicuramente una risoluzione ai problemi precedentemente presentati e, di conseguenza, la necessità di sviluppare nuovi modelli, metodi e architetture per materializzare ancora di più i vantaggi dei DaaS.

5.4 Come cambiano i modelli di BDA con l'avvento del Covid-19?

Ciò che è accaduto negli ultimi mesi con la diffusione del Coronavirus in tutto il mondo ha avuto come conseguenza un cambiamento nei comportamenti degli individui a livello umano ma anche economico. È per questo che alcuni modelli delle aziende potrebbero non essere più rappresentativi per la situazione attuale, specialmente quelli di analisi predittiva, poiché in qualche modo bisogna tenere in considerazione nelle previsioni lo shock economico verificatosi globalmente.

Rezzani (2020) menziona alcune linee guida volte a verificare la tenuta dei modelli predittivi precedenti al Covid-19 e a modificarli eventualmente.

In primo luogo l'azienda deve procedere con un'analisi di *quanto* e *se* le variabili interne del modello preso in considerazione siano state intaccate dallo shock esterno. L'analista o gli analisti aziendali devono porsi domande concatenate, tra cui: ci sono variabili economiche e/o comportamentali su cui lo shock ha avuto impatto? Che variazione si è osservata? Ci sono altri modelli costruiti su queste variabili? Che impatto ha un modello falsato sul business aziendale?

Una volta trovata risposta ad ognuna di quelle domande, individuate le variabili oggetto di un consistente cambiamento a causa del Coronavirus, si ragiona su come inserire le modifiche nell'algoritmo per avere una rappresentazione e predizione più fedele della realtà e i rischi collegati a una ridefinizione del modello.

Terminato anche questo passaggio, gli analisti passano ad attuare alcune trasformazioni matematiche prendendo la funzione della variabile interessata e rendendola più robusta agli shock, senza modificare il significato intrinseco. Se si tratta invece di serie storiche bisogna assolutamente evitare di applicarvi strategie di *differencing e rolling statistics* perché fonti di amplificazione e reiterazione degli errori nella predizione.

In ultimo si effettua la vera e propria modifica del modello utilizzando diversi strumenti. Uno tra questi è il proxy, che simula la risposta dei modelli progettati una volta che sono sottoposti a shock estremi. Questi ultimi ricreano sinteticamente eventi passati ma con caratteristiche simili a quelle che lo shock, in questo caso il Coronavirus, ha avuto.

Nella **Figura 11** si può osservare direttamente una rappresentazione grafica dell'*algoritmo decisionale dell'importanza* delle variabili che spiega direttamente gli step da seguire.

⁵⁷ Un Security Operations Center (SOC) è un centro da cui vengono forniti servizi finalizzati alla sicurezza dei sistemi informativi dell'azienda stessa (SOC interno) o di clienti esterni.

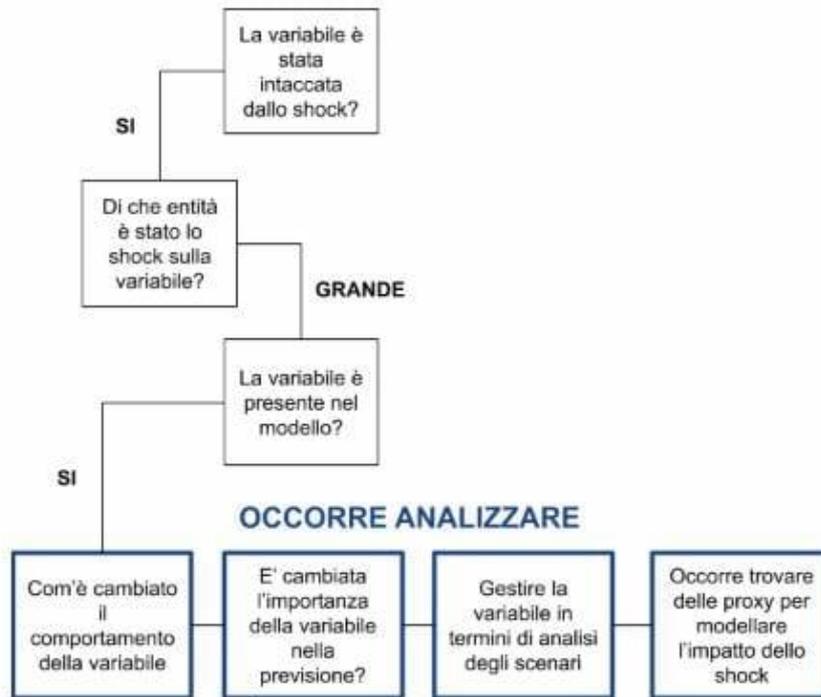


Figura 11: rappresentazione grafica dell'algoritmo decisionale

Gli altri strumenti da usare vanno in base alla durata delle conseguenze portate dallo shock nell'ambito del modello considerato, alle aspettative future per quanto riguarda la ripetizione dell'evento estremo e alla rimozione o meno di alcuni dati rilevati durante il periodo critico.

CONCLUSIONI

Lo scopo di questa ricerca è capire se davvero la Big Data Analytics può essere considerata il nuovo oro, tanto anche da spodestare l'importanza primaria che oggi assume il petrolio.

Il focus nel corso della tesi è stato posto su vari aspetti della gestione aziendale; per questo motivo le conclusioni vertono in tal senso, pur essendo molto chiaro che ormai l'Advanced Analytics e i big data permeano a 360° il mondo che ci circonda e molti successi nelle applicazioni aziendali possono essere traslati agevolmente nei più disparati aspetti della vita quotidiana.

Il fenomeno della BDA è ad oggi in piena espansione grazie a modelli predittivi, prescrittivi e automatizzati, sviluppati tramite ML, AI, Cloud Computing e Data Mining e, come visto nel quinto capitolo, c'è ancora spazio per numerosi miglioramenti che permetteranno di analizzare i meccanismi della gestione aziendale sotto nuove prospettive.

Fino ad ora è comunque possibile affermare con sicurezza che la Big Data Analytics ha completamente rivoluzionato le modalità di *decision-making*, così come la cultura decisionale del top management che adesso viene supportato dalle indicazioni che fornisce l'analisi dei dati e non più soltanto da intuizione ed esperienza. Allo stesso tempo è notevole il contributo apportato dalla sezione marketing con il relativo Customer Relationship Management. Il caso Nike dimostra come la personalizzazione dei servizi grazie alla *predictive analytics* porti valore aggiunto sia ai clienti che all'azienda, traducibile in un aumento dei ricavi e ad una maggiore *customer loyalty*. In più, lo studio di una strategia basata sull'elaborazione dei dati generati dai dispositivi IoT come ha fatto UPS, prevede l'ottimizzazione dei processi di logistica e trasporto, l'area del Supply Chain Management che più ha beneficiato della BDA.

Non mancano sicuramente gli aspetti critici che hanno portato numerosi studiosi a vedere sotto una luce negativa i big data e la loro elaborazione, per via della grande incertezza circa l'elaborazione concordata di una definizione univoca e rigorosa. Ma le più grandi perplessità nascono in materia di trattamento dei dati personali e di tutela della privacy, che oggi, essendo stato riconosciuto dall'UE un diritto fondamentale delle persone fisiche, gode di notevole risonanza tra i cittadini. È doveroso menzionare però che sono stati fatti molti passi in avanti in particolare in Europa attraverso l'emanazione del GDPR, direttiva che uniforma la disciplina dei dati personali in tutti i Paesi dell'Unione, soprattutto a seguito di scandali che hanno visto una violazione della privacy di milioni di persone in tutto il mondo (l'esempio Cambridge Analytica è il più eclatante). Ora le imprese sono molto più responsabilizzate nella gestione dei big data che generano i rispettivi clienti e nella potenziale condivisione con terze parti di informazioni ritenute personali.

Rimane però un certo scetticismo perché, nonostante i progressi fatti, ci sono ancora delle migliorie da apportare, come per esempio l'incoraggiamento alle imprese riguardo nuovi modelli di business, con un controllo dei dati che non spinga le società a guardare solo i propri interessi. Proprio le iniziative per migliorare l'attuale legislazione europea potrebbero essere oggetto di interessante approfondimento da parte di ricerche future.

La prova più evidente che la Big Data Analytics può essere considerata una risorsa indispensabile per l'economia globale è il focus sul settore dei trasporti aerei presentato nel quarto capitolo. In esso si osservano numerosissime applicazioni pratiche e di successo implementate negli ultimi anni, in cui ci si concentra su vari elementi della gestione aziendale. Bisogna così riconoscere che le compagnie aeree sono state protagoniste della rivoluzione digitale e sono le prime a credere fermamente, attraverso i loro top manager o i data scientist, che la strada intrapresa è quella giusta

e che ritengono l'analisi dei dati una priorità assoluta nello sviluppo delle strategie a livello corporate e del mantenimento di un vantaggio competitivo a lungo termine.

In particolare, tra le compagnie c'è chi usando la BDA migliora il proprio servizio clienti (British Airways, Delta Airlines, Vueling), chi vuole ottimizzare i processi nella SC (SouthWest Airlines & EasyJet), chi implementa nuove strategie di prezzo (EasyJet), chi si preoccupa della sicurezza degli aerei (SouthWest Airlines), e chi studia la disponibilità dei servizi (Vueling). Se si pensasse ad un'ipotetica unione di tutte le strategie che sono, nella realtà, perseguite separatamente, l'azienda oggetto di tutti questi cambiamenti diventerebbe completamente *business-driven* dal punto di vista della gestione aziendale e vedrebbe una crescita senza precedenti dei ricavi, accompagnata da una riduzione dei costi. Sarebbe la ricetta ideale per l'aumento dei margini di profitto, un elemento che le compagnie aeree puntano a migliorare dati i risultati non soddisfacenti in alcuni anni.

In conclusione, è giusto affermare che la Big Data Analytics sta assumendo un valore assoluto che viene comunemente attribuito all'oro, ma con due riserve da menzionare. La prima si scioglierà a patto che si arrivi una volta per tutte alla descrizione univoca del concetto dei big data e al suo riconoscimento globale; la seconda verrà sciolta quando sarà presente in tutti i Paesi che utilizzano la BDA una regolamentazione della privacy adatta a tutelare la diffusione impropria dei dati personali e a prevedere un loro corretto uso da parte di tutte le aziende coinvolte nel fenomeno.

Proprio come l'oro, il valore attribuito ai dati è riconosciuto universalmente. La Big Data Analytics rappresenta una nuova forma di energia, quella digitale: così come l'oro nei secoli precedenti era e tuttora è una materia prima preziosissima, un bene rifugio, allo stesso modo lungo tutto il Ventunesimo secolo la corretta analisi dei dati costituirà una fonte primaria di valore inestimabile per le organizzazioni che decideranno di operarla.

BIBLIOGRAFIA

- AGCM. (2019). Indagine conoscitiva sui big data (Rep.). AGCM, AGCOM, Garante per la protezione dei dati personali. Disponibile su https://www.agcm.it/dotcmsdoc/allegati-news/IC_Big_data_imp.pdf
- Atlantic Technologies (Ed.). (22 aprile 2020). Amazon e la rivoluzione del Big Data Analytics. Disponibile su <https://atlantic-technologies.com/it/blog/amazon-big-data-analytics/>
- Barseghian, A. (7 ottobre 2019). Council Post: How Nike Is Using Analytics To Personalize Their Customer Experience. Disponibile su <https://www.forbes.com/sites/forbestechcouncil/2019/10/07/how-nike-is-using-analytics-to-personalize-their-customer-experience/#462458461611>
- Baumann, P., & Riedel, M. (5 settembre 2019). Big Data - Definition, Importance, Examples & Tools. Disponibile su <https://www.rd-alliance.org/group/big-data-ig-data-development-ig/wiki/big-data-definition-importance-examples-tools>
- Bellini, M. (3 gennaio 2020). Big Data: Cosa sono, come utilizzarli, soluzioni ed esempi applicativi. Disponibile su <https://www.bigdata4innovation.it/big-data/big-data-analytics-data-science-e-data-scientist-soluzioni-e-skill-della-data-driven-economy/>
- Biswas, S. (18 aprile 2020). Nike – just Do it with Data science and Demand sensing. Disponibile su <https://digital.hbs.edu/platform-digit/submission/nike-just-do-it-with-data-science-and-demand-sensing/>
- Bobriakov, I. (22 novembre 2019). Top 8 Data Science Use Cases in Marketing. Disponibile su <https://medium.com/activewizards-machine-learning-company/top-8-data-science-use-cases-in-marketing-d62291ce62b6>
- Brownlow, J., Zaki, M., Neely, A., & Urmetzer, F. (2015). Data and Analytics - Data-Driven Business Models: A Blueprint for Innovation. University of Cambridge.
- Carbone, M. R. (11 febbraio 2020). Big Data Analytics, tutte le prerogative di uno strumento di successo. Disponibile su https://www.bigdata4innovation.it/big-data/big-data-analytics-tutte-le-prerogative-di-uno-strumento-di-successo/#Le_caratteristiche_del_Big_Data_le_5_V
- Chiheb, F., Boumahdi, F., & Bouarfa, H. (2019). A New Model for Integrating Big Data into Phases of Decision-Making Process. *Procedia Computer Science*, 151, 636–642. doi: 10.1016/j.procs.2019.04.085
- Darvazeh, S. S., Vanani, I. R., & Musolu, F. M. (2020). Big Data Analytics and Its Applications in Supply Chain Management. *New Trends in the Use of Artificial Intelligence for the Industry 4.0 - IntechOpen* doi: 10.5772/intechopen.89426
- De Mauro, A. (2019). *Big data analytics: analizzare e interpretare dati con il machine learning*. Milano: Apogeo.
- European Banking Authority. (2020). Report on Big Data and Advanced Analytics. EBA. Disponibile su https://eba.europa.eu/file/609786/download?token=Mwkt_BzI
- Faggioli, G. (Febbraio 2018). Big data e privacy: la protezione dei dati personali è possibile? Disponibile su https://blog.osservatori.net/it_it/big-data-e-tutela-dei-dati-personali-è-possibile
- Favaretto, M., Clercq, E. D., Schneble, C. O., & Elger, B. S. (2020). What is your definition of Big Data? Researchers' understanding of the phenomenon of the decade. *Plos One*, 15(2). doi: 10.1371/journal.pone.0228987
- Feliu, C. (2019). *Datumize*. Disponibile su https://cdn2.hubspot.net/hubfs/4029218/DAT-IC_eBook/Ebook_7_Big_Data_in_the_Airline_Industry/Big_Data_in_the_Airline_Industry_5_Case_Studies_from_the_industry_leaders

- (1).pdf?__hssc=211445770.1.1592153552859&__hstc=211445770.e33d0fcfaa236d18a9fcfb0ad9e930c3.1592153552859.1592153552859.1592153552859.1&__hsfp=3900029635&hsCtaTracking=1151ce38-f1d7-4283-a157-cbcde30f473d|2144c5f8-413d-41bc-8d99-1a7fae2045b9
- Few, S. (2018). Big data, big dupe: a little book about a big bunch of nonsense. El Dorado Hills, CA: Analytics Press.
 - GlobalData Technology. (29 gennaio 2020). Future of big data: What are the leading data trends? Disponibile su <https://www.verdict.co.uk/big-data/>
 - Hazelcast. (25 aprile 2020). What Is Data-as-a-Service (DaaS)? Disponibile su <https://hazelcast.com/glossary/data-as-a-service/>
 - J. Isaak and M. J. Hanna (Agosto 2018) “User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection,” in Computer, vol. 51, no. 8, pp. 56-59. doi: 10.1109/MC.2018.3191268.
 - Jagota, A. (30 dicembre 2019). Customer Segmentation Data Science Modeling. Disponibile su <https://towardsdatascience.com/customer-segmentation-data-science-modeling-210fb36c90bb>
 - Jeble, Shirish & Kumari, Sneha & Patil, Yogesh. (2018). Role of Big Data in Decision Making. Operations and Supply Chain Management: An International Journal. 11. 36. doi: 10.31387/oscm0300198.
 - Khvoynitskaya, S. (31 gennaio 2020). The future of big data: 5 predictions from experts for 2020-2025 Disponibile su <https://www.itransition.com/blog/the-future-of-big-data>
 - LathamDrive. (n.d.). Decision-Making at a Startup: Strategic, Tactical and Operational: Resources. Disponibile su <https://www.lathamdrive.com/resources/insights/decision-making-at-a-startup-strategic-tactical-and-operational>
 - Lindes, J. (5 giugno 2019). How to use artificial intelligence to improve data quality - PredictX. Disponibile su <https://www.predictx.com/blog/artificial-intelligence-data-quality/>
 - Marr, B. (18 ottobre 2019). What’s The Difference Between Structured, Semi-Structured And Unstructured Data? Forbes. Disponibile su <https://www.forbes.com/sites/bernardmarr/2019/10/18/whats-the-difference-between-structured-semi-structured-and-unstructured-data/#4cd22d942b4d>
 - Mayer-Schönberger, V., Cukier, K. (2013). Big Data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà. Milano: Garzanti.
 - McDaniel, S. (22 agosto 2019). Big Data for Supply Chain Management - Talend. Disponibile su <https://www.talend.com/resources/big-data-supply-chain/>
 - Nguyen, T., Zhou, L., Spiegler, V., Ieromonachou, P., & Lin, Y. (2018). Big data analytics in supply chain management: A state-of-the-art literature review. Computers & Operations Research, 98, 254–264. doi: 10.1016/j.cor.2017.07.004
 - Pearlman, S. (14 marzo 2019). Che cosa sono i big data e in che direzione si stanno muovendo - Talend. Disponibile su <https://it.talend.com/resources/future-big-data/>
 - Peterson, K. – UPS (29 gennaio 2020). UPS To Enhance Orion With Continuous Delivery Route Optimization. Disponibile su <https://www.globenewswire.com/news-release/2020/01/29/1977072/0/en/UPS-To-Enhance-Orion-With-Continuous-Delivery-Route-Optimization.html>
 - Politecnico di Milano - Osservatorio Big Data & Business Analytics (19 novembre 2019). Il mercato dei Big Data Analytics in Italia vale 1,7 miliardi di euro, 23%. Disponibile su https://www.osservatori.net/it_it/osservatori/comunicati-stampa/mercato-big-data-analytics-italia-valore-trend-comunicato

- Press, G. (9 maggio 2013). A Very Short History Of Big Data. Disponibile su <https://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data/#14b97ae465a1>
- Raj, J. (1 giugno 2019). Role of AI and Machine Learning in Data Quality. Disponibile su <https://www.intellectyx.com/blog/role-of-ai-machine-learning-in-data-quality/>
- RCN Business. (n.d.). Future Trends of the IoT, Big Data, AI Convergence. Disponibile su <https://www.rcn.com/business/insights-and-news/insights-articles/iot-big-data-ai-trends/>
- Rezzani, A. (22 aprile 2020). Come cambiano i modelli predittivi a causa del COVID-19. Disponibile su <https://www.dataskills.it/come-cambiano-i-modelli-predittivi-a-causa-del-covid19>
- Roby, K. (11 maggio 2020). UPS: How VR, IoT, AI, and big data powers the logistics company's digital transformation. Disponibile su <https://www.techrepublic.com/article/ups-how-vr-iot-ai-and-big-data-powers-the-logistics-compa>
- Saetta, B. (2 giugno 2018). Regolamento generale per la protezione dei dati: Protezione dati personali. Disponibile su <https://protezionedatipersonali.it/regolamento-generale-protezione-dati>
- SAS. (n.d.). Big Data, Bigger Marketing. Disponibile su https://www.sas.com/en_lu/insights/big-data/big-data-marketing.html
- Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, 70, 263–286. doi: 10.1016/j.jbusres.2016.08.001
- Statistical Analysis System – SAS Institute (n.d.). Intelligenza Artificiale: che cos'è e come funziona. Disponibile su https://www.sas.com/it_it/insights/analytics/what-is-artificial-intelligence.html
- Statistical Analysis System – SAS Institute (n.d.). Machine Learning: che cos'è e perché è importante. Disponibile su https://www.sas.com/it_it/insights/analytics/machine-learning.html
- van Rijmenam, M. (14 agosto 2019). How Big Data Turns CRM into Something Truly Valuable. Disponibile su <https://medium.com/dataseries/how-big-data-turns-crm-into-something-truly-valuable-c44f55482557>
- Waldmeir, P., & Hodgson, C. (8 maggio 2018). How airlines aim to use big data to boost profits. Disponibile su <https://www.ft.com/content/f3a931be-47aa-11e8-8ae9-4b5ddcca99b3>

INDICE DELLE FIGURE

Fig 1: De Mauro, A. (2019). Big data analytics: analizzare e interpretare dati con il machine learning. Milano: Apogeo.

Fig 2: Lavecchia, V. (2019). Disponibile su <https://vitolavecchia.altervista.org/che-cosa-sono-e-caratteristiche-dei-big-data-in-informatica/>

Fig 3: Todorovich, P. (30 ottobre 2018). Come usare i big data in azienda: Metodologie e strumenti di analisi. Disponibile su <https://www.zerounoweb.it/analytics/big-data/come-usare-i-big-data-in-italia/>

Fig 4: Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. Journal of Business Research, 70, 263–286. doi: 10.1016/j.jbusres.2016.08.001

Fig 5: Today (19 marzo 2018). Finti test su Facebook: vogliono i vostri dati e sarete voi a darglieli gratis. Disponibile su <http://www.today.it/tech/facebook-cambridge-analytica.html>

Fig 6: Darvazeh, S. S., Vanani, I. R., & Musolu, F. M. (2020). Big Data Analytics and Its Applications in Supply Chain Management. New Trends in the Use of Artificial Intelligence for the Industry 4.0 - IntechOpen doi: 10.5772/intechopen.89426

Fig 7: Waldmeir, P., & Hodgson, C. (8 maggio 2018). How airlines aim to use big data to boost profits. Disponibile su <https://www.ft.com/content/f3a931be-47aa-11e8-8ae9-4b5ddcca99b3>

Fig 8: Saunders, A. A. (26 gennaio 2017). How Airlines are Using Big Data. Disponibile su <https://www.exastax.com/big-data/how-airlines-are-using-big-data/>

Fig 9: Khvoynitskaya, S. (31 gennaio 2020). The future of big data: 5 predictions from experts for 2020-2025 Disponibile su <https://www.itransition.com/blog/the-future-of-big-data>

Fig 10: Hazelcast. (25 aprile 2020). What Is Data-as-a-Service (DaaS)? Disponibile su <https://hazelcast.com/glossary/data-as-a-service/>

Fig 11: Rezzani, A. (2020, April 22). Come cambiano i modelli predittivi a causa del COVID-19. Disponibile su <https://www.dataskills.it/come-cambiano-i-modelli-predittivi-a-causa-del-covid19>