



UNIVERSITÀ POLITECNICA DELLE MARCHE

Facoltà di Ingegneria

CORSO DI LAUREA IN INGEGNERIA GESTIONALE

---

**TECNICHE DI BIG DATA ANALYTICS PER LA  
GESTIONE DELLE FILIERE LOGISTICO  
PRODUTTIVE**

**BIG DATA ANALYTICS TECHNIQUES FOR THE  
MANAGEMENT OF PRODUCTION CHAINS**

Relatore:

*Ch.mo Professore Maurizio Bevilacqua*

Tesi di Laurea di:

*Chiara Scardera*

---

Anno Accademico 2019-2020



# Ringraziamenti

Ringrazio il Professor Maurizio Bevilacqua e l'Università Politecnica delle Marche per la disponibilità e per avermi accompagnato in questo percorso. Ringrazio i miei genitori, perché i loro sacrifici hanno permesso il mio cammino, per avermi accompagnato e consigliato, per aver da sempre stimolato la mia curiosità verso il mondo. Ringrazio i miei fratelli Giuseppe e Beatrice per avermi insegnato la condivisione e per avermi sopportato in questo ultimo periodo. Ringrazio i miei nonni e la mia famiglia tutta per l'affetto e per avermi dato la forza di non mollare. Ringrazio Nonna Rina per l'amore che ha sempre dato alla sua famiglia attraverso ogni suo singolo gesto. Ringrazio le mie amiche di sempre per essere state sempre un sostegno e una certezza della mia vita, so che potrò sempre contare su di voi.

# Indice

<b>Introduzione</b>	<b>1</b>
<b>1 Big Data</b>	<b>4</b>
1.1 Analisi delle caratteristiche . . . . .	5
1.2 Evoluzione . . . . .	7
1.3 Perché sono importanti . . . . .	9
1.4 Problematiche e sfide . . . . .	10
<b>2 Supply Chain e Supply Chain Management</b>	<b>12</b>
2.1 Evoluzione della Supply Chain . . . . .	13
2.2 Supply Chain Management . . . . .	16
2.3 Organizzazione della Supply Chain . . . . .	18
2.4 Modello SCOR . . . . .	19
2.5 Tecnologie . . . . .	21
<b>3 Applicazione della Big Data Analytics nel supply chain management</b>	<b>27</b>
3.1 BDA e gestione delle relazioni con i fornitori . . . . .	28
3.2 BDA e progettazione rete della supply chain . . . . .	29
3.3 BDA e ricerca e sviluppo dei prodotti . . . . .	30

3.4	BDA e pianificazione della domanda . . . . .	32
3.5	BDA e la gestione degli approvvigionamenti . . . . .	32
3.6	BDA e gestione del magazzino . . . . .	33
3.7	BDA e logistica . . . . .	34
<b>4</b>	<b>Analisi di dati</b>	<b>36</b>
4.1	Evoluzione della Big Data Analytics . . . . .	36
4.2	Tipologie di dati . . . . .	37
4.3	Tecniche di analisi nella supply chain . . . . .	39
4.3.1	Analisi statistica . . . . .	39
4.3.2	Simulazione . . . . .	40
4.3.3	Ottimizzazione . . . . .	42
4.4	Data Mining . . . . .	42
4.5	Visualizzazione . . . . .	47
4.6	Machine Learning . . . . .	47
<b>5</b>	<b>Tecniche di supporto alla Big Data Analytics</b>	<b>49</b>
	<b>Conclusioni</b>	<b>52</b>
	<b>Bibliografia</b>	<b>54</b>

# Elenco delle figure

2.1	Gestione e integrazione del supply chain management . . . .	18
3.1	Integrazione dei dati all'interno della supply chain . . . . .	35

# Introduzione

L'argomento che si considera in questo lavoro è quello dei Big Data, del loro impatto in vari ambiti pratici, soffermandoci in modo particolare sulle applicazioni e sulle conseguenze in campo aziendale. Le tecniche statistiche per analizzare i Big Data costituiscono uno sviluppo originale, un'estensione e un ampliamento delle tecniche statistiche. I metodi impiegati, infatti, sono molto complessi e fanno parte di un mondo in continua e veloce evoluzione; non verrà svolta un'analisi dettagliata, perché richiede conoscenze approfondite in campo statistico e informatico ma un excursus e una rappresentazione delle principali tecniche di analisi dei Big Data. La trattazione dell'argomento parte con la definizione dei Big Data, delle loro principali caratteristiche e della rilevanza che assumono all'interno della filiera logistico-produttiva. Verrà quindi definito in primo luogo l'obiettivo che la presenza dei Big Data e la loro corretta analisi permette di raggiungere che è quello di estrarre informazioni aggiuntive rispetto a quelle che ottenibili da piccole serie di dati. Successivamente si parlerà di Supply Chain e Supply Chain Management. Tutte le imprese fanno parte di una catena di approvvigionamento, dalle materie prime fino al consumatore finale. La quota parte di supply chain che deve essere direttamente gestita dipende da diversi fattori tra cui la complessità del prodotto, il numero dei fornitori disponibili, la disponibilità di

materie prime. Il supply chain management permette la gestione della filiera produttiva, nella quale diverse imprese operano come un'unità commercialmente integrata, verso i fornitori di materie prime e verso i clienti, gestendo in comune le fasi di acquisizione delle materie prime (pagamento ai fornitori, approvvigionamento), della loro lavorazione e distribuzione (fatturazione ai clienti, gestione ordini), nonché della pianificazione dei processi interni alla catena del valore. Si passerà poi ad elencare le tecnologie necessarie affinché possa i dati possano essere integrati e utilizzati nella filiera andando a considerare sia le tecnologie di base e semplici che quelle più recenti e sofisticate che devono il loro sviluppo a Internet. Nella parte centrale dell'elaborato si parlerà di come e di quali tecniche di analisi big data sono usate nelle diverse fasi della catena produttiva, dalla progettazione del prodotto e della linea, alla scelta dei fornitori e dei partner, fino alla gestione della logistica e della distribuzione. Il fulcro del lavoro è invece l'analisi dei Big Data; verrà innanzitutto definita l'evoluzione delle tecniche di analisi determinata dallo sviluppo delle tecnologie e dell'ambiente economico. Gli strumenti di Big Data Analytics sono dei software, come ad esempio database e software utili per acquisire ed elaborare informazioni oppure applicativi dedicati per specifici processi aziendali. Le tecniche e i metodi di analisi invece si basano sulle diverse tipologie di analisi dei dati che possono essere effettuate. Il processo di analisi permette di operare un'analisi predittiva, ovvero permette di conoscere anticipatamente cosa accadrà: ciò diventa possibile poiché, se si hanno a disposizione un modello e dati storici a sufficienza, si può determinare una tendenza future con basi o fondamenti statistici. Sulla base di queste previsioni è possibile poi intervenire sul futuro mediante un'analisi



prescrittiva, ovvero si vanno a cercare le condizioni affinché un certo evento accada.

# Capitolo 1

## Big Data

Quando si parla di Big Data ci si riferisce a una vasta mole di informazioni che è possibile ricavare dall'ambiente esterno. Non esiste una definizione univoca riguardo i Big Data; le diverse definizioni, infatti, variano a seconda dell'ambiente che si sta analizzando, dello scopo dell'analisi e dei dati che si hanno a disposizione. Alcune definizioni pongono l'attenzione sulle elevate dimensioni di Big Data (Boubeta-Puig et al., 2012), altre invece sottolineano la loro complessità (Chen et al., 2012) e la loro eterogeneità (Liu, 2013), altre infine evidenziano le possibilità che il loro utilizzo può aprire per conoscere gli ambienti in cui le aziende si muovono (Lee, Mandick, Wang, Wang, and Zang, 2014). Le definizioni più recenti considerano che oggi la trasmissione dei dati avviene soprattutto attraverso dati "non-strutturati", mentre le prime descrizioni di Big Data prendevano in considerazione solo i dati strutturati. La differenziazione tra dati strutturati e non è una delle tante classificazioni che i Big Data possono subire; i dati strutturati sono organizzati secondo schemi e tabelle rigide e sono indicati per i modelli di gestione relazionale delle informazioni, mentre i dati non strutturati, ovvero

immagini, testi, file audio e video, non hanno uno schema e per essere utilizzati e analizzati devono essere processati e convertiti. Esiste poi un terzo tipo di dati, ovvero quelli semi-strutturati, che hanno caratteristiche comuni sia ai dati strutturati che non. L'importanza dei Big Data è esplosa con lo sviluppo delle nuove tecnologie, in particolare con la diffusione dei social media e del fenomeno dell'Internet of Things (IOT), che ha aumentato sia le possibilità di raccolta dei dati che le fonti da cui ricavarli, oltre che la velocità nella loro creazione e le possibilità di utilizzo. Parallelamente alla crescente importanza, si è estesa la quantità di dati raccolti, cui però non è seguito un adeguato miglioramento della capacità di immagazzinare i dati raccolti.

## 1.1 Analisi delle caratteristiche

Secondo la letteratura è possibile analizzare i Big Data secondo tre dimensioni principali, a cui con il tempo e con il progresso tecnologico si sono aggiunte delle altre, ovvero: volume, velocità e varietà. La prima dimensione si riferisce alla quantità di dati che un'organizzazione o un individuo è in grado di creare e raccogliere. Ogni secondo social media, e-commerce, sensori, flussi fisici, informativi e monetari generano un elevato volume di dati strutturati, ma soprattutto non strutturati, come immagini, audio e video, che devono essere immagazzinati e processati. Tale quantità di dati non può essere processata elaborata dai sistemi convenzionali di gestione dei dati a causa della limitata capacità di questi ultimi, per cui è necessaria l'adozione di sistemi informatici ad hoc che possano garantire non solo un completo immagazzinamento dei dati ma anche una corretta elaborazione. Per velo-

cità si intende la velocità alla quale i dati sono generati e processati. Grazie al progressivo miglioramento delle tecnologie che permettono di trattare i dati, essa è cresciuta nel tempo fino a consentire la possibilità di processare in real-time i dati. L'ultima dimensione prende in considerazione il numero delle tipologie di dati che possono essere creati. Tale dimensione è proporzionale all'avanzamento delle tecnologie; maggiore, infatti, è la possibilità di creare dati, più ampia sarà la *varietà* di dati che è possibile raccogliere. Naturalmente, l'analisi di dati non strutturati è più difficoltosa in quanto questi dati, a causa della loro diversità e disomogeneità, devono essere rielaborati e riadattati per poter essere analizzati in modo uniforme. Alle tre dimensioni descritte è possibile aggiungerne altre in modo da poter visualizzare meglio le opportunità che l'analisi dei Big Data può apportare a chi li utilizza. Le altre dimensioni sono: veracità, valore, variabilità e deperimento. Con *veracità* si intende l'inaffidabilità e l'incertezza delle fonti che rendono necessario un controllo sui dati per poterli rendere un valido strumento da utilizzare nei vari processi a cui sono destinati. Inaffidabilità e incertezza sono dovute a incompletezza, latenza, inaccuratezza e soggettività dei dati raccolti; ad esempio, le impressioni dei consumatori sono inaffidabili perché sono legate alle opinioni umane che non possono essere valutate in maniera oggettiva. Ad ogni modo, per poter trattare anche i dati affetti da incertezza, sono stati sviluppati delle tecniche e degli strumenti statistici che valutano i dati in intervalli di sicurezza. Un'altra dimensione è quella del *valore* che permette di evidenziare l'importanza dell'utilizzo dei Big Data per aumentare i ricavi, diminuire i costi e prendere le decisioni migliori per le aziende. L'analisi dei dati permette quindi di aumentare il valore degli stessi in relazione alla loro

utilità e alla loro importanza nella strategia dell'impresa. La *variabilità* si riferisce alle variazioni, spesso imprevedibili, che il flusso di dati può subire e che devono essere registrate attraverso software. Legata alla variabilità, vi è la *complessità* ovvero la diversità delle fonti da cui è possibile ricavare i dati e che rende difficile il processo di raccolta, mantenimento e studio dei dati. È possibile ridurre la complessità attraverso piattaforme uniformi e raccolta dei dati in real-time. L'ultima dimensione è quella che si riferisce al *deperimento* dei dati e del loro valore nel tempo. Poiché i dati vengono raccolti in tempo reale, l'intervallo di tempo in cui devono essere utilizzati è molto importante; basti pensare ai dati raccolti in ambito medico nel quale l'immediato studio dei dati è fondamentale per poter salvare le vite dei pazienti. Ogni dimensione è legata e interagisce con le altre attraverso un legame di proporzionalità diretto o indiretto. Ad esempio, l'aumento della dimensione della velocità ha effetti diretti sia sul volume che sulla varietà dei dati, ma anche sulle altre dimensioni (aumenta il valore dei dati, la loro complessità e variabilità, ma diminuisce la veracità).

## 1.2 Evoluzione

Il termine Big Data è stato utilizzato in letteratura per la prima volta nel 1997 da Michael Cox e David Ellsworth; con esso si intendeva il fenomeno per cui i dati erano troppo ingombranti per poter essere contenuti nei tradizionali database. In ogni caso, l'idea di collezionare grandi quantità di dati per agevolare l'organizzazione dei servizi o per migliorare processi produttivi nasce quando si inizia a parlare di Business Intelligence, con l'obiettivo di ottenere vantaggio competitivo sui competitor collezionando e analizzan-

do in maniera strutturata informazioni rilevanti per il business. Nel 1965 viene creato il primo data center in USA, e soltanto pochi anni dopo uno sviluppatore crea il primo framework per un database relazionale, ovvero un magazzino virtuale in grado di contenere tabelle di dati collegate tra loro. Ma è con la nascita di Internet nel 1991 che i dati diventano accessibili a tutti e ovunque nel mondo e la loro crescita diventa esponenziale. I primi dati che circolavano sul web subito dopo l'avvento di Internet erano quelli relativi all'e-commerce, poiché i principali fruitori e frequentatori del web erano principalmente le aziende, mentre il contributo degli utenti era marginale. Dal 2004, quando è esploso il fenomeno dei social media, il volume dei dati ha subito un'impennata, poiché milioni di utenti hanno potuto far circolare immagini, video e file audio. Ciò ha provocato anche un cambiamento del tipo di dati scambiati in rete; oltre ai tradizionali dati strutturati vengono scambiati anche quelli non strutturati. I social media sono una fonte molto utilizzata dalle aziende per raccogliere informazioni sugli utenti attraverso lo studio delle pagine visitate, delle preferenze e dei gusti dei consumatori, della permanenza sulle pagine, delle posizioni registrate. Negli ultimi cinque anni si è assistito ad un ulteriore cambiamento provocato dalla nascita del fenomeno dell'Internet of Things che si riferisce ad un ambiente in cui ogni suo oggetto può generare, raccogliere e scambiare dati con un altro oggetto o con l'esterno. Con la rapida crescita dell'IoT, i dispositivi e i sensori sorpasseranno i social media e i siti di e-commerce come fonte primaria di big data.

### 1.3 Perchè sono importanti

I Big Data offrono un valido aiuto alle aziende nel creare nuovi business, nello sviluppare nuovi prodotti e servizi e nel migliorare i processi già esistenti. Avvalersi dell'analisi dei Big Data crea molteplici benefici, innanzitutto da un punto di vista dei costi che le imprese devono sostenere. Osservare in tempo reale il processo produttivo e vedere immediatamente eventuali problemi riduce i tempi di reazione e permette di intervenire più rapidamente per poter correggere le problematiche. Una efficiente analisi dei dati, inoltre, consente di effettuare migliori previsioni di domanda, di tracciare in modo più preciso i beni durante il trasporto e di gestire in maniera ottimale la distribuzione dei prodotti. Un altro esempio di riduzione dei costi può avvenire attraverso la supervisione degli impianti produttivi attraverso il controllo dei consumi rapportati ai tempi di funzionamento effettivi. L'analisi dei Big Data inoltre consente di osservare il mercato in cui le aziende si muovono e guidarle al suo interno con maggiore consapevolezza e conoscenza. Le aziende infatti possono raccogliere informazioni sulle preferenze dei consumatori e carpire le opinioni sui prodotti sul mercato. Per fare questo gli strumenti più immediati e diffusi sono i social media perché fotografano in modo univoco e diretto il mercato rendendo il processo di raccolta dati più snello rispetto a una tradizionale ricerca di mercato. In questo modo le aziende possono raccogliere informazioni riguardo anche a problematiche e impressioni sui prodotti, migliorando la loro offerta.

## 1.4 Problematiche e sfide

L'utilizzo di Big Data all'interno dell'organizzazione è una pratica recente che ha di fronte alcune sfide da fronteggiare. Alcune di queste riguardano la qualità e la sicurezza dei dati. È importante che i dati raccolti rispettino delle specifiche valutazioni per poter essere utilizzati nel prendere decisioni per l'azienda. La qualità dei dati deve essere garantita in ogni fase del processo di analisi per non alterare i risultati degli algoritmi e per essere un valido supporto alle aziende. I dati non strutturati e quelli raccolti da fonti diverse sono quelli che tendono più velocemente a decadere e non assicurano una affidabile validità. Dati di non alta qualità non vengono usati dalle aziende per non inficiare le prestazioni e per non prendere le decisioni sbagliate. È necessario infatti un controllo sui dati che pervengono alle aziende, il quale deve avvenire sulla base di parametri fissi e deve valutare la qualità dei dati, riparare dati errati e garantire un compromesso tra i costi, proporzionale alla loro qualità e i guadagni effettivi derivanti dal loro utilizzo. Un'altra fondamentale sfida da affrontare è quella di garantire la sicurezza dei dati degli utenti, ovvero la protezione delle informazioni e dei sistemi informativi da accessi e usi non autorizzati, da modifiche illegittime in modo da assicurare confidenzialità, integrità e affidabilità a chi fornisce i dati alle aziende. Nel contesto dell'analisi dei Big Data la sicurezza dei dati è importante sia nella fase di raccolta e conservazione dei dati che in quella del loro utilizzo attraverso gli algoritmi che dovranno analizzarli. Un rischio legato all'utilizzo dei dati potrebbe essere quello di sovrastimare i dati e affidarsi completamente a loro. Avere una larga disponibilità di dati comporta il rischio di considerare anche quelli non appropriati e non adatti



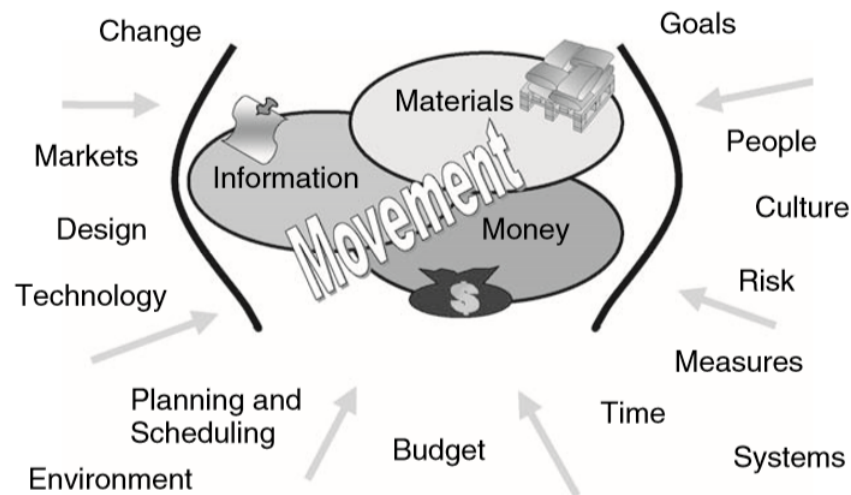
restituendo un'analisi fuorviante e distante dai reali obiettivi.

## Capitolo 2

# Supply Chain e Supply Chain Management

La Supply Chain può essere qualificata come “un insieme selezionato, duraturo di entità autonome e indipendenti sotto il profilo proprietario, ma che sono accomunate dall’operare insieme attraverso l’integrazione di alcuni processi aziendali affinché sia possibile rendere disponibili i prodotti, i servizi e le informazioni che aggiungono valore per i clienti, a partire dai consumatori finali risalendo fino ai produttori di materie prime”. Essa prevede il flusso di materiali, accompagnato da un flusso di informazioni e capitali; tali flussi sono sia equiparati, quindi nessun flusso è più importante di un altro, che interconnessi tra loro e non possono esistere senza gli altri, per cui una carenza di uno dei tre flussi si riflette anche sugli altri due andando a inficiare le prestazioni della filiera. Il giusto bilanciamento sulle quantità dei flussi è difficile da trovare a causa dei diversi fattori che intervengono nel ciclo produttivo come ad esempio i cambiamenti dell’ambiente, i finanziamenti e il budget a disposizione dell’azienda, fino ad arrivare alla situazione socio-

politica del mercato che può spostare la domanda del mercato e le necessità dei clienti. Ogni attività, infatti, deve essere allineata agli obiettivi finanziari; in altre parole, lo scopo della supply chain non è tenere traccia delle conseguenze di una decisione circoscritta ad un ambito su tutte le attività, ma deve essere anche quello di valutare le conseguenze di ogni decisione sui ricavi e i costi. Un altro aspetto da sottolineare è che nonostante il processo produttivo sia composto da molti step e fasi consequenziali, il cliente vede il processo come unico e il fallimento di una fase, di una operazione si riflette sull'intero processo.



**Chart 1.4** The Forces Affecting the Resources That Define a Supply Chain

## 2.1 Evoluzione della Supply Chain

Fino ai primi anni '70 il focus delle aziende era sull'efficienza del lavoro interno intesa in termini di quantità più che sulla qualità dell'organizzazione. Infatti, si pensava che la chiave del successo di una impresa fosse nella mo-

le di prodotti realizzati senza considerare i costi generati; efficienza voleva dire quantità. I metodi che permettevano di pianificare i livelli produttivi erano basati sull'EOQ (Economic Order Quantity, che calcolava la quantità ottima di prodotti da reintegrare a magazzino sulla base dei costi di immagazzinamento e di emissione degli ordini), oppure sul MRP (Material Requirements Planning, che a partire dal piano principale di produzione e dall'esplosione della distinta base permetteva di conoscere nel dettaglio tempistiche e quantitativi delle varie componenti dei prodotti) fino ad arrivare all'ERP (Enterprise Resources Planning, che integrava le informazioni rilevanti per l'azienda). In ogni caso l'impresa era focalizzata verso l'interno e i miglioramenti si cercavano all'interno dell'organizzazione. Dagli anni '80, con la nascita di nuove filosofie produttive come il Just In Time, si è passati a considerare meno il lavoro come cardine dell'efficienza produttiva e a porre l'attenzione sui materiali e sull'efficienza delle macchine, anche se il focus dell'organizzazione era ancora interno. Nel decennio successivo, con l'avvento di Internet, delle possibilità di scambio veloce di informazioni e della possibilità di connettersi e interagire con l'esterno con più facilità, si è compresa l'opportunità di considerare l'influenza dei fattori esterni come strumento per riadattare i processi produttivi, migliorare la loro efficienza e aumentare il valore dei prodotti. Il mondo esterno da mero destinatario dei prodotti diventa una fonte da cui è possibile ricavare e scambiare informazioni e dati e la sfida che le aziende hanno intrapreso e sostenuto in quegli anni è stata quella di trovare un equilibrio che permettesse di riorganizzare i processi produttivi in base ai cambiamenti dell'ambiente circostante. Nel corso dell'ultimi due decenni si è assistito ad un rinnovato e crescente

interesse attorno ai problemi legati alla gestione della filiera produttiva al fine di fornire un servizio che potesse rispondere alle complesse e diverse esigenze del consumatore odierno che hanno richiesto una crescente efficienza, flessibilità e capacità innovativa alle aziende. I cambiamenti sono dovuti ovviamente allo sviluppo delle tecnologie che permettono lo scambio diretto e immediato delle informazioni sia all'interno dell'azienda che verso gli altri attori del mercato in modo che l'integrazione delle informazioni sui diversi aspetti dei processi aziendali sia economica, immediata e di conseguenza anche efficiente. Un altro fattore è il cambiamento della concezione che i clienti hanno dei prodotti che utilizzano e richiedono. I consumatori infatti domandano prodotti sempre più personalizzati, a prezzi bassi e disponibili in tempi brevi e ciò ha provocato una diminuzione del ciclo di vita del prodotto a cui le aziende si sono dovute adattare fornendo risposte rapide al cliente attraverso grandi investimenti per rendere il processo produttivo più snello e flessibile. La nuova concezione di supply chain va nella direzione di vedere la singola azienda come parte di una rete di imprese e attività tutte strettamente connesse tra loro e la cui cooperazione e collaborazione deve fornire al cliente finale prodotti, servizi e informazioni che creano valore. Il successo di un'impresa deriva quindi, non solo dalla propria capacità di generare valore, ma anche dall'efficienza del sistema con cui essa si interfaccia e dalla competitività e affidabilità degli altri attori, come fornitori e clienti, presenti nel sistema.

## 2.2 Supply Chain Management

Il supply chain management è una disciplina o “filosofia che permette di gestire l’intero flusso di materiali e informazioni dai fornitori ai clienti finali” (Cooper-Ellram, 1990). Questa prima definizione di supply chain management mette in evidenza la continuità tra le fasi di un ciclo produttivo e la necessità di supervisionarlo dall’inizio alla fine per poter garantire una corretta allocazione delle risorse. Come per i Big Data, non è possibile definire in modo univoco il concetto di supply chain management perchè in letteratura esistono moltissime definizioni che differiscono l’una dall’altra per il concetto che si è voluto mettere in evidenza riguardo le funzionalità e l’essenza del SCM. Alcuni studiosi hanno voluto porre l’accento su come aumentare il vantaggio competitivo utilizzando competenze e tecnologia al servizio di una gestione integrata del processo produttivo che unisce sia i partner esterni che interni alla filiera (Tan- Kannan, 1998). Una definizione più improntata a descrivere un processo produttivo industriale infine definisce la gestione della filiera produttiva come un “insieme di approcci utilizzati per integrare in maniera efficiente fornitori, produttori, magazzini, in modo che la merce sia prodotta nelle giuste quantità, nei giusti luoghi e tempi in modo da minimizzare i costi globali soddisfacendo al tempo stesso i requisiti in termini di livello di servizio” (Simchi-Levi, 2000). Nonostante le diverse sfumature di significato che le definizioni descrivono, l’elemento che le accomuna è il passaggio da una visione dell’azienda concentrata e chiusa su se stessa a una aperta e proiettata verso l’esterno. I concetti di interazione e integrazione sono alla base dello sviluppo di questa disciplina e di un nuovo modo di vedere le aziende, interconnesse su diversi livelli con i loro partners,

ovvero con tutti quei soggetti non interni alla filiera produttiva ma facenti parte di un progetto integrato con la catena produttiva. L'integrazione deve andare oltre la condivisione delle informazioni e della conoscenza tra i diversi attori della filiera, ovvero tra fornitori, produttori, distributori e clienti, i quali devono essere messi nella condizione di scambiarsi informazioni riferite alla domanda, alla situazione di magazzino, ai piani produttivi, ai piani promozione, alle previsioni e al programma di spedizione, essendo altresì in grado di coordinare la gestione delle scorte; essa deve infatti completarsi con lo scambio di informazioni che deve avvenire tra le diverse imprese del mercato, tra l'impresa e i fornitori e i clienti. L'obiettivo è infatti quello di garantire la presenza e l'efficienza di canali comunicativi che proiettino l'azienda verso l'esterno e le permettano di superare i suoi confini. La possibilità di condividere informazioni in modo trasversale rispetto a tutti i livelli dell'azienda fa in modo che si possa parlare di integrazione verticale. In questo modo le informazioni viaggiano velocemente tra un attore e l'altro della catena produttiva rendendo tutti i soggetti coscienti e consapevoli del proprio ruolo e partecipi all'interno della rete produttiva. Un altro beneficio è la riduzione dei tempi che intercorrono tra l'acquisizione delle informazioni e le decisioni, poiché le informazioni vengono raccolte contestualmente al processo. I vantaggi dell'integrazione verticale si riflettono in un controllo da parte dell'azienda su ogni fase del processo produttivo facilitando ad esempio la programmazione della produzione per poter rispondere a eventuali cambiamenti della domanda oppure riducendo i costi che possono derivare sia dalle giacenze in magazzino, sia dai tempi di non valore aggiunto che non creano valore ma rallentano la produzione e sia dalla non

corretta comunicazione con i clienti riguardo le richieste.

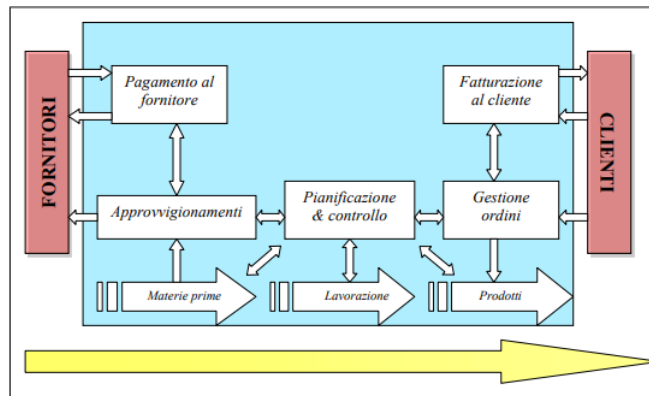


Figura 2-1: *Gestione ed integrazione del Supply Chain Management*

**Figura 2.1:** Gestione e integrazione del supply chain management

## 2.3 Organizzazione della Supply Chain

Le attività chiave svolte all'interno della catena produttiva sono:

- gestione delle relazioni con il cliente: identifica il cliente chiave in funzione dei bisogni e delle loro esigenze. I vari segmenti individuati vengono divisi in gruppi target al fine di comprenderne le esigenze ridurre la variabilità della domanda;
- gestione dei servizi resi al cliente: definisce modalità e tempi di consegna dei prodotti; le informazioni sono raccolte in tempo reale e in stretta collaborazione con le funzioni di produzione e logistica;
- gestione della domanda: vengono bilanciate le richieste del singolo cliente con le capacità e le competenze che l'azienda è in grado di fornire. Voluti tempi e localizzazione della domanda sono coordinate



con le attività di approvvigionamento, produzione e distribuzione al fine di rispondere al meglio alle richieste del cliente;

- gestione del flusso fisico lungo il processo produttivo: la produzione può seguire le richieste del mercato in un ambiente di tipo pull oppure essere avviata a prescindere dagli ordini. Il primo scenario garantisce una maggiore flessibilità e costi di produzione minori;
- gestione dei rapporti con i fornitori: sviluppare e definire una rete di relazioni a lungo termine con un numero ristretto e selezionato di fornitori nell'ottica di reciproci vantaggi;
- processo di ricerca e sviluppo: mira a immettere sul mercato nuovi prodotti allo scopo di raggiungere un più alto vantaggio competitivo; fondamentale in questa fase è la conoscenza del mercato e la possibilità di avere un time to market, ovvero il tempo che intercorre tra ideazione del prodotto e la commercializzazione, basso;
- gestione dei flussi di ritorno: ha come obiettivo pianificazione, implementazione e controllo di efficienza dei flussi di materie prime e prodotti, e dei correlati flussi informativi.

## 2.4 Modello SCOR

Il modello SCOR (Supply Chain Operations Reference) è uno strumento gestionale che permette la gestione della catena produttiva dai fornitori ai clienti in quanto definisce quattro attività principali in ogni supply chain –source, make, deliver, return- a cui si aggiunge una quinta -plan- che ac-

compagna ogni singola attività. Il modello inoltre definisce tre livelli di dettaglio: top, nel quale le aziende stabiliscono i loro obiettivi strategici in base alle loro aree di azione, configuration, nella quale le aziende configurano la loro supply chain sulla base dei loro obiettivi strategici, e infine process element, in corrispondenza del quale si definiscono le particolari operazioni della catena produttiva, gli input e gli output. Le diverse attività si attuano in diversi orizzonti temporali a seconda del livello di dettaglio.

- plan: processi che esaminano domanda ed offerta per definire una linea d'azione che soddisfi i requisiti di approvvigionamento, produzione e consegna;
- source: processi che procurano beni e servizi per soddisfare la domanda pianificata e quella effettiva;
- make: processi che trasformano il prodotto in uno stato finito per soddisfare la domanda pianificata e quella effettiva;
- deliver: processi che forniscono beni e servizi per soddisfare la domanda pianificata e quella effettiva, compresa la gestione degli ordini, il trasporto e la distribuzione;
- return: processi inerenti alla restituzione dei prodotti o il reclamo sui servizi per qualunque motivo, compreso il supporto cliente successivo alla consegna.

## 2.5 Tecnologie

Come già detto, l'introduzione della tecnologia e in particolar modo di Internet ha modificato in maniera radicale la gestione della catena produttiva, garantendo un significativo miglioramento della produttività, ma ha anche rappresentato e rappresenta tutt'ora un valido strumento per la valutazione e di supporto alle decisioni in settori strategici quali la previsione della domanda, la pianificazione integrata della produzione e del processo di approvvigionamento, lo sviluppo di prodotti aderenti con le richieste dei clienti e infine la gestione di ordini e trasporto per poter rispondere in modo rapido alle richieste dei clienti. Le tecnologie utilizzate dalle imprese, secondo una ricerca dell'University of Illinois at Chicago's Center for Supply Chain Management and Logistics, vanno da quelle più mature e affermate come sensori, software di ottimizzazione, cloud computing, magazzini automatizzati e data warehouse a quelle emergenti ma in rapidissimo sviluppo come la stampa 3D, l'Internet of Things e il Machine Learning. Inoltre, di fondamentale importanza sono i social media e l'analisi dei Big Data.

- cloud computing: è una tecnologia che permette di elaborare, archiviare e memorizzare dati grazie all'utilizzo di hardware e software distribuiti all'interno della rete. I software sono stati classificati in SaaS (software as a service), PaaS (platform as a service) e IaaS (infrastructure as a service). Per SaaS s'intende un servizio attraverso il quale il cliente avrà accesso a un'interfaccia di utilizzo del software via browser, senza la necessità installare programmi applicativi o avere macchine con determinate caratteristiche di performance, abbattendo in modo sostanziale i costi sotto tre punti di vista:

- investimento hardware;
- acquisto licenze software;
- upgrade software;

Tra i vantaggi del Software as a Service c'è senz'altro la possibilità di utilizzare il software ovunque ci si trovi e su qualsiasi dispositivo. Il PaaS permette a un fornitore esterno di mettere a disposizione l'infrastruttura, sia hardware sia software in modo che l'azienda o l'ente possa dedicarsi unicamente allo sviluppo di un'applicazione o di un programma. In PaaS è possibile creare e sviluppare qualunque programma e applicativo. Alcune delle funzioni che questo servizio offre sono, ad esempio: sistema operativo, ambiente di sviluppo, server, hosting, accesso alla rete, archiviazione. I vantaggi per le aziende che decidono di ricorrere a questo sistema sono, essenzialmente, tre:

- riduzione dei costi: il costo di implementazione delle infrastrutture hardware e software necessarie per lo sviluppo dell'applicativo va totalmente a carico di chi offre il servizio;
- migliore gestione delle risorse, derivante dalla non necessità di allocare risorse interne per queste attività specifiche dal momento che software e hardware sono demandati all'esterno;
- flessibilità, è possibile creare delle piattaforme assolutamente personalizzate unicamente in base alle esigenze dell'azienda o dell'ente.

L'IaaS permette all'azienda di avere a disposizione tutte le funzioni dell'infrastruttura hardware di cui necessitano senza doversi sobbar-

care gli oneri derivanti dalla gestione ed è quindi alla base del servizio di cloud. I benefici del cloud computing includono bassi costi di implementazione e tempi più veloci per creare valore; inoltre i risparmi hanno permesso di agevolare e rafforzare la condivisione trasversale di informazioni riguardanti il livello di inventario, la capacità e le informazioni sulla logistica tra i vari attori della filiera produttiva. Ad esempio, i dati che riguardano la movimentazione dei flussi fisici vengono caricati nel cloud in modo da essere visibili da vari soggetti e in ogni livello della catena che possono in questo modo monitorare in tempo reale il flusso dei materiali, quindi da e verso quali nodi le materie prime o i prodotti si stanno spostando, quali sono i livelli di inventario, se vi sono degli stockout o se il flusso si è bloccato in qualche sua fase;

- data warehouse: l'ammontare dei dati generati e raccolti cresce in modo esponenziale grazie alla generazione da parte di sistemi automatizzati di dati real-time, di conseguenza ricorrere a sistemi standard di archiviazione dei dati risulta essere poco pratico. Per questo sono stati introdotti all'interno della filiera produttiva e della logistica dei magazzini virtuali di dati che conservano le informazioni e i dati parallelamente alle varie fasi del processo. Si sono resi necessari dei computer più veloci che hanno accelerato l'uso di soluzioni di questo tipo nell'ambito della supply chain e della logistica e ciò ha trainato l'utilizzo di Big Data raccolti e analizzati contestualmente ai processi. Questa tecnologia riduce il tempo di processamento e aumenta la flessibilità di raccolta e analisi dei dati, ma può causare problemi nel trasferimento

di grandi quantità di dati oltre a richiedere una buona compatibilità tra le piattaforme di warehouse e i software che analizzano i dati. La virtualizzazione dei dati inoltre permette di immagazzinare anche dati non strutturati come immagini, video e documenti;

- sensori: le tecnologie che utilizzano i sensori sono state recentemente riconosciute per la loro importanza e il loro potenziale nel panorama della supply chain in quanto possono estendere il controllo dell'azienda sui prodotti, possono monitorare la movimentazione degli ordini e lo spostamento dei sistemi di movimentazione all'interno degli impianti. Essi, infatti, sono molto utilizzati per prevedere, gestire e ottimizzare i tempi degli spostamenti dei materiali o degli ordini all'interno dell'impianto produttivo in quanto permettono di governare gli spostamenti dei veicoli automatizzati oppure per controllare le giacenze del magazzino ed effettuare gli ordini. Nell'ambito del picking l'utilizzo di sensori e codici permette di effettuare anche in maniera automatizzata gli ordini attraverso il riconoscimento dei diversi articoli in base alle liste in cui sono riportati i vari codici relativi ad ogni ordine. Prevedere i ritardi delle consegne permette ai gestori dei nodi logistici di dare notifica immediata ai clienti dei ritardi, di registrare i riordini di magazzino e di schedare le operazioni di cross-docking. Questa soluzione combina i dati dei sensori con quelli derivanti dall'organizzazione della logistica e della catena produttiva per garantire la riduzione dei tempi di trasporto e di viaggio e una pianificazione della distribuzione degli ordini che tende a eliminare le discrepanze rispetto alla situazione reale. La tecnologia maggiormente usata è quella

attraverso i codici RDIF, ovvero una tecnologia utile per l'identificazione e memorizzazione automatica di informazioni inerenti a oggetti, animali o persone basata sulla capacità di memorizzazione di dati da parte di particolari etichette elettroniche (tag) e sulla capacità di questi di rispondere all'interrogazione a distanza da parte di appositi apparati fissi o portatili (reader). L'identificazione avviene mediante radiofrequenza, grazie alla quale un reader è in grado di comunicare e aggiornare le informazioni contenute nei tag che sta interrogando;

- social media: vengono utilizzati per ottenere informazioni da diverse fonti per avere una visione d'insieme del mercato e per prendere decisioni in modo più accurato. Attraverso i social media le aziende possono aumentare la comunicazione con i clienti, carpire i loro orientamenti e le loro richieste, generare nuova domanda, orientare i clienti, mitigare i rischi, aumentare la produttività e di conseguenza i ricavi. Data la massiccia presenza sui social dei consumatori, dei fornitori ma anche dei competitors una eventuale assenza dell'azienda le comporterebbe enormi svantaggi;
- IoT: l'applicazione dell'IoT impatta significativamente sulla gestione della SC, riducendone i costi, aumentando l'accuratezza dell'inventario e permettendo la tracciabilità della merce. Supply chain e IoT insieme garantiscono un monitoraggio migliore nella capacità di rilevazione, soprattutto tramite sensori che utilizzano classici tag a radiofrequenza o RFID, di tutte le merci in transito lungo la filiera. Dal loro ingresso nello stabilimento, ad esempio, dentro un pallet, al disimballo con tecniche di picking fino al confezionamento in vista dell'uscita alla volta

dei canali commerciali, ogni spostamento è rigorosamente tracciato così da evitare perdite di materiale o inefficienze.



## Capitolo 3

# Applicazione della Big Data Analytics nel supply chain management

Nell'ambito della produzione, la vasta quantità di dati è generata sia dai canali esterni, tramite le informazioni e i dati provenienti da fornitori, clienti e mercato, che da quelli interni attraverso i diversi macchinari e i sensori posizionati all'interno dell'impianto. Usare i Big Data per rendere più snella la produzione porta dei benefici anche al settore della distribuzione e delle vendite.

### 3.1 BDA e gestione delle relazioni con i fornitori

Le relazioni con i fornitori devono essere stabilite nell'ambito della pianificazione strategica e sono un fattore chiave per creare nuovo valore e ridurre i rischi di fallimento. Usare le tecniche di analisi Big Data permette di raccogliere informazioni accurate riguardo i modelli di spesa in modo da aiutare a stipulare i migliori contratti di collaborazione con i fornitori. Le informazioni che possono essere raccolte riguardano il ROI (Return on investment che indica la redditività e l'efficienza economica dell'investimento) dei potenziali fornitori oppure riguardano la conoscenza della struttura dell'azienda partner per capire la sua capacità ad adattarsi a eventuali cambiamenti futuri. Una tecnica che permette di valutare i fornitori e permetterne la scelta è quella che si basa sull'utilizzo di modelli di **fuzzy synthetic evaluation** e di **AHP** (processi di analisi gerarchica). Questi modelli sono dei metodi di supporto alle decisioni nel caso in cui a disposizione del decisore vi siano informazioni conflittuali e complesse; essi quindi consentono di ottenere una soluzione di compromesso in modo trasparente. I metodi di analisi multicriterio supportano il decisore nella fase di organizzazione e sintesi di informazioni complesse e spesso di natura eterogenea. Tale metodologia permette al decisore di analizzare e valutare diverse alternative, monitorandone l'impatto sui differenti attori del processo decisionale. L'AHP consente di assegnare delle priorità ad una serie di alternative decisionali, mettendo in relazione valutazioni di tipo qualitativo e quantitativo, altrimenti non direttamente confrontabili, e combinando scale multidimensionali di misure

in una singola scala di priorità. Secondo la logica fuzzy ai parametri vengono assegnati dei valori di verità normalizzati a uno; le valutazioni logiche vengono effettuate con questi valori e successivamente si riportano i risultati di nuovo sottoforma di variabili continue. I modelli fuzzy sono robusti e tolleranti al rumore, gestiscono anche informazioni imprecise, non usano complesse formule matematiche ma semplici regole logiche. Gli indici fuzzy permettono di catturare anche le sfumature e le incertezze delle variabili qualitative e per questo contengono molte informazioni.

## **3.2 BDA e progettazione rete della supply chain**

La progettazione della supply chain è una decisione strategica che include tutte le decisioni riguardanti la selezione dei partner della filiera e definisce le politiche e le strategie di lungo termine dell'azienda. È importante in questa fase tener conto della soddisfazione del cliente e dell'efficienza della filiera. Lo scopo della fase di progettazione della supply chain è definire una rete di membri che possano incontrarsi e convergere sugli stessi obiettivi di lungo termine dell'azienda stessa e per fare ciò è necessario:

- definire gli obiettivi strategici di lungo termine
- definire lo scopo del progetto
- determinare le forme di analisi e i relativi strumenti da adottare
- realizzare il miglior progetto possibile.

Uno studio condotto nel 2016 denominato "Distribution network design with big data: model and analysis" propone come modello da adottare per

localizzare i centri di distribuzione di un'azienda un modello di **programmazione non lineare intera mista**. Il problema che si voleva risolvere era quello di identificare il corretto numero di centri di distribuzione da aprire e la corretta assegnazione del numero di clienti da poter soddisfare avendo come obiettivo la minimizzazione dei costi e come vincoli la domanda da dover soddisfare, i costi del trasporto e le date di consegna degli ordini. Oltre ai dati già disponibili, gli studiosi si avvalsero anche dei dati random generati dai datasets e dei dati provenienti da strumenti di marketing intelligence.

### **3.3 BDA e ricerca e sviluppo dei prodotti**

Uno dei maggiori obiettivi della produzione manifatturiera è quello di assicurarsi che i prodotti realizzati siano conformi alle preferenze e ai gusti dei consumatori, i quali cambiano con facilità e a ritmi sempre più veloci andando a ridurre il ciclo di vita del prodotto stesso. Per questo è necessario per le aziende avere degli strumenti che le permettano di predire e misurare le aspettative e le preferenze dei prodotti. La mancanza di tali informazioni può rappresentare un'importante problematica in quanto le aziende non conoscerebbero a fondo l'ambiente e i clienti con cui devono interagire, non potrebbero così rispondere alle esigenze del mercato e potrebbero essere superate dalle concorrenti. Gli strumenti da utilizzare devono garantire un monitoraggio continuo dei comportamenti e degli orientamenti dei consumatori e per fare ciò vengono create grandi quantità di dati, appunto Big Data. Il processo che permette di trasformare i bisogni dei consumatori nelle specifiche di progetto è quello di ingegnerizzazione che utilizza i dati, le informazioni ricavabili dal mondo reale ma trasformate in dati compres-

bili e utilizzabili per il processo di decision-making. Il recente sviluppo delle tecniche e degli strumenti di analisi dei dati hanno aperto nuove possibilità per generare informazioni e realizzare un prodotto qualitativamente adatto per le esigenze del mercato. Anche se diversi software di analisi sono stati sviluppati proprio per estrarre informazioni dall'ambiente produttivo e dal mercato, avvalersi di questi metodi e strumenti è ancora a una fase preliminare e i designer devono ancora fronteggiare diverse sfide come ad esempio la difficoltà di scegliere il miglior strumento di analisi. L'utilizzo dei big data inoltre consente di modificare il processo di ingegnerizzazione del prodotto in quanto consente:

- sviluppo del prodotto informato: le aziende conoscono non solo come i consumatori utilizzano il prodotto ma anche quali potrebbero essere i problemi legati ad esso e le problematiche che dovrebbero essere risolte. A tal proposito è molto importante l'IoT poiché permette uno scambio di informazioni diretto tra prodotti e azienda;
- ingegnerizzazione potenziata: la migliore relazione che si instaura tra prodotti, consumatori e aziende offre la possibilità alle aziende di aumentare la competitività e la qualità dei loro prodotti secondo le aspettative dei consumatori. I big data permettono di raccogliere elevate quantità di informazioni dai consumatori, molte di più di quelle che è possibile raccogliere tramite i tradizionali canali di vendita. Tali informazioni saranno poi a disposizione dalle aziende per garantire ai loro clienti un migliore servizio;
- sviluppo del prodotto più veloce: maggiori quantità di informazioni, soprattutto se già conservate nei Cloud, velocizzano e rendono più

economica la fase di ingegnerizzazione.

### **3.4 BDA e pianificazione della domanda**

Molte aziende sono interessate a conoscere le previsioni di domanda e a pianificare la produzione attraverso i big data. Avere consapevolezza della lealtà dei consumatori, conoscere gli orientamenti del mercato e settare il prezzo ottimale rappresentano una sfida per le aziende, ma anche un obiettivo da raggiungere e per fare ciò si devono avvalere di avanzati sistemi software e di algoritmi. I modelli di previsione della domanda possono essere di tipo qualitativo o quantitativo; i primi sono basati su opinioni e giudizi di esperti o su ricerche di mercato, mentre i modelli quantitativi si basano su dati certi e modelli matematici, come ad esempio i modelli basati su serie storiche o su indicatori economici. La big data analytics permette di identificare i trend dei nuovi mercati e determinare alla radice le cause di problematiche, fallimenti e difetti.

### **3.5 BDA e la gestione degli approvvigionamenti**

L'approvvigionamento di materiali, manodopera e tutto ciò che permette la produzione consiste in una serie di attività di stipula di contratti che deve avvenire nella fase di pianificazione strategica. I dati che vengono generati sono ingenti e richiedono sistemi avanzati di gestione e immagazzinamento, oltre che professionalità che siano in grado di analizzare tali dati ed estrarre

da essi informazioni che siano di grande rilevanza per prendere le decisioni. Oggi questi dati, siano essi generati internamente o esternamente, sono disponibili in real-time alle organizzazioni che possono utilizzarli al meglio.

### 3.6 BDA e gestione del magazzino

Utilizzare sistemi di immagazzinamento capaci raccogliere informazioni aumenta l'efficienza delle operazioni e di conseguenza la redditività dell'intera catena produttiva. Un tale tipo di sistema è capace di registrare le quantità già presenti in magazzino e quelle necessarie, minimizzare l'inventario sulla base di dati storici di consumo e controllare spese attraverso una verifica del magazzino e dei prodotti a giacenza. I principali benefici dell'utilizzo di tecniche di analisi dei big data sono:

- miglioramento dell'efficienza delle operazioni: grazie alla possibilità di effettuare controlli continui sui magazzini è possibile conoscerne la reale situazione, apportare delle modifiche ed evitare ritardi o penurie;
- massimizzare vendite e ricavi: seguire il flusso fisico dei materiali o delle richieste di servizio consente di leggere le logiche che governano le richieste e di capire in anticipo come si sta muovendo la domanda e in quale direzione sta andando. Il corretto tempismo di analisi dei dati è la chiave per effettuare investimenti mirati e tempestivi e che sono in grado di portare un vantaggio economico e competitivo all'azienda;
- aumentare la soddisfazione del cliente: strettamente collegato alla massimizzazione delle vendite, vi è la capacità di poter soddisfare e capire i gusti dei consumatori a cui deve seguire l'abilità di rispondere in modo

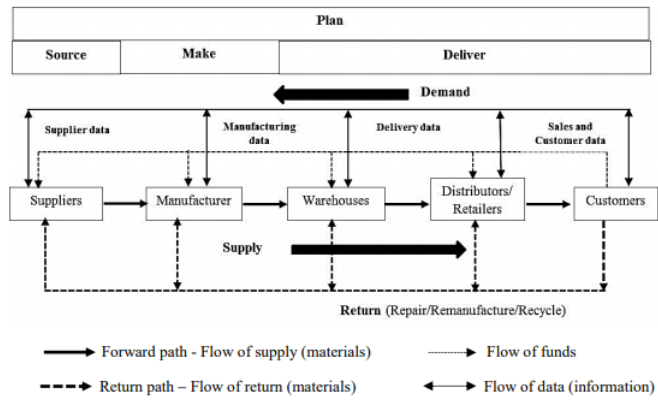
rapido ed efficiente alle richieste dei clienti. Per fare ciò è opportuno che il livello dei magazzini sia in grado di rispondere correttamente alla domanda di prodotti;

- ridurre i costi: utilizzare dei software cloud, come ad esempio il software SaaS, permette di ridurre i costi di mantenimento e lo spazio occupato sugli hardware.

### **3.7 BDA e logistica**

Il compito fondamentale dei dati in questa fase della supply chain è quello di permettere la comunicazione e la collaborazione tra i diversi settori, ovvero tra fornitori, trasportatori, azienda, distributori e venditori. Oggi, a causa dell'alto volume di dati generati da diverse fonti come ad esempio sensori, GPS e riferimenti RFID, oltre che dall'interazione di altre tecnologie tutte legate ad Internet, sono necessarie delle tecniche molto potenti per analizzare tempestivamente i dati raccolti. Le aziende hanno bisogno di piattaforme che siano capaci di immagazzinare una vasta quantità di dati provenienti da sistemi interni ed esterni, di garantire la gestione e la validazione di essi per aumentarne l'affidabilità e poterli quindi metterli a disposizione del business infine di analizzarli. L'analisi segue un approccio pervasivo, che quindi integra dati provenienti da sistemi e devices differenti e di diversa natura. Un'altra caratteristica delle piattaforme è quella di avere un'organizzazione orizzontale che si può adattare alle diverse esigenze dei settori e può quindi immagazzinare dati diversi tra loro.





**Figure 1.** Data driven general supply chain structure

**Figura 3.1:** Integrazione dei dati all'interno della supply chain

# Capitolo 4

## Analisi di dati

Per big data analytics si intende l'utilizzo combinato di matematica e statistica per essere in grado di analizzare grandi quantità di dati e dar loro un maggiore valore; tale valore è tanto maggiore quanto è più alta la possibilità di ricavare informazioni e la rilevanza che tali informazioni assumono all'interno dei processi per cui sono destinati. Ciò è possibile grazie alle enormi capacità di calcolo dei sistemi informatici disponibili a un costo minore rispetto al passato.

### 4.1 Evoluzione della Big Data Analytics

Già a partire dagli anni Sessanta le tecnologie disponibili consentivano la raccolta dei dati riguardanti i processi aziendali su supporti magnetici e le analisi che potevano essere svolte erano di tipo statistico, limitate a estrazioni dei dati a consuntivo e su base aggregata. Con l'avvento dei data base relazionali e del linguaggio SQL, l'analisi dei dati assume una maggiore dinamicità poiché è possibile estrarre i dati sia in modo aggregato che con

maggior livello di dettaglio. Si prendono in considerazione i dati operazionali, ovvero quelli generati dall'attività operativa dell'azienda; nel caso in cui le attività possano essere gestite su un solo livello si può utilizzare il software ERP, un sistema di gestione che integra tutti i processi di business rilevanti di un'azienda (vendite, acquisti, gestione magazzino, contabilità ecc.). Spesso però i dati si presentano in formati diversi tra loro, poiché provenienti da fonti, livelli e attività diversi tra loro; per questo motivo a partire dagli anni Novanta sono stati creati dei database disegnati appositamente per le analisi e che consentivano di integrare i dati provenienti dai vari sistemi operazionali. Nasce così il data warehouse, un database che contiene dati integrati, consistenti e certificati relativi a tutti i processi di business dell'azienda e che costituisce il punto di partenza per le attività analitiche del sistema di Business Intelligence. La Business Intelligence è un sistema di modelli, metodi, processi e strumenti che rendono possibile la raccolta regolare e organizzata del patrimonio dati generati da un'azienda. La Big Data Analytics va oltre e permette di raccogliere ed elaborare i dati provenienti dai consumatori e dai fornitori in tempo reale per acquisire vantaggio sul mercato.

## 4.2 Tipologie di dati

La Big Data Analytics è costituita da 3 tipologie di analisi interpretative:

- **Analisi descrittiva:** permette di rispondere alla domanda “che cosa è successo?”. Per essa si intende la trasformazione dei dati grezzi in una forma che li renda facili da capire e interpretare, riorganizzare, ordinare, e manipolare per generare informazioni utili. Viene quindi

utilizzata per visualizzare i dati nel giusto contesto, identificare le informazioni rilevanti e valutare la qualità dei dati, oltre che riconoscere ipotesi e limiti di quanto ricavato. Essa è utile perché consente di studiare i dati storici, quindi quelli del passato, e di capire a partire da essi come potrebbero influenzare i risultati futuri, senza mai stabilire le relazioni di causa ed effetto tra dati e situazione effettiva, per questo motivo l'interazione con questi tipi di dati è semplice immediata e di conseguenza di larga diffusione. L'analisi descrittiva viene utilizzata quando si ha bisogno di capire a livello aggregato cosa succede nell'azienda in modo da poter redigere report aziendali riguardo la situazione finanziaria, le vendite, le operazioni effettuate e la produzione utilizzando dati storici.

- **Analisi predittiva:** permette di rispondere alla domanda “cosa potrebbe accadere in futuro?”. Essa consente quindi di eseguire l'analisi dei dati storici con lo scopo di prevedere gli eventi e le tendenze futuri attraverso la costruzione di un modello matematico in grado di rilevare la macro-tendenza futura. L'analisi predittiva va oltre quella descrittiva, che si fermava alla semplice osservazione dei dati passati che non erano utilizzati e analizzati a fondo, e tenta sia di scoprire le cause degli eventi e i fenomeni che di scovare informazioni che non sono ancora note. Tale tipo di analisi viene utilizzata per prevedere i modelli di acquisto, i comportamenti dei consumatori e, in ambito della supply chain, la domanda dei consumatori, i livelli di inventario e le operazioni future. Le principali tecniche che consentono di svolgere una analisi predittiva sono il Data Mining e il Machine Learning.

- **Analisi prescrittiva:** permette di rispondere alla domanda “come si dovrà rispondere a quegli eventi futuri?”. Essa guida le decisioni dell’impresa basate sia sull’analisi descrittiva che predittiva e combinate con simulazioni, modelli matematici e tecniche di decision-making. Lo scopo di questo tipo di analisi è quello di affiancare il decisore e guidarlo nella scelta della strategia giusta da adottare per l’impresa; per tale motivo l’analisi sarà tanto più efficiente se il decisore è in grado di porre le giuste domande ad esempio sul tipo di offerta proporre al segmento a cui l’azienda si rivolge e di conseguenza quale dovrà essere le modalità di distribuzione dei prodotti.

## 4.3 Tecniche di analisi nella supply chain

### 4.3.1 Analisi statistica

L’analisi statistica consiste in due tipi di analisi: descrittiva e inferenziale. Nella prima i dati passati sono usati per descrivere le caratteristiche di un fenomeno che vengono riassunte all’interno di una tabella o all’interno di grafi. Tale tipo di analisi viene utilizzata soprattutto nell’analisi descrittiva in quanto le relazioni tra i dati sono già note e l’analisi si ferma alla pura conoscenza dei dati e della situazione passata e corrente del fenomeno e non si cercano le relazioni future. Attraverso l’analisi statistica inferenziale invece si cerca di individuare le proprietà del fenomeno per predire il comportamento dello stesso nel futuro. L’analisi statistica viene utilizzata quando vi è incertezza nella raccolta dei dati, come ad esempio nella distribuzione, nell’inventario e nell’analisi del rischio e in particolare per monitorare il flusso

di materiali lungo la filiera produttiva. Dati il volume, la varietà, la veracità e la velocità dei Big Data, la supply chain ha bisogno di tecniche di analisi robuste e semplici e che siano aggiornate rispetto a quelle tradizionali; un esempio di algoritmo utilizzato è quello di analisi parallela che utilizza specifici metodi come quello del gradiente coniugato, dei minimi quadrati o del Mann-Whitney U testing per modellizzare e comparare le densità di probabilità e le distribuzioni quadratiche dei dati. Il test di Mann-Whitney è un test non parametrico, cioè che può essere applicato anche quando non si conosce la distribuzione dei dati, mediante il quale si può verificare se la provenienza di campioni indipendenti ma continui diversi è la stessa.

### 4.3.2 Simulazione

Le imprese ricorrono a strumenti di simulazione per ottimizzare la fase di progettazione sia del prodotto che della linea produttiva allo scopo di diminuire i costi e rendere questa fase più snella e produttiva. Ricorrere alla simulazione prima di realizzare un prodotto significa poterlo visualizzare in digitale e apportare con facilità delle modifiche al disegno già nelle prime fasi, eliminando così il rischio di avviare la produzione di un prodotto dalle caratteristiche fisiche errate. Per quanto riguarda invece la progettazione della linea, utilizzare la simulazione attraverso l'analisi "what-if", ovvero mandando in esecuzione la linea virtualizzata e analizzando i risultati, permette di verificare la presenza di eventuali errori prima che essi possano verificarsi durante la produzione rallentando o bloccando la linea; in questo modo si potrà modificare e migliorare la linea prima che sia stato effettuato l'investimento oppure si potranno valutare diverse alternative tra le configu-

razioni della linea progettate. La modellazione e le tecniche di simulazione dovrebbero essere utilizzate per sviluppare dei progetti che siano guidati da grandi quantità di dati che permettono di conoscere l'ambiente in cui il prodotto dovrà affacciarsi. Un modello di analisi di sensibilità può aiutare a identificare le cause della performance; ad esempio, in un impianto produttivo il tempo perso a causa della rottura di un macchinario può essere variato e il suo impatto sul tempo ciclo di produzione può essere esaminato per quantificare i tempi e trovare il miglior compromesso per diminuire il tempo di lavorazione. L'analisi di sensibilità può essere condotta anche per far variare i diversi parametri che condizionano la filiera produttiva, come ad esempio le variazioni di domanda, i costi di trasporto, il numero di clienti, la capacità produttiva, l'efficienza del trasporto, e le politiche operazionali che impattano sulle performance. Per analisi di sensibilità si intende la possibilità di valutare gli effetti sui risultati forniti dal modello, cioè la funzione matematica che descrive analiticamente il sistema fisico, sia esso un prodotto o una linea produttiva, in seguito a variazioni delle variabili di ingresso. Tale analisi può essere di due tipi: per scenari futuri, se uno scenario rappresenta una tra le possibili combinazioni di valori assunti dalle variabili indipendenti, o "what if", in quanto si va a valutare cosa cambia se cambiano i valori assunti dai parametri decisionali. La simulazione può essere applicata ad ogni livello di analisi, sia descrittivo, che predittivo o prescrittivo oppure può essere utilizzata come generatore di dati, soprattutto nel caso in cui risulta difficile raccogliere e conservare grandi quantità di dati, oppure come validatore del modello. Modelli di simulazione reali possono essere utilizzati come fabbriche virtuali di dati che generano dati nelle

stesse misure e formati di quelli già a disposizione per avere una maggiore flessibilità e una maggior livello di dettaglio nelle simulazioni stesse.

### **4.3.3 Ottimizzazione**

Le tecniche di ottimizzazione consentono di estrarre le logiche interne e le conoscenze da una vasta quantità di dati generati dai complessi sistemi e modelli di rappresentazione di linee produttive, prodotti o qualsiasi sistema si voglia rappresentare. L'obiettivo, che può consistere nella riduzione dei costi o nel soddisfacimento della domanda, deve essere raggiunto considerando i vincoli del sistema, come ad esempio la capacità produttiva o la disponibilità delle risorse. Inoltre, nel caso in cui tali tecniche venissero applicate in una supply chain in cui collaborano diversi soggetti indipendenti tra loro/esterni, esse permettono l'organizzazione tra i diversi soggetti, la gestione accurate dell'intera catena e il controllo della performance al fine di permettere ad ogni soggetto di raggiungere i propri obiettivi.

## **4.4 Data Mining**

Il data mining è il processo di scoperta delle relazioni, dei pattern e delle informazioni, non conoscibili a priori, all'interno di grandi basi di dati. Lo scopo è quello di indagare le cause e gli schemi ricorrenti all'interno dei dati raccolti per dare un ordine alle informazioni contenute nei dati e aumentare la conoscenza del fenomeno, del processo che si vuole studiare. Gli obiettivi di predizione e descrizione possono essere raggiunti usando diversi metodi di data mining illustrati di seguito.



- **Classificazione:** utile per apprendere un modello che mappa un elemento in base alle sue caratteristiche in una dell'insieme di classi predefinite;
- **Regressione:** calcola la relazione stimata tra una variabile dipendente e una o più variabili esplicative. Con l'analisi di regressione, è possibile definire la relazione tra le variabili scelte e prevedere i valori in base al modello. Questo metodo è usato nel predire la domanda dei consumatori per un nuovo prodotto in relazione all'investimento in pubblicità;
- **Clustering:** usato per identificare oggetti simili l'uno all'altro; il grado di similitudine tra oggetti diversi può essere deciso in base alle caratteristiche possedute dai dati. L'effetto del clustering è quello di suddividere i dati in un insieme finito di raggruppamenti che ne rappresenta il modello. La similarità può essere calcolata basandosi su fattori quali il comportamento esibito negli acquisti o la reattività a certe azioni in modo da individuare sottopopolazioni omogenee tra i consumatori. Le informazioni ottenute dal raggruppamento sono utili per capire meglio consumatori e fornire quindi servizi maggiormente personalizzati;
- **Riepilogazione:** include metodi che individuano una descrizione compatta dei dati analizzati o di un loro sottoinsieme. Vengono così fornite informazioni sintetiche sui dati. Le tecniche di riepilogazione sono spesso applicate all'analisi dei dati interattiva e esplorativa e nella generazione automatica dei report;

- **Analisi dei dati storici:** consente di applicare le tecniche di regressione ai dati storici, utilizzando anche variabili temporali; sono molto utili nell'analisi predittiva;
- **Modellazione delle dipendenze:** consiste nel trovare un modello che descriva significative dipendenze tra variabili. I modelli di dipendenza esistono a due livelli: quello strutturale specifica (spesso in forma grafica) quali variabili sono localmente dipendenti tra loro, e il livello quantitativo del modello specifica le forze delle dipendenze usando qualche scala numerica.

I metodi e gli strumenti per il data mining sono diversi e possono cambiare e integrarsi a seconda delle esigenze; in ogni caso i principali sono:

- **rete neurale:** è un programma ispirato alla struttura del cervello umano in cui è possibile riconoscere molti nodi (neuroni) connessi da relazioni di causalità (sinapsi) e nel quale le informazioni viaggiano in parallelo.

I vantaggi principali sono:

- forte interconnessione;
- parallelismo tra i nodi;
- grande memoria di archiviazione;
- architettura non programmata, ma capace di evolversi con l'esperienza, ampliando la propria capacità di risolvere determinate tipologie di problemi.

In una rete neurale, partendo dalle unità di input, a cui vengono forniti i dati del problema da risolvere, la computazione si propaga in parallelo nella rete fino alle unità di output, che forniscono il risultato.

Una rete neurale ad apprendimento supervisionato viene addestrata fornendo una certa quantità di input (problemi) ed output (soluzioni), in modo che rilevando le associazioni apprenda a produrre autonomamente output corretti. Una rete neurale ad apprendimento non supervisionato invece viene addestrata soltanto con input costituiti da selezionate tipologie di dati. Esaminando i medesimi, la rete apprende a cogliere analogie e differenze, operando classificazioni. Grazie all'elevata capacità di calcolo parallelo, queste 2 categorie di reti neurali sono in grado di processare in modo proficuo ed efficiente i big data, effettuando classificazioni, associazioni e clustering.

- albero di decisione: è grafo in cui partendo dalla radice si procede ad effettuare una classificazione tramite un percorso che è di volta in volta una scelta fra varie diramazioni, o sottoinsiemi (nodi), i cui rami sono le alternative che conducono alle diverse foglie (risultati o classi). Un albero decisionale implementato in modo corretto deve avere dimensioni adeguate, il che significa non eccessive: un numero troppo alto di variabili renderebbe caotico e lento un algoritmo che è invece veloce ed efficiente. Nel data mining, gli alberi di decisione sono utilizzati per operazioni di segmentazione, classificazione, regressione e serie storiche;
- Naive Bayes: è un algoritmo di apprendimento supervisionato che si avvale del teorema di Bayes che descrive la probabilità di un elemento di appartenere o meno a una certa classe o la probabilità di un evento di verificarsi in funzione dei suoi attributi o delle condizioni di partenza correlate all'evento. La peculiarità è che ogni caratteristica

viene valutata in autonomia: le probabilità di un attributo di risultare attinente a una classe (e quindi il suo contributo alla classificazione dell'elemento o alla verifica dell'evento) non dipendono dalla relazione con altri attributi. Questa tipologia di classificatori è piuttosto robusta nei confronti del rumore (ossia dati errati, inutili, etc) e riesce ad essere efficiente anche in presenza di dati incompleti.

Il processo del Data Mining in realtà è solo la fase centrale di un processo il cui nome corretto è Knowledge Discovery in Databases (KDD) e che consiste di tre fasi:

- **Data preprocessing:** si occupa di preparare i dati originali per poterli processare automaticamente nella fase di data mining. Importante in questa fase è comprendere il dominio applicativo dei dati in modo da conoscere i dati grezzi posseduti e le loro correlazioni. Ciò permette di capire le necessità informative del fruitore finale del processo e di identificare il tipo di analisi da attuare e in base ad essa determinare il formato dei dati utile all'elaborazione: l'obiettivo è produrre un dataset target, cioè un insieme di dati ottenuti dai dati originari ma idoneo alla fase successiva. In generale i dati hanno bisogno di essere bonificati (data cleaning) perché possono presentare errori derivanti anche dalle misurazioni e dalle osservazioni fatte oltretutto da operazioni ordinarie su di essi; questo significa procedere a correzioni, rimozioni o modifiche di dati incompleti o errati;
- **Data processing:** è il processo vero e proprio, quindi il Data Mining; in questa fase si sceglie quale metodo di analisi è il più adatto in base all'obiettivo da raggiungere e quale algoritmo applicare;

- Data postprocessing: fornisce gli strumenti per poter analizzare la conoscenza ottenuta. L'informazione che si ottiene è data dall'output degli algoritmi di data mining usati nella fase centrale; tale informazione deve essere interpretata e spiegata per renderla comprensibile nel caso in cui il fruitore sia una persona oppure per trasformarla in un formato adatto se il fruitore finale è un dispositivo. Quando le informazioni sono destinate alle persone nel processo vengono utilizzate tecniche di visualizzazione.

### Il processo di KDD (1)

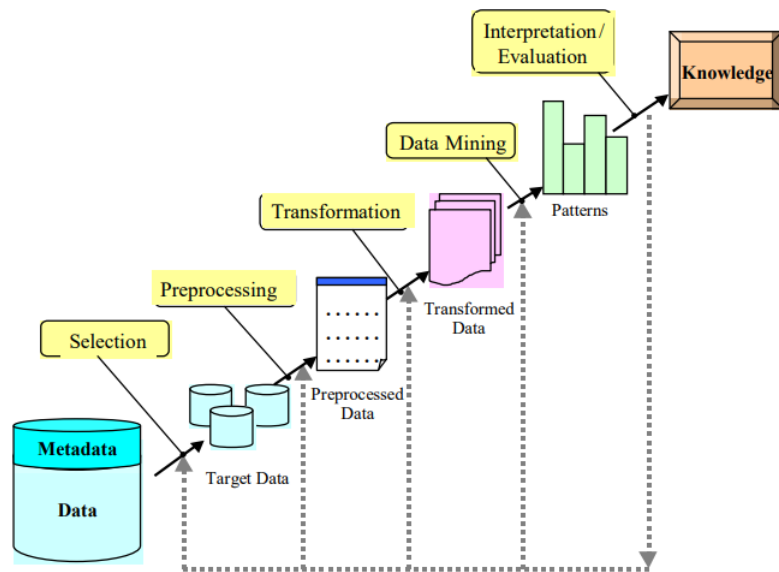


Figura 4.1: Il processo KDD

## 4.5 Visualizzazione

Questa tecnica viene utilizzata per rendere i dati più comprensibili utilizzando mappe di colore, che permettono di evidenziare la densità di avvenimento di un evento attraverso l'intensità dei colori, flowchart, tabelle e diagrammi

che valutano e rappresentano flussi fisici, informativi e performance. I dati quindi vengono tradotti in immagini, diagrammi e animazioni per un facile interpretazione e una facile comunicazione.

## 4.6 Machine Learning

Il machine learning permette ai computer di imparare dall'esperienza, ovvero di migliorare le prestazioni del programma dopo lo svolgimento di un compito o il completamento di un'azione. Gli algoritmi di Machine Learning, ovvero algoritmi di autoapprendimento basati sui dati in input, vengono utilizzati quando non ci sono strategie predefinite o non si conosce il modello di risoluzione del problema; la logica è quella per la quale i sistemi possono imparare dai dati, sviluppando una propria logica utilizzando algoritmi generici. Dal punto di vista informatico, quindi, non si dice alla macchina cosa deve fare, ma le viene fornito un set di informazioni che, mediante un algoritmo generico, le consentono di imparare a svolgere compiti e attività richieste con un intervento umano ridotto al minimo.

## Capitolo 5

# Tecniche di supporto alla Big Data Analytics

La crescente domanda e importanza dell'analisi dei dati hanno richiesto l'esistenza di piattaforme, software e programmi adatti a contenere ed elaborare una vasta e complessa quantità di dati. Fra i principali strumenti utilizzati in ambito di Big Data Analytics ci sono:

- MapReduce: Questa struttura fu pensata per fornire un modello di elaborazione parallela e implementazione distribuita per processare grandi quantità di dati. Attraverso MapReduce, le richieste di dati vengono divise e distribuite su nodi paralleli e processate in modo distribuito (la fase Map). I risultati dell'elaborazione vengono poi raccolti e consegnati (fase Reduce). Il framework ebbe molto successo, così che l'algoritmo MapReduce fornì l'ispirazione per la creazione di nuove soluzioni. Ad esempio, una implementazione del framework fu adottata da un progetto open source di Apache, Hadoop, che ancora

oggi è lo strumento più diffuso per la Big Data Analytics. Attualmente la strada maggiormente intrapresa per implementare soluzioni di Big Data Analytics è l'uso di architetture a livelli multipli, cioè di architetture parallele e distribuite in cui i dati vengono disposti su più unità di elaborazione. Il calcolo parallelo processa i dati molto più velocemente, consentendo una maggiore rapidità di ottenimento dei risultati. In questo tipo di architettura i dati sono gestiti in DBMS distribuiti, che lavorano parallelamente implementando l'uso di framework MapReduce come Hadoop. L'utente finale, attraverso un front-end application server, tiene traccia del processo di analisi grazie agli strumenti forniti dal framework.

- Hadoop: è una struttura per diversi componenti software che consente di scomporre i compiti di calcolo (jobs) in processi in più parti, di distribuirli su diversi nodi di un cluster e lasciarli così lavorare parallelamente. Nelle architetture Hadoop più grandi sono utilizzati migliaia di singoli computer. In questo modo si ha il vantaggio che ogni computer del cluster deve mettere a disposizione solo una piccola parte delle risorse hardware necessarie. Il lavoro con grandi quantità di dati non richiede un computer di calcolo dalle prestazioni elevate, ma si può svolgere grazie a svariati server standard che ottimizzano i costi.
  - flessibilità ad immagazzinare i dati indipendentemente dal fatto che siano strutturati, semi-strutturati o non strutturati. L'immagazzinamento non è vincolato dal tipo di dati;
  - eccelle nel processare i dati di natura complessa e di elevate dimensioni;



- economico e scalabile rispetto ai sistemi di immagazzinamento dei dati tradizionali;
- tollerante al guasto, così se per qualche motivo un nodo dovesse smettere di funzionare non è un problema perché i dati verrebbero processati da un altro nodo.

# Conclusioni

In questo elaborato sono stati trattati vari argomenti riguardanti le tecniche di analisi di dati volte a rendere i dati originati da e verso le aziende profittevoli da un punto di vista economico e di efficienza. Fondamentale è stato riconoscere quanto la forza dei Big Data ha impattato sulla conoscenza del mondo e, in particolare, in ambito economico sulla gestione delle filiere logistico produttive. Le imprese che hanno già messo in atto progetti di data analysis e data management lo hanno fatto per migliorare l'engagement con il cliente, incrementare le vendite, ridurre il time to market, ampliare l'offerta di nuovi prodotti e servizi, ottimizzare l'offerta attuale al fine di aumentare i margini, ridurre i costi e identificare nuovi mercati. Ovviamente la possibilità di raccogliere e analizzare i dati nel modo corretto è stata data dall'evoluzione delle tecnologie, dalla crescente importanza di Internet e dal progresso informatico che ha fatto sì che si potessero raccogliere e utilizzare informazioni di una crescente rilevanza per il business. Di pari passo all'evoluzione tecnologica si sono affinate le tecniche e le modalità di analisi dei dati; si è passati infatti da metodi basati prevalentemente sull'analisi statistica e sullo studio di modelli matematici a tecniche che utilizzano software informatici e algoritmi iterativi quasi autonomi che integrano dati provenienti da diverse fonti. Un'importante spinta al cambiamento e una

nuova risorsa per migliorare il business e la conoscenza del mondo è rappresentata dall'IoT. Raccogliere ed analizzare tutti i dati provenienti dalla sensoristica di produzione e trarne le informazioni per il proprio business attraverso gli strumenti di business analytics è la nuova chiave della competitività. Essere più competitivi oggi significa essere in grado di analizzare tutte le variabili di processo come energia consumata, temperature e pressioni, velocità e tutte le grandezze fisiche misurabili all'interno del ciclo produttivo. L'Industrial Internet of Things (IIoT) è in grado di mettere in comunicazione ogni linea di processo, ogni fase produttiva, ogni singola macchina grazie a diverse tipologie di sensoristica (RFID, accelerometri, termometri) posizionate nei punti sensibili e critici. Grazie ai dati generati dagli oggetti e trasmessi al sistema, è possibile ottenere un monitoraggio accurato dell'impianto, della qualità del sistema, dell'efficienza energetica e di una tempestiva segnalazione delle criticità impostate.

# Bibliografia

- [1] STEPHEN KAISLER, FRANK ARMOUR, 2016. Big Data: Issues and Challenges Moving Forward
- [2] SAEID SADEGHI DARVAZEH, IMAN RAEESI VANANI, FARZANEH MANSOURI MUSOLU. Big Data Analytics and Its Applications in Supply Chain Management
- [3] GERHARD PLENERT, What Is a Supply Chain? , Reinventing Lean - Introducing Lean Management into the Supply Chain, Chapter I
- [4] GERHARD PLENERT, What Is Supply Chain Management? , Reinventing Lean - Introducing Lean Management into the Supply Chain, Chapter II
- [5] DAVID LOSIN, 2013, Big Data Tools and Techniques, Big Data Analytics From Strategic Planning to Enterprise Integration with Tools, Techniques, NoSQL, and Graph, Pages 61-72
- [6] DEEPAK JAIN, ISHANT MEHTA, JISHNU MITRA, SAURABH AGRAWAL, 2016. Application of Big Data in Supply Chain Management, 5th International Conference of Materials Processing and Characterization;

- [7] QI LI, ANG LIU. Big Data Driven Supply Chain Management
- [8] RAY Y.ZHONG, STEPHEN T.NEWMAN, GEORGE Q.HUANG, SHULINLAN, 2016. Big Data for supply chain management in the service and manufacturing sectors: Challenges, opportunities and future perspectives, Computers and Industrial Engineering Volume 101, Pages 572-591
- [9] IN LEE. Big data: Dimensions, evolution, impacts, and challenges
- [10] Big data and predictive analytics applications in supply chain management, Computers and Industrial Engineering;