



CONFINDUSTRIA
Veneto SIAV S.p.A.

Fondirigenti
Innovatori per formazione

BIG DATA

ANALYTICS



Review tecnologica, costruzione del network dei centri per
l'innovazione del Veneto e strutturazione dei relativi servizi
CIG Z7F1EEBEE8 – 15/11/2017

| BIG DATA: LA DEFINIZIONE |

«Il vero valore dei big data non deriva dai dati nella loro forma grezza, ma dalle informazioni che risultano dalla loro elaborazione e analisi»

Thomas H. Davenport in Big Data in Big Companies

Cosa sono i Big Data?

Il termine big data si riferisce a un ampio volume di dati indirizzati alle imprese. Nonostante l'importanza quantitativa, la vera rivoluzione è la capacità di analizzare tutte queste informazioni per produrre valore e conoscenza.

I big data richiedono tecnologie e metodi specifici, capaci di estrarre nozioni aggiuntive rispetto a quelle ottenibili da piccole serie di dati, di trattare così tante variabili in poco tempo e con poche risorse.

I big data possono essere riassunti in cinque caratteristiche fondanti:

Volume. Considera la quantità di dati che le organizzazioni raccolgono da una ampia varietà di sorgenti.

Velocità. I dati fluiscono, si generano, si acquisiscono, si aggiornano e si elaborano a una velocità senza precedenti: vanno gestiti in tempo reale.

Variabilità. I flussi informativi hanno picchi periodici e significato differente a seconda del contesto e del luogo virtuale in cui vengono reperiti.

Complessità. I dati arrivano da molteplici fonti, rendendo difficili collegamenti e gerarchie.

Varietà. Considera le differenti tipologie di dati sia rispetto alla fonte che al formato:

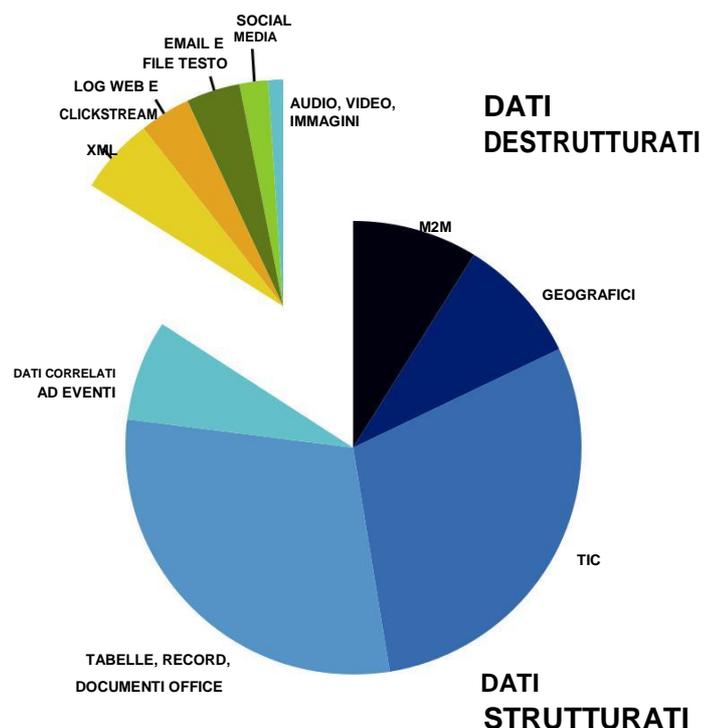
Dati Strutturati: rispettano regole predefinite come la tipologia di contenuto, il formato e la lunghezza. Sono facili da archiviare e interpretare.

Dati Semi-strutturati: non sono conformi a un database, ma sono suddivisibili in record abbastanza strutturati. **Dati non strutturati:** non seguono regole, quindi sono di difficile interpretazione (documenti di testo, email, video, audio e transazioni).

L'ultima differenza è tra dati generati dagli utenti e dati generati automaticamente.

A proposito di Big Data | 1

Il 90% dei dati oggi esistenti sono stati generati negli ultimi 2 anni, con un flusso così crescente che le informazioni accumulate hanno superato l'ordine dei zettabyte (1021 byte), segnando un record per la civiltà. Il ritmo con cui queste informazioni sono prodotte è talmente alto che ogni due giorni viene creato un volume di dati pari alla quantità di informazioni generate dall'umanità intera fino al 2003. Ogni minuto vengono inviate più di 200M di email, su Youtube più di 60h di contenuti, 300mila tweet e fatte più di 9M di telefonate. Si stima che entro il 2020 verranno creati 35 zettabyte di dati (ovvero 35 mila miliardi di gigabyte).



Il grafico mostra la suddivisione dei dati aziendali, distinguendo tra strutturati e non.



| BIG DATA ANALYTICS |

Il processo di elaborazione dati è costituito da fasi cicliche, poiché in continuo aggiornamento. È possibile individuare cinque step consequenziali: acquisizione, immagazzinamento, organizzazione, integrazione e analisi.

Oggi l'innovazione non si basa più sul reperimento dei dati, bensì sull'abilità di gestione, analisi, sintesi, visualizzazione e valorizzazione di queste informazioni. Questo processo prende il nome di Big Data Analytics.

L'analisi dei dati può condurre a diversi livelli di conoscenza e questo processo è correlato alla tipologia di modelli di analytics messi in campo. È possibile identificare quattro categorie principali:

Descriptive Analytics. Strumenti orientati a descrivere la situazione attuale e passata dei processi aziendali. Permettono di visualizzare in modo sintetico i principali indicatori di prestazione.

Diagnostics Analytics. Indaga e spiega la ragione per la quale è accaduto un determinato evento nel passato.

Predictive Analytics. Analisi dei dati per rispondere a domande relative a cosa potrebbe accadere nel futuro.

Prescriptive Analytics. Applicazioni big data avanzate che, insieme all'analisi dei dati, sono capaci di proporre soluzioni operative/strategiche sulla base delle analisi svolte.

A quali business si rivolge?

I big data non devono essere pensati come prerogativa del settore IT. Infatti, se l'Information Technology rappresenta il grande starter da cui partire con gli strumenti necessari, d'altra parte i big data sono necessari e utili nei mercati business più disparati, dalle automobili, alla medicina, dal commercio all'astronomia, dalla biologia alla chimica farmaceutica, dalla finanza al gaming. Nessun settore in cui esiste un marketing e dei dati da analizzare può dirsi indenne da questa rivoluzione. L'utilizzo dei big data supporta lo sviluppo aziendale, così come la relazione con il cliente, dando valore a molteplici realtà.

FINANCIAL SERVICES



IT



RETAIL



CONSUMER



Il grafico evidenzia l'aumento del fatturato grazie all'utilizzo dei big data, tra il 1999 e il 2009.

● Aziende leader Big Data

● Aziende che non utilizzano Big Data

Le figure professionali

Le principali figure professionali, coinvolte nel processo, sono il Data Analyst e il Data Scientist che, pur svolgendo mansioni differenti, risultano complementari nel valorizzare il potenziale informativo. La differenza sostanziale tra i due è il metodo utilizzato per soddisfare il business:

Data Analyst. Riduce un argomento complesso in più piccole parti, meglio comprensibili;

Data Scientist. Fa previsioni future basandosi sui dati a disposizione.

Le piattaforme

Le piattaforme big data sono strumenti che combinano le caratteristiche e le capacità di più servizi in una singola soluzione. Il beneficio, connesso all'usilio di questi strumenti, è ridurre la complessità di svariate applicazioni facendole convergere in un'unica procedura logica. È possibile individuare quattro principali funzioni:

1. Processa e gestisce ogni sorta di dato
3. Fa fronte a molteplici richieste
4. Facilita l'integrazione dei dati aziendali

| PUNTI DI PARTENZA |

Dove reperire i dati

Streaming Data. Questo fenomeno, conosciuto anche come Internet of Things (IoT), è l'insieme dei dati che raggiungono i sistemi IT da una rete di dispositivi collegati. Le aziende possono raccogliere questo tipo di informazioni e decidere quali analizzare subito e quali, invece, conservare per analisi successive.

Social Media Data. Le informazioni provenienti dalle interazioni sui social divengono sempre più interessanti, in particolare per il marketing, l'area vendite e il supporto clienti. Questa tipologia di dati si presenta solitamente in forma destrutturata o semi-strutturata: oltre alla vastità dei dati, la sfida è riuscire a classificarli.

Fonti pubbliche. Un'ulteriore mole di dati proviene da fonti open data come dati.gov.it, datiOpen.it, data.gov o European Union Open Data Portal.

Definire la strategia

Una volta iniziato a sfruttare le informazioni, è necessario prendere alcune decisioni:

Come immagazzinare e gestire i dati

Molti anni fa lo spazio dedicato allo storage sarebbe stato un problema, oggi però esistono opzioni low-cost per memorizzare i dati.

Quanti dati analizzare

Alcune organizzazioni non escludono nessun dato dalla loro analisi, ciò è possibile con le nuove tecnologie altamente performative. Un altro approccio è determinare in anticipo quali dati siano rilevanti prima di analizzarli.

Come usare le informazioni scoperte

Più conoscenza significa maggiore fiducia nelle decisioni aziendali, ma il volume di informazioni in possesso dell'azienda necessita di una strategia.



Settori di utilizzo dei big data: distribuzione nel mercato italiano.

Scegliere le tecnologie

Quali sono le tecnologie più adatte a ricavare il meglio dai big data e dai big data analytics? È possibile prendere in considerazione:

1. Memoria di archiviazione capiente ed economica.
2. Processori più veloci.
3. Piattaforme open source, economicamente accessibili come Hadoop.
4. Elaborazione parallela, clustering, virtualizzazione, griglie di grandi dimensioni, connettività e capacità di trasmissione alte.
5. Cloud e altre risorse per archiviazioni flessibili.

A proposito di Big Data | 2

Durante il progetto Mountain View condotto nel 2008, Google – analizzando i termini di ricerca digitati dagli utenti sul proprio motore – era riuscito a prevedere l'avanzamento dei focolai di influenza negli USA, più velocemente di come lo stesso ministero della salute non fosse riuscito a fare utilizzando i record di ammissione ospedaliera nelle strutture pubbliche e private del Paese.

| QUAL È IL POTENZIALE |

Perché i big data sono importanti?

Aziende ed enti possono raccogliere dati da molteplici fonti e analizzarli per trovare risposte che permettono di tagliare i costi, ridurre i tempi, sviluppare nuovi prodotti e ottimizzare l'offerta, prendere decisioni più consapevoli.

Rispetto al passato, grazie alle nuove tecnologie è possibile:

1. Processare qualunque tipo di dati, non solo quelli prettamente strutturati.
2. Inserire dati grezzi, senza che debbano essere puliti per massimizzare la loro utilità.
3. Usufruire di dispositivi di largo consumo ed economici.

Le opportunità in azienda

Nello specifico, quando ai big data si uniscono gli analytics è possibile:

1. Determinare in tempo reale le cause di guasti, avarie o difetti.
2. Creare offerte nei punti vendita basate sulle abitudini dei clienti.
3. Ricalcolare interi portafogli di rischio in pochi minuti.
4. Individuare comportamenti fraudolenti prima che colpiscano la propria organizzazione.

Le opportunità sul mercato

In generale, l'utilizzo dei big data:

1. Incentiva il social mining, ossia la capacità di misurare e prevedere crisi economiche, tendenze di consumatori, diffusioni di pensieri e opinioni, distribuzione di risorse.
2. Aiuta la razionalizzazione della spesa pubblica, il miglioramento dei servizi offerti alla cittadinanza, la riduzione di sprechi con conseguente ottimizzazione delle risorse.
3. Incrementa lo sviluppo e la ricerca tecnologici.

Possibili difficoltà

I principali rischi imputabili ai big data sono:

Privacy. Proprietà e sfruttamento dei dati: a chi appartengono i dati e per quali scopi li utilizzano?

Qualità dei dati. Completezza, consistenza, accuratezza, assenza di duplicazioni.

Trasparenza. Chi può fare cosa con i dati?.

Monopoli ed Asimmetrie. Controbilanciare il potere delle grandi corporation della conoscenza.

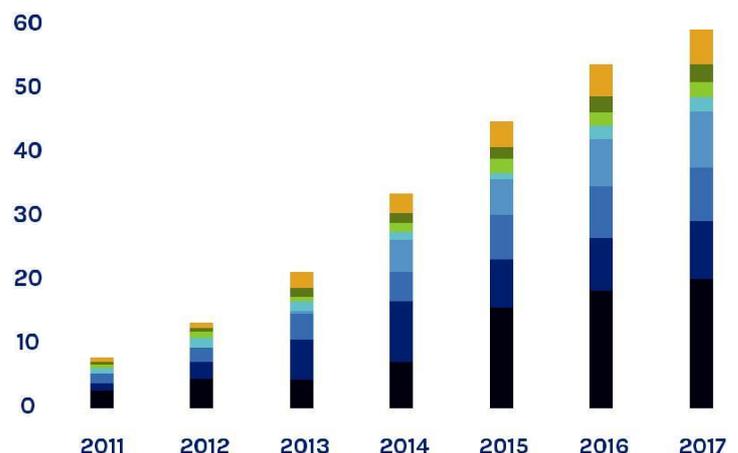
Condivisione. Difficoltà nel mettere a disposizione degli altri le informazioni.

A proposito di Big Data | 3

Con l'aiuto dei big data, gli istituti di credito hanno individuato alcune associazioni inusuali per valutare il rischio finanziario di una persona.

Secondo alcune ricerche, chi compra i feltrini per i mobili rappresenta una clientela migliore perché più attenta e propensa a colmare i propri debiti nei tempi giusti.

Il grafico mostra l'incremento mondiale nell'utilizzo dei big data, dal 2011 al 2017.



| DATA VISUALIZATION |

«By visualizing information, we turn it into a landscape that you can explore with your eyes, a sort of information map.
And when you're lost in information, an information map is kind of useful»
David McCandless (designer) in TED talk

La Data Visualization è definita come l'esplorazione visuale/interattiva – e la relativa rappresentazione grafica – di dati di qualunque dimensione, natura e origine. È esattamente il modo in cui il cervello umano elabora le informazioni; esso non funziona come un calcolatore e non è in grado di processare in parallelo centinaia di migliaia di dati generati in azienda.

Grazie a questo metodo è possibile:

1. Identificare le aree di business in cui apportare modifiche o miglioramenti.
2. Comprendere i fattori chiave che influenzano il comportamento dei clienti.
3. Elaborare un efficace posizionamento dei prodotti.
4. Effettuare previsioni sui volumi di vendita.

Come utilizzarla?

Consapevoli dell'importanza dei dati, è importante capire obiettivi e bisogni. Quindi, predisporre l'azienda alla data visualization richiede di:

1. Selezionare e comprendere i dati che si intendono visualizzare, incluse le loro dimensioni.
2. Determinare cosa visualizzare e quale tipo di informazioni comunicare.
3. Conoscere l'audience e analizzare come elabora l'informazione visiva.
4. Utilizzare una visualizzazione che trasmetta le informazioni nella maniera più appropriata e semplice per il target in esame.

<https://www.youtube.com/watch?v=aeHqYLgZP84>

A proposito di Big Data | 4

Il primo utilizzo documentato del termine “Big Data” risale al 1997, ed emerse in una ricerca redatta da due scienziati della NASA, Michael Cox and David Ellsworth, in merito a una ben nota sfida degli anni '90: i supercomputer generavano un altissimo numero di informazioni – nel loro caso simulazioni di flussi d'aria attorno ai velivoli – che non potevano essere processate e visualizzate. Denominarono la difficoltà “the problem of big data”.

Quando utilizzarla?

La data visualization aiuta le aziende a dare un senso ai loro dati, indipendentemente dalla loro dimensione e dal loro settore di mercato. Ecco come:

Comprendere velocemente le informazioni Utilizzando una rappresentazione grafica, le aziende sono in grado di visualizzare enormi volumi di dati in modo semplice e ottenere le informazioni necessarie a prendere le giuste decisioni di business.

Data la facilità di analizzare le informazioni in un formato grafico, rispetto a quanto avviene con i fogli di calcolo, è possibile comprendere le difficoltà o rispondere alle domande in modo tempestivo.

Identificare relazioni e modelli

Anche i volumi più estesi di dati complessi acquistano un senso se visualizzati graficamente: le aziende, in questo

modo, possono riconoscere più semplicemente le correlazioni tra i dati.

Evidenziare i trend emergenti

Utilizzare la data visualization per scoprire i trend attuali, sia della propria azienda che del mercato, può fornire un vantaggio competitivo rispetto alla concorrenza e, di conseguenza, migliorare il processo decisionale.

Diventa più semplice evidenziare le variabili che influiscono sulla qualità dei prodotti o sull'abbandono dei clienti, nonché rispondere alle difficoltà.

Condividere le proprie scoperte con gli altri

Durante la fase di condivisione, utilizzare rappresentazioni grafiche dall'alto impatto visivo è importante poiché crea coinvolgimento e diffonde più velocemente il messaggio, agevolando la comprensione.



Riccardo Scandellari

L'Intelligenza Artificiale ci rende più umani

C'è un ambito in cui l'essere umano non entrerà mai in competizione con la macchina: la relazione. Creare contenuti, ideare storie coinvolgenti, nutrire e condividere le emozioni di chi gli è simile.

Sono d'accordo con il neuroscienziato dell'Istituto Max Planck, Danko Nikolic, che ha affermato che "non si può superare l'intelligenza umana, l'AI potrà avvicinarsi, ma superarla mai". Stiamo assistendo all'alba di una nuova era in cui le macchine affiancheranno l'uomo nel ragionamento e non solo nei calcoli.

Alcuni quotidiani americani, per arrivare primi sulla notizia, hanno sperimentato forme di giornalismo automatizzato, più rapido rispetto ai tempi imposti dai giornalisti in carne e ossa. Nessun computer ha ancora vinto il premio Pulitzer e temo non succederà mai.

Non importa quanto sei intelligente, importa come lo sei

Le macchine, rispetto alle persone, hanno vantaggi indiscussi. Riescono ad analizzare un'enorme quantità di dati in brevissimo tempo e non si fanno influenzare. I numeri, se inseriti correttamente e non inquinati, una volta elaborati, restituiscono un quadro chiaro e impietoso di informazioni precise a cui affidarsi. Questo mestiere non apparterrà più all'essere umano. Le persone e i professionisti del futuro dovranno adattarsi a nuove mansioni, non in concorrenza con le macchine.

Howard Gardner, lo studioso che ha teorizzato l'esistenza di ben 9 tipi di intelligenza, ci avverte che "non importa quanto sei intelligente, importa come lo sei".

Non sempre "intelligence" è sinonimo di intelligenza

C'è un ambito in cui l'essere umano non entrerà mai in competizione con la macchina, la relazione. Creare contenuti, ideare storie coinvolgenti e nutrire le emozioni dei propri simili, è la sua vocazione esclusiva. I limiti e le debolezze degli uomini non sono inferiori a quelli delle macchine. Alan Turing diceva che "se ci si aspetta che la macchina sia infallibile, allora essa non può essere anche intelligente", perché l'intelligenza è la prerogativa di chi si assume un rischio, anche a fronte di dati contrari, seguendo l'istinto che detta regole che la logica non comprende.

Quando la relazione è più importante di un algoritmo

Le persone che svolgono ruoli surrogabili dai calcolatori devono compiere, fin da ora, una scelta. Hanno l'urgenza di spostarsi verso mansioni e ruoli a cui le macchine non potranno mai ambire: la creatività e la relazione. All'essere umano sono richiesti compiti difficilmente replicabili da un algoritmo, quali creare una visione, prendere decisioni e costruire relazioni autentiche con i suoi simili. Robot e Computer non sono stati creati per sostituirci, ma per facilitare il nostro lavoro e creare le condizioni per un ulteriore progresso.

Chi ha compreso che gli strumenti di comunicazione digitale ottengono la conversione dei contenuti in relazioni, ha un enorme vantaggio competitivo. Questi strumenti sono una palestra per la creatività, un luogo in cui le idee e le informazioni si trasformano in incontri e opportunità. Anche se all'intelligenza artificiale verranno assegnati compiti di relazione automatica e di informazione, ad esempio attraverso servizi di messaggistica, non riuscirà mai a generare empatia e vicinanza con gli esseri umani. Le persone si fidano dei loro simili, non di avanzati risponditori automatici. La Marketing Automation non risolve un problema, lo semplifica. L'acquirente si fida di più dei feedback di altri esseri umani e basa le sue convinzioni sulla passione e la competenza che un suo simile riuscirà a trasmettergli. Analogamente, la comunicazione non può essere ridotta a schemi. Serve un approccio creativo in modo da distinguerci dalla comunicazione sintetica.

Stephen Hawking ha detto che "gli esseri umani, limitati dalla loro lenta evoluzione biologica, non potranno competere e saranno sostituiti dall'Intelligenza Artificiale". Si riferiva certamente ad esseri umani noiosi e facilmente replicabili, ragionieri dello sportello della monotonia, amministratori contabili del tedio e della ripetitività.