

APPUNTI VIDEO

Maddalena Maggiore

https://www.youtube.com/watch?v=b-kj-Hqbl_8

BUSINESS INTELLIGENCE (BI): articolo utile per approfondire quanto comunicato durante il video [Business intelligence: cos'è e perché è importante \(tableau.com\)](#)

<https://www.youtube.com/watch?v=PJ2Gxvan6zQ>

- TOOLS: strumenti di interfaccia per analisi semplici – poca programmazione – facilmente accessibili alle aziende.
- DATA SCIENTIST: modelli previsionali.
- BUSINESS INTELLIGENCE: estrae i dati da modelli storici.
- SQL: linguaggio di interrogazione dei DATA BASE – richiesto anche per l'utilizzo dei TOOLS.
- R/ PYTHON/ JAVA: linguaggi di programmazione.

<https://www.youtube.com/watch?v=Ph5DHoWHW2c>

Struttura piramidale del *modus operandi* di un Data Scientist:

- DATA COLLECTION: capacità di individuare quali dati possono essere estratti e come.
- ETL (Extract Transform Load): sistemi per estrarre, trasformare e ricaricare i Big Data (es. Spark; Hadoop; etc.) – dalle risorse alla Data Warehouse.
- DATA EXPLORATION/ DATA CLEANSING: attività messe in atto per individuare, corruzioni, anomalie, rumori.
- ANALYTICS: sviluppo delle metriche, delle correlazioni, dei features (funzionalità) da utilizzare, etc.
- (SIMPLE) MACHINE LEARNING: modelli semplici di algoritmi per testare le ipotesi del passaggio precedente – Logistic Regression; Random Forest Simplified; etc.
- ARTIFICIAL INTELLIGENCE (AI): modelli complessi di Machine Learning, portando il processo di Data Scientist allo State of the Art.

<https://www.youtube.com/watch?v=BAkVWqs4GyA>

<https://www.youtube.com/watch?v=36roiXQ38uk>

BIG DATA: sono enormi quantità di dati diversi tra loro; sono definiti attraverso quattro principali categorie:

- VOLUME: l'imponente quantità a disposizione.
- VARIETA': tipologie di dati.
- VELOCITA': quanto rapidamente cambiano i dati nel tempo.
- VARIABILITA': quando la velocità con cui i dati sono prodotti è incostante.

L'innovazione introdotta da questi dati sta nella capacità di collegarli tra loro attraverso algoritmi e trarne previsioni per un futuro.

Trovano la loro applicazione attraverso diversi contesti:

- INDUSTRIA 4.0: nell'automazione industriale i Big Data consentono la correlazione tra i dati dei clienti, la pianificazione e la manutenzione dei macchinari, perfezionando l'efficienza operativa e le tempistiche.
- SALUTE: attraverso dispositivi di monitoraggio, personalizzando ausili e tutori e migliorando diagnosi, trattamenti e cure preventive.
- VEICOLI/ TRASPORTI: autoveicoli autonomi e intelligenti, ottimizzazione dei percorsi e la velocizzazione dei servizi.
- AMBIENTE: modelli previsionali metereologici e di qualità dell'aria.
- SPORT: analisi di atleti e squadre, dei flussi di gioco, dei movimenti tattici, della prevenzione degli infortuni.
- SETTORE ENERGETICO: per calibrare domanda e offerta e ridurre gli sprechi.
- SETTORE FINANZIARIO: individuazione delle azioni fraudolente; sviluppo delle analisi di mercato e delle tendenze commerciali delle imprese e per le imprese (es. BLOCK CHAIN: [La Blockchain spiegata in maniera semplice: significato e applicazioni \(osservatori.net\)](#)).
- GRANDE DISTRIBUZIONE: per realizzare piani di offerta e scontistica efficaci e personalizzati, che permettano di evincere le abitudini di consumo e di spesa; effettuare il giusto posizionamento dei prodotti in termini di luoghi, tempi e costi.
- DOMOTICA.
- TELECOMUNICAZIONI.

Tutta questa tipologia di dati ha quattro aspetti in comune:

- I dati rivelati da comportamenti di imprese e utenti vengono raccolti.
- Questi dati vengono trattati da appositi algoritmi.
- Elaborazione di modelli predittivi sulla base degli algoritmi sopra citati.
- Estrazione del valore economico del dato.

Da chi vengono gestiti? Da alcuni colossi definiti Big Tech.

Questo accentrimento di un elevato numero di dati nelle mani di pochi ha fatto emergere alcuni macro-problemi oggi in via di dibattito:

- Regolamentazione della Cyber Security.
- Manipolazione delle informazioni.
- Normazione delle dinamiche economiche ed etiche della gestione di questa nuova risorsa.
- Non esiste un trattato internazionale che disciplini l'utilizzo di internet e del dato.

Ma quanto vale un dato? I criteri per stabilire il valore di un dato sono:

- La varietà.
- Il numero di azioni che lo hanno generato.
- Impiego del dato attuale e prospettico.

A chi appartiene il dato? I diritti di appropriazione del dato ad oggi sono ambigui.

Come funziona il mercato del dato?

- Cessione dei dati dai consumatori ai broker di dati.
- Cessione dei dati dai broker ai soggetti che li utilizzano per scopi commerciali.

Dove si crea il valore del dato?

- Nella capacità intrinseca della profilazione della nostra domanda su prodotti e servizi, customizzandoli.
- Nel miglioramento costante degli algoritmi che imparano con l'ingresso di un maggior numero di dati, arrivando a una precisione quasi chirurgica sull'elaborazioni delle prospettive future e attuali.

Come si estrae il valore del dato?

- Il sistema acquisisce i dati per evitare un'*over collection*, li prepara subito e li conserva per gli usi successivi (Data Silos; Data Warehouse; etc.).

- Il dato grezzo (Data Exhaust come Cookies, Files Temporanei, etc./ Impronte Digitali) viene condiviso e la Varietà ne determina qualità e utilizzo (Dati Strutturati, Dati Non Strutturati, Dati Semi-strutturati).
- Analisi, esplorazione, trasformazione e modellazione (predittiva) del dato grezzo attraverso gli algoritmi che li processano e rivelano le *Correlation Insights*.
- Immagazzinamento che garantisca una facile scalabilità e memorizzazione.

Più l'Algoritmo è efficace e sofisticato nell'estrazione del dato, maggiore è il valore dei dati acquisiti. Maggiore è il volume dei dati, maggiore è l'efficienza degli algoritmi. È necessario poi anche ottimizzare il sistema garantendo una sempre maggiore velocità dei processi.

Che cosa si intende per Dati personali? (Direttiva 95/46/CE)

Per dati personali si intende qualsiasi informazione concernente una persona fisica identificata o identificabile (interessato).

Si considera identificabile la persona fisica che può essere identificata, direttamente o indirettamente, con particolare riferimento a un identificativo come un nome, un numero di identificazione, dati relativi all'ubicazione, un identificativo online o a uno o più elementi caratteristici della sua identità fisica, fisiologica, genetica, psichica, economica, culturale o sociale.

È vietato trattare particolari categorie di dati, definiti come dati personali che rivelano l'origine razziale o etnica, le opinioni politiche, le convinzioni religiose o filosofiche, o l'appartenenza sindacale, nonché trattare dati genetici, dati biometrici intesi a identificare in modo univoco una persona fisica, dati relativi alla salute o all'orientamento sessuale della persona, fatte salve determinate eccezioni (cfr. articolo 10 del regolamento (UE) 2018/1725).

<https://www.youtube.com/watch?v=L501HXFshQk>

- DATA ENGINEER: si occupa di strutturare e architettare l'infrastruttura che gestirà e immagazzinerà il dato; ha competenze solide in ambito informatico di data base, machine learning, tecnologie API (Application Programming Interface – interfaccia di programmazione delle applicazioni).
- DATA SCIENTIST: ovvero il Data Researcher, che ha generalmente competenze di statistica, che si occupa di capire come i dati sono correlati tra loro e, di conseguenza, tirar fuori delle informazioni significative.

- DATA ANALYST: la persona che si occupa di tirar fuori degli insights, delle intuizioni, dai dati, con competenze sia in ambito strategico che in campo informatico.
- DATA VISUALIZATION EXPERT: crea dashboard fruibili ed estetiche per rendere il dato accessibile e godibile anche ai non esperti
- DATA STRATEGY MANAGER (DATA SCIENCE MANAGER): colui che definisce l'architettura del team e del progetto di analisi dei dati in azienda, decidendo tipologia e modalità dei dati utilizzati.