



# BIG DATA e ANALYTICS

## Creare valore dai dati

Daniele Miorandi (U-Hopper)

**U·HOPPER**



# Il relatore (ovvero io)



- papà a tempo pieno
- AD presso U-Hopper
- CRO & co-founder presso Thinkinside
- dottorato in Ingegneria delle Telecomunicazioni
- ~ 200 pubblicazioni scientifiche, 8,400+ citazioni
- 5 brevetti
- 4 aziende fondate (2 chiuse), 2 in fase di decollo
- Titolare corso Big Data Technologies @UniTN



# Due parole su U-Hopper



## cosa facciamo?

aiutiamo le aziende  
ad **estrarre valore** dai  
dati

## come lo facciamo?

sviluppando soluzioni di Big Data  
Analytics, Business Intelligence ed  
Intelligenza Artificiale



# Cosa impareremo



- Cosa sono i big data
  - E cosa li caratterizza
- L'importanza dell'uso dei dati nelle decisioni e nei processi aziendali
- “Mi servono i big data?”
- Big data vs big data analytics
- Tre casi d'uso
  - Retail
  - Bancario
  - Energia
  - Trasporto pubblico locale



# Le origini del termine dato



In latino **DATUM** ==>  
**DARE** in italiano

# Il nostro approccio



1. ESTRARRE VALORE DAI DATI
2. APPLICANDO METODI DI BIG DATA ANALYTICS
3. USANDO TECNOLOGIE DI BUSINESS INTELLIGENCE
4. SVILUPPANDO ALGORITMI DI INTELLIGENZA ARTIFICIALE

# Una moltitudine di dati...



**~50**

miliardi  
di oggetti  
connessi a  
Internet entro il  
2030  
(Statista)

**~2,5**

quintilioni  
di bytes vengono  
generati ogni  
giorno  
su Internet  
(IBM)

**90%+**

di tutti i dati  
a ns disposizione  
sono stati  
generati negli  
ultimi 5 anni  
(IBM)

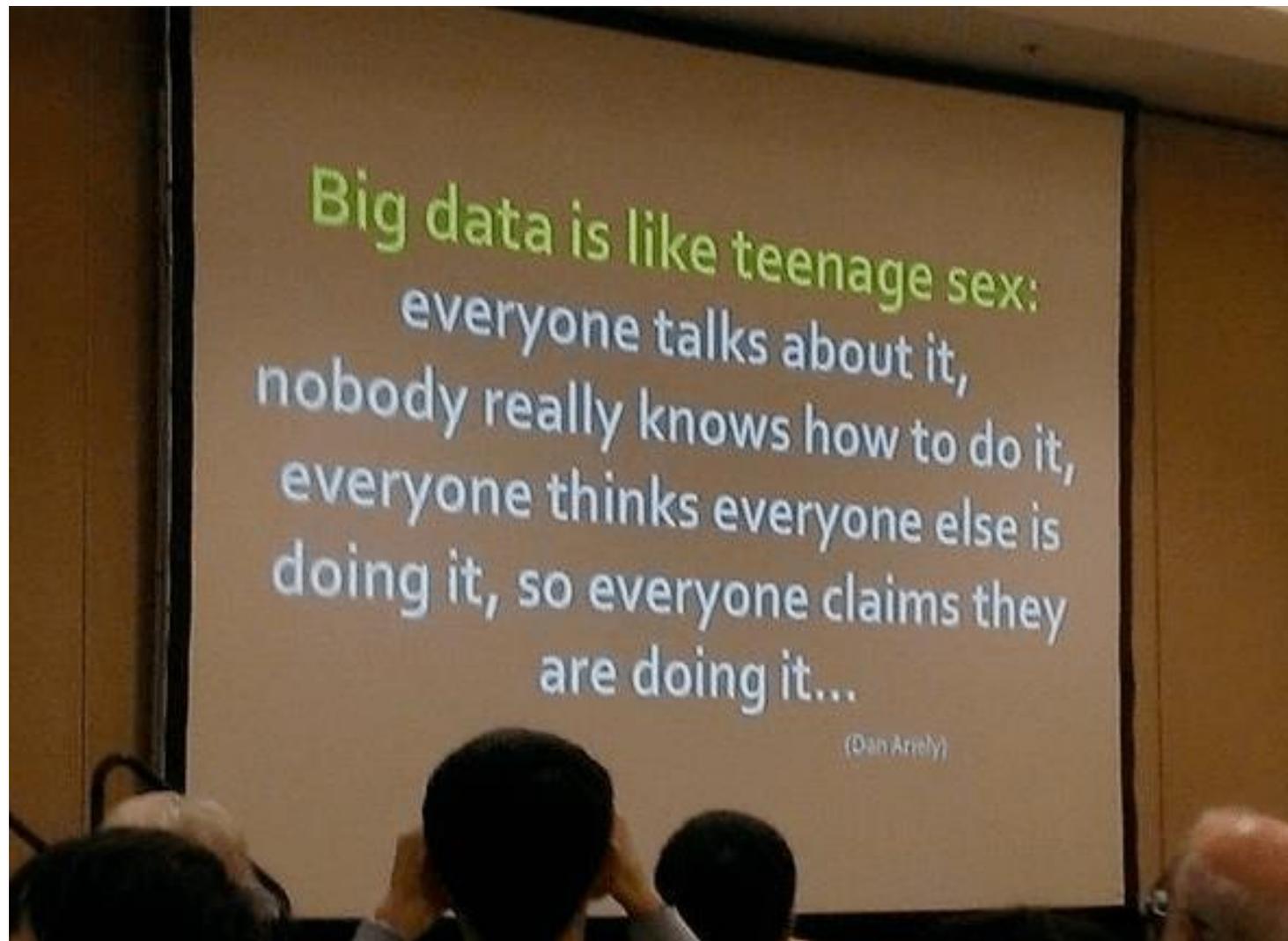
# Dobbiamo imparare ad usare i dati!



Fate in modo che diventi un'abitudine discutere i problemi basandosi sui dati e rispettando i fatti che essi dimostrano.

(Kaoru Ishikawa)

# Tutti parlano di BIG DATA



# BIG DATA vs. SMALL DATA



# Cosa sono i SMALL DATA



Si parla di *small data* quando i dati sono abbastanza "piccoli" (in termini di volume e varietà) da poter essere compresi dall'uomo senza necessità di analisi complesse.

# BIG o SMALL DATA, cosa è preferibile?



- In realtà, la domanda giusta da porsi è: qual è l'informazione che ci forniscono i dati? Quali dati possono essere utilizzati al meglio per estrarre conoscenza?
- In altre parole, le dimensioni non contano. I big data non sono necessariamente migliori dei piccoli dati!
- In realtà, gli *small data* sono generalmente preferiti, in quanto non richiedono competenze e strumenti specifici per poterli gestire. Ma ci sono fenomeni che devono essere compresi su larga scala, quindi...

# Ma perché BIG?



- Il concetto 'Big Data' nasce dal fatto che ogni cosa che noi facciamo lascia una traccia digitale (dato) che noi (e terzi) possono usare e analizzare
- Il termine 'Big Data', quindi, si riferisce alla nostra abilità di utilizzare un numero sempre crescente di dati

# Le 3V nei BIG DATA



- VOLUME (terabyte, petabyte, ...)
- VELOCITY (da static data a data stream)
- VARIETY. (mix di dati ? strutturati ? semi-strutturati ? non-strutturati)

# Volume



...si riferisce alla grande quantità di dati generati ogni secondo. (Volume ~ Dimensioni)

- Non stiamo parlando di Terabyte ma di Zettabyte o Brontobyte.
- La stessa quantità di dati generati **fino 2000** sarà presto pari alla quantità di dati generata **ogni minuto**.
- I nuovi strumenti per big data utilizzano sistemi distribuiti in modo da poter archiviare e analizzare i dati in database sparsi ovunque nel mondo.
- D: Qual è il file/DB più grande che hai mai utilizzato?

# Velocity



...si riferisce alla velocità con cui vengono generati nuovi dati e alla velocità **con cui essi vengono trasferiti.**

- Pensiamo ai messaggi sui social media che diventano virali in pochi secondi.
- La tecnologia ci consente ora di analizzare i dati man mano che vengono generati senza mai inserirli in database
- Velocità → Azione in tempo reale

# Variety



...si riferisce ai diversi tipi di dati che possiamo ora utilizzare.

- In passato ci siamo concentrati solo su dati strutturati che si adattavano perfettamente a tabelle o database relazionali, come i dati finanziari.
- Ora invece, l'80% dei dati mondiali non è strutturato (testo, immagini, video, voce, ecc.)
- Con le tecnologie Big Data oggi possiamo analizzare e riunire dati di diverso tipo come messaggi, conversazioni sui social media, foto, dati di sensori, video o registrazioni vocali.

# 3 Tipologie di dati



- Strutturati
- Non-strutturati
- Semi-strutturati

# Dati strutturati



- **Dati che seguono un modello predefinito o sono organizzati secondo una logica prestabilita**
- Sono immediati da analizzare
- Sono conformi al formato tabulare. Esempi di dati strutturati li troviamo in file Excel o database SQL
- Dipendono dall'esistenza di un data model - cioè un modello che descrive come i dati vengono salvati, processati e come è possibile accedervi
- I dati strutturati sono considerati come la forma più tradizionale di salvataggio dati

# Dati non-strutturati



- Dati che NON HANNO un modello predefinito NÉ sono organizzati secondo una logica predefinita
- Esempi:
  - Testo (email, articoli, etc.)
  - Contenuto multimediale (audio, video)
- Dati non strutturati non possono essere rappresentati in *tabelle*
- Sono difficili da analizzare (perché non siamo a conoscenza della loro struttura)

# Dati semi-strutturati



- Non hanno un data model, ma sono comunque organizzati
- **Elements within the data are separated (tags/markers)**
- Può esserci una gerarchia strutturale dei dati
- “Self-describing structure”
  - struttura adatta per essere processata da una macchina
- Non può essere rappresentata in una tabella (non si possono conoscere a priori le colonne)

# Le 5 P dei BIG DATA



Per creare valore dai Big Data, le aziende hanno bisogno di adottare la strategia delle 5Ps:

- Purpose (scopo)
- People (persone)
- Process (processi)
- Platforms (piattaforme)
- Programmability (programmabilità)

# Purpose (scopo)



- Punto di partenza: occorre sapere cosa stai cercando
- Meglio sempre concentrarsi sul problema e sulla domanda, NON sulla tecnologia o sulla soluzione
- Esempio 1: John gestisce un e-commerce. Come può targetizzare meglio i suoi clienti con offerte specifiche, così da aumentare le sue vendite e margini?
- Esempio 2: Susan è responsabile della politica ambientale nella Regione X. C'è un incendio che si sta diffondendo nella sua regione in questo momento. Come può decidere dove inviare i vigili del fuoco e/o se alcuni cittadini devono lasciare la propria casa?

# People (persone)



- Quali persone in azienda lavoreranno su questo problema?
- Quali competenze sono necessarie?
- È necessaria una riorganizzazione o una nuova assunzione?

# Process (processi)



- Quali processi interni sono "toccati" dal problema?
- In che modo una soluzione basata sui dati cambierebbe tali processi?
- Abbiamo bisogno di introdurre nuovi processi per utilizzare un approccio basato sui dati?
- È necessaria una modifica della procedura?

# Platforms (piattaforme)



- Di quali tecnologie abbiamo bisogno per supportare un approccio basato sui dati?
- Abbiamo bisogno di nuove tecnologie per supportare i processi nuovi/migliorati?
- Li abbiamo? O dobbiamo comprarne di nuovi? O dovremmo esternalizzare del tutto la soluzione?

# Programmability (programmabilità)



- Abbiamo un approccio programmatico per la gestione dei processi nuovi/migliorati?
- Quali passaggi possono essere automatizzati in fase esecutiva?
- Disponiamo di funzionalità di monitoraggio dei processi?
- Sappiamo come intraprendere azioni correttive sulla base dei dati osservati? Possiamo farlo in modo programmatico o automatizzato?

# BIG DATA ECONOMY



# BIG DATA vs. BIG DATA ANALYTICS



## BIG DATA

la capacità di gestire i dati caratterizzati dalle 3V e la relativa infrastruttura IT

## BIG DATA ANALYTICS

la capacità di elaborare big data su larga scala (in molti casi attraverso l'apprendimento automatico) per estrarre insight e informazioni

# Ogni giorno i dati crescono



**“ From the dawn of civilization until 2003, humankind generated five exabytes of data. Now we produce five exabytes every two days...and the pace is accelerating.”  
(E. Schmidt, former CEO of Google)**

# Le nostre attività generiamo dati



- l'ascolto di musica → dati
- la lettura di un libro → dati
- attività sportiva → dati
- perché? Digital music players, eBooks readers e smart watches raccolgono una grande quantità di dati sulle nostre attività (activity data)
  - Il tuo smartphone colleziona un'enorme quantità di dati cts relativamente al tuo utilizzo
  - il web browser che usi di solito salva tutte le tue ricerche
  - l'istituzione finanziaria associata alla tua carta di credito raccoglie dati che indicano dove fai shopping
  - i negozi hanno dati circa i tuoi acquisti.
- riesci a pensare ad un'attività che non generi dati?

# Le nostre conversazioni...



- Le nostre conversazioni sono registrate digitalmente
- Emails → anche, ovviamente
- Social media (IG/FB/Twitter/Whatsapp/TikTok/Tinder...)
- Conversazioni telefoniche (???)
- Alexa/Echo etc.



# I contenuti...

Quante foto/video fai al giorno?

Dove le salvi/carichi/condividi?

**CCTV cameras (!)**

[Privacy....]

**SIAMO DEI PROSUMER\***

\* PROSUMER = PRODUCER + CONSUMER

# E i sensori



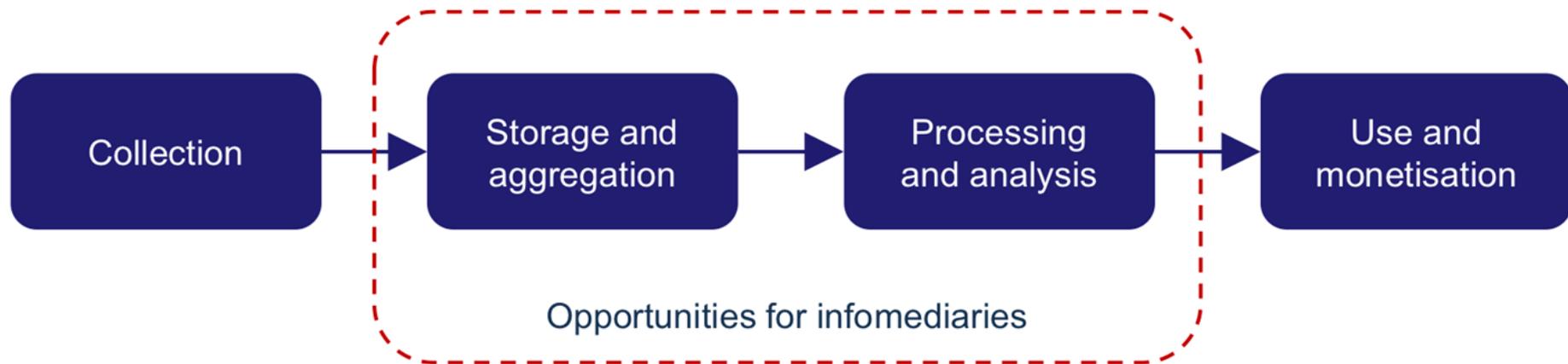
- Siamo sempre più circondati da sensori che raccolgono e trasmettono dati.
- Non pensiamo solo ai sensori in un impianto di produzione o chimico.
- Pensiamo invece ai nostri smartphone: contengono un sensore GPS per tracciare esattamente dove ci troviamo ogni secondo della giornata, includono un accelerometro per tracciare la velocità e la direzione in cui ti stai muovendo, ecc.
- Ma non solo: esistono sensori in una miriade dispositivi e prodotti (!)

# Cos'è una soluzione BIG DATA



- Non ci sono soluzioni “a scaffale” di big data analytics
- Ma ci sono delle tecnologie abilitanti che sono ormai standard
- Stanno emergendo delle piattaforme per supportare big data analytics
- Ad oggi, la big data analytics richiede comunque uno sforzo di customizzazione (gestiti come progetti, non prodotti)

# La catena del valore dei BIG DATA



# Il modello di business emergente



- Le diverse fasi della catena del valore presentano attori che utilizzano diversi modelli di business
- Tendenza comune: *servitizzazione* ("pay per use")
  - Per la raccolta: pay per data items
  - Per stoccaggio/agggregazione: pay per volume
  - Per elaborazione/analisi: pay per operation (chiamate API)
  - Per la monetizzazione: canone mensile

# Il mercato dei BIG DATA



## Traditional Analytics (BI)

vs

## Big Data Analytics

### Focus on

- Descriptive analytics
- Diagnosis analytics

- **Predictive analytics**
- **Data Science**

### Data Sets

- Limited data sets
- Cleansed data
- Simple models

- Large scale data sets
- More types of data
- Raw data
- Complex data models

### Supports

**Causation:** what happened, and why?

**Correlation:** new insight  
More accurate answers



# Alcuni ostacoli



1. Mancanza di sicurezza dei dati
2. Costi elevati
3. Know-how mancante (in particolare: PMI)
4. Mancanza di standard comuni
5. Quadro normativo

# Caso aziendale #1



## STREET SHOPPING DEMAND

- Analisi degli spostamenti delle persone in determinate fasce orarie in luoghi circoscritti (es. centro città, centri commerciali, etc.);
- Incrocio degli spostamenti durante eventi meteorologici
- Analisi del tempo di stazionamento di una vetrina (dwell time)



## RISULTATO

- Alta variabilità di una visita di un punto vendita durante eventi meteorologici specifici
- Pedonabilità di un punto vendita variabile in base al luogo geografico e fasce orarie relative (es. centro città è più pedonabile durante la pausa pranzo, un centro commerciale tra le 18:00 e le 20:00)

**SOLUZIONE = ADOTTARE L'APERTURA DI UN PUNTO VENDITA IN MODO DINAMICO**

# Caso aziendale #2



## LIFE-EVENTS PER IL MARKETING

- I dati del CRM sono dati strutturati, ma non possono fornire informazioni aggiuntive oltre l'anagrafica di un cliente
- Arricchimento dei dati con dati non-strutturati (es. fonte da social network)
- Recupero di informazioni legati ai cosiddetti life-events (es. compleanno, promozione in azienda, nascita di un figlio)

**PRINCIPALE GRUPPO  
BANCARIO ITALIANO**

## RISULTATO

- Arricchimento di un profilo utente grazie a terze fonti
- Nuove meccaniche di marketing automation per stimolare l'acquisto di impulso

**SOLUZIONE = COMUNICAZIONE PERSONALIZZATA NELL'ACCESSO AI SERVIZI ATM**



# Caso aziendale #3



## MANUTENZIONE PREDITTIVA DI IMPIANTI FOTOVOLTAICI

- Analisi dello storico degli interventi di manutenzione
- Analisi delle causali dei ticket di intervento e tempi di risoluzione
- Analisi dei costi di intervento a «vuoto» rispetto agli interventi «standard»

## RISULTATO

- Creazione di un modello di analisi predittiva degli interventi
- Ottimizzazione del ciclo di gestione delle manutenzioni

**SOLUZIONE = NUOVO PROTOCOLLO DI MANUTENZIONE IN BASE A DATI PREDITTIVI E ABBAANDONO GRADUALE DELLA MANUTENZIONE PROGRAMMATA**



# Caso aziendale #4



## ANALISI IN TEMPO REALE DELL'OCCUPAZIONE DI UN MEZZO PUBBLICO

- Analisi dello storico di occupazione di un mezzo pubblico per fascia oraria e linea
- Incrocio con i dati metereologici
- Incrocio con eventi specifici (es. manifestazioni) in base ai dati di traffico giornalieri

## RISULTATO

- Creazione di un modello di analisi predittiva dei passeggeri per linea/fascia oraria
- Indicazione del livello di occupazione in base alle disposizioni COVID-19

**SOLUZIONE = SISTEMA DINAMICO DI INDICAZIONE DELLE SOGLIE DI OCCUPAZIONE DI UN MEZZO PUBBLICO IN TEMPO REALE**



**GRAZIE  
DELL'ATTENZIONE**

