



# BIG DATA

THE BUSINESS OF INFORMATION

**I** CODING  
THE  
FUTURE

Information  
is the oil of the  
21st century,  
and Analytics is the  
combustion engine

— Peter Sondergaard, Senior Vice President, Gartner

#### Autori



Gianluca Pericoli  
PRODUCT OWNER  
& DATA SCIENTIST



Alessandro Zandonà  
DATA SCIENTIST



Renzo Taffarelo  
BUSINESS ANALYST



Gaia Conti  
DIGITAL MARKETING &  
COMMUNICATION



Alberto Bianchi  
DATA SCIENTIST



Ariele Pirona  
VISUAL COMMUNICATION

# Indice

04

BIG DATA: UNA BREVE  
INTRODUZIONE

12

TREND DI ADOZIONE DEI  
BIG DATA

15

CURIOSITÀ

16

LEGAME TRA BIG DATA E  
INTELLIGENZA ARTIFICIALE

20

COME SI USANO I BIG DATA

24

DATA SCIENTIST E IL  
DREAM TEAM

25

TOOL DI DATA SCIENCE

26

INTERLOGICA - LESSON  
LEARNED

32

GLOSSARIO

# BIG DATA

## Una breve introduzione

Big Data e Analytics sono due delle evoluzioni più profonde e pervasive del mondo digitale degli ultimi anni. Un trend destinato a crescere e a incidere profondamente sia sulle nostre vite che sul modo di fare impresa.

Molti gesti che compiamo quotidianamente generano una mole di dati spropositata: smartphone interconnessi, click su un sito web, smart meter, device IoT, richieste vocali a Google, Siri o Alexa. Enormi volumi di dati eterogenei per fonte e formato analizzabili in tempo reale: tutti questi sono Big Data.

Ma quello di Big Data è un concetto ancora recente che ad oggi non ha ancora una sua definizione standard. Di sicuro, la prima caratterizzazione dei Big Data è legata alla grandezza dei dati, il cui punto di riferimento è già nell'ordine dei terabyte (1024 gigabyte), dei petabyte (1024 terabyte) e anche di più, fino agli zettabyte (miliardi di terabyte). Ma per dare un senso al concetto, gli esperti lo hanno suddiviso in 3 macro segmenti rilevanti, definite le 3V di crescita dei Big Data:

**Varietà** - le informazioni archiviate nei Big Data sono molto variegata e ognuna ha

una provenienza specifica

**Velocità** - serve estrema velocità per archiviare, salvare e catalogare questa mole di informazioni

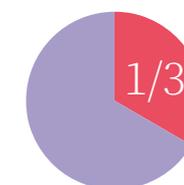
**Volume** - servono soluzioni tecnologiche che permettano di gestire le grandi dimensioni (in termini di bytes) dei Big Data

Queste caratteristiche principali nel tempo si sono traslate in ulteriori 10 segmenti - le 10 V dei Big Data - per descrivere il concetto nel dettaglio.

Quando si parla di Big Data è necessario anche comprendere i cambiamenti che essi comportano: dal processo di raccolta e gestione dei dati, alle tecnologie a supporto del ciclo di vita del dato, allo sviluppo di nuove competenze per la valorizzazione del dato. La gestione e l'analisi dei Big Data richiedono da un lato l'impiego di tecnologie adeguate, dall'altro la presenza di figure professionali che ne massimizzino

Valore del mercato

1,4  
MILIARDI DI €



Delle grandi imprese ha introdotto competenze di Machine Learning.

Dati delle grandi imprese \*

44% Usa Analytics Real-Time o Near Real-Time.

56% Ha già in organico Data Analyst

46% Ha già in organico Data Scientist

42% Ha già in organico Data Engineer

fruibilità e utilità - di qui la centralità della figura del Data Scientist.

Le informazioni estratte dall'analisi di tutti questi dati aiutano ad indirizzare le azioni di business: gli esperti di marketing a indirizzare le campagne in modo più strategico, gli operatori sanitari a identificare epidemie e gli ambientalisti a comprendere la sostenibilità futura, etc.

In relazione al mercato, è possibile ottenere un notevole vantaggio competitivo grazie a decisioni tempestive e maggiormente informate, e questo vale certamente per le grandi organizzazioni, ma anche per le piccole e medie imprese.

Secondo il report sugli andamenti del mercato ICT di Assintel pubblicato a marzo 2019, è stata prevista una crescita dell'1,6% in Italia durante il corso dell'anno. L'andamento incrementale è trainato da grandi e medie aziende che investono

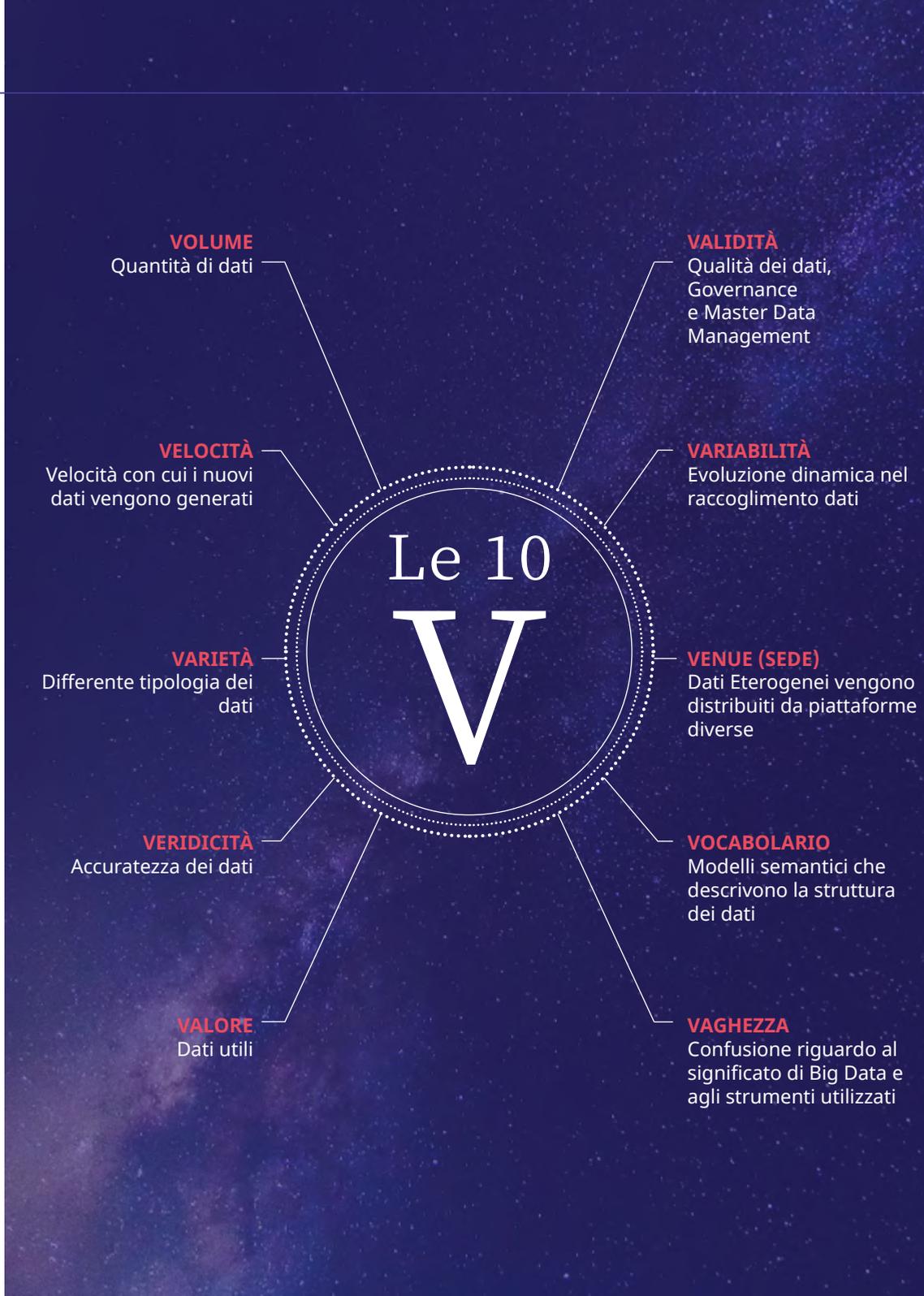
in tecnologie relative agli "Acceleratori dell'Innovazione", come Internet Of Things, Intelligenza Artificiale, Big Data e Analytics, che progrediscono a ritmi decisamente sostenuti.

Rispetto ad un mercato che vale 1,4 Miliardi di euro e in crescita costante, la situazione italiana rispetto all'uso dei Big Data non è certo rosea. La maggior percentuale degli investimenti, infatti, è ascrivibile alle grandi imprese. Un terzo di queste ha introdotto competenze di Machine Learning, il 44% utilizza Analytics Real-Time o Near Real-Time. Il 56% ha già in organico Data Analyst, il 46% Data Scientist, il 42% Data Engineer.

Solo una bassissima percentuale di piccole e medie imprese, invece, ha deciso di avviare progetti di Big Data Analytics. Un dato poco incoraggiante visto il numero rilevante di PMI che compongono il tessuto economico italiano.

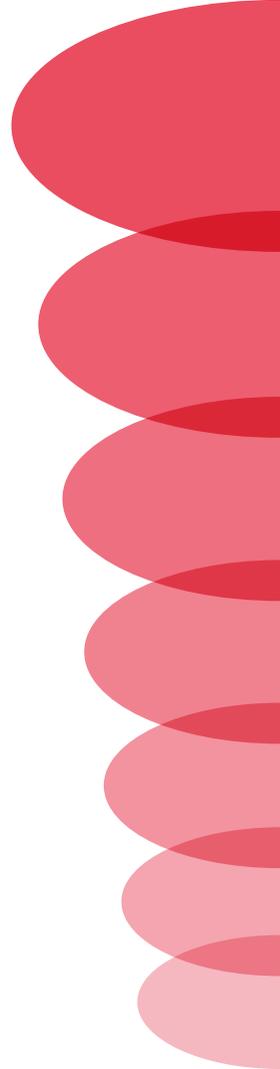
## Da dove arrivano i dati?

-  SMARTPHONE INTERCONNESSI
-  SOCIAL
-  CLICK SU UN SITO WEB
-  SMART METER
-  DEVICE IOT
-  RICERCHE SU GOOGLE
-  RICHIESTE VOCALI A GOOGLE, SIRI O ALEXA



## Quanto pesano i dati?

- ZETTABYTE**  
10<sup>21</sup>
- EXABYTE**  
10<sup>18</sup>
- PETABYTE**  
10<sup>15</sup>
- TERABYTE**  
10<sup>12</sup>
- GIGABYTE**  
10<sup>9</sup>
- MEGABYTE**  
10<sup>6</sup>
- CHILOBYTE**  
10<sup>3</sup>



# Big Data: big timeline



**2008**  
Google elabora 20 Petabyte di dati in un solo giorno

**2010**  
Il presidente esecutivo di Google, Eric Schmidt, afferma: "Al giorno d'oggi il numero di dati creati ogni due giorni corrisponde al numero totale di quelli creati dall'inizio della civiltà umana fino al 2003".

**2014**  
L'88% dei dirigenti aziendali ritiene che l'analisi dei Big Data sia una priorità assoluta per la propria attività.

**2018**  
Forrester Research conclude: "Il 70% delle imprese prevede di elaborare soluzioni di intelligenza artificiale (AI) nei prossimi 12 mesi, rispetto al 40% nel 2016 e al 51% nel 2017." Inoltre, prevede che "il 50% delle imprese adotterà una strategia di analisi dei Big Data cloud-first" e oltre il 50% dei Chief Data Officer (CDO) riferirà ora direttamente al proprio CEO.

**2022**  
Cisco nella dodicesima edizione dello studio "Mobile Visual Networking Index" sostiene che l'avvento del 5G unito al crescente numero di device e di connessioni porterà a un conseguente aumento del volume dati.

Sulle sole reti mobili italiane ci saranno 165 milioni tra device e connessioni IoT, con una crescita dell'11,1% rispetto al 2017, quando ce n'erano poco più di 97,6 milioni.

**2020**  
Traffico di dati: i Big Data crescono a un tasso annuale del 40%.

**2021**  
BCC Research stima che il valore di mercato della Smart City raggiungerà in quest'anno i 775 miliardi di dollari, con un tasso di crescita annuo del 18%.

**2015**  
Bernard Marr, guru dei Big Data, rivolgendosi ai leader del settore e del governo asserisce che "I Big Data non sono un capriccio", e che il vento del cambiamento è alle porte. "Si è solo all'inizio di una rivoluzione che toccherà ogni azienda e ogni forma di vita su questo pianeta", scrive. Aveva ragione. Ogni anno da allora è stato definito, da importanti media, "L'Anno dei Big Data".

**2016**  
Circa 3 miliardi di persone nel mondo hanno accesso alla rete. Le loro attività e preferenze – terabyte di dati – sono in larga parte raccolte per poter in seguito essere esplorate ed utilizzate, quantificando e misurando comportamenti collettivi per comprendere e predire il funzionamento della nostra società e le tendenze del mercato.



## Trend di adozione dei Big Data

La trasformazione digitale abilitata dai Big Data Analytics nelle aziende si concretizza in cambiamenti in diversi ambiti che comprendono tecnologia, processi, organizzazione e cultura aziendale.

Anche il 2019 si conferma un anno di crescita per il mercato dei Big Data Analytics in Italia; le organizzazioni ne hanno compreso il potenziale relativo alle opportunità strategiche.

I Big Data sono fondamentali per analisi approfondite di tipo interpretativo o predittivo a supporto dei processi decisionali e come vantaggio competitivo, in tutti i settori produttivi. L'evoluzione del mercato, infatti, porta i consumatori a chiedere servizi sempre più smart capaci contemporaneamente di risolvere i problemi in tempi rapidi.

D'altra parte, le stesse aziende, sempre più stressate dalla competizione globale, hanno la necessità di razionalizzare le spese e perseguire un percorso di trasformazione verso il concetto di data-driven company, offrendo di conseguenza risposte più rapide alle richieste del business aziendale.

Mentre le PMI coprono ancora un esiguo 12% del mercato, il vero motore della

crescita poggia sulle spalle delle Grandi Imprese (sopra i 249 addetti), che ne rappresentano l'88%.

Lato PMI, il problema maggiore risiede principalmente nella scarsa comprensione della materia, accentuata dall'eterogeneità caratteristica dei Big Data, e da una immaturità tecnologica che non permette di beneficiare degli infiniti vantaggi di progetti ad essi legati. Molto spesso i nodi centrali stanno a monte: la mancanza di conservazione dei dati, presupposto senza il quale non è possibile attuare alcuna strategia, una scarsa comprensione del valore reale del business dei Big Data, ed un approccio di analisi dei dati limitato e tradizionale.

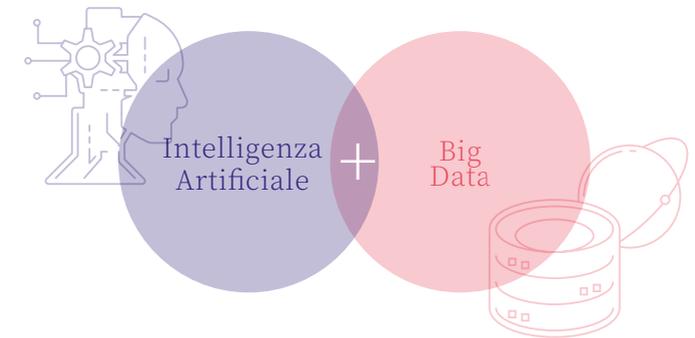
Per le Grandi Imprese, invece, le maggiori difficoltà si rintracciano nella mancanza di competenze e di figure interne, nell'integrazione dei dati e nella stima dei benefici dell'investimento. (fonte: Dati dell' Osservatorio Big Data Analytics & Business Intelligence della School of

Management del Politecnico di Milano).

Se si guarda alla consapevolezza e alla maturità tecnologica delle piccole e medie imprese, dalla ricerca emerge come il 10% continui ad avere una comprensione scarsa o nulla di quali vantaggi i Big Data potrebbero apportare e come l'approccio all'analisi dei dati sia limitato e tradizionale.

Le nuove sfide si giocano quindi sulla velocità e sull'intelligenza delle soluzioni Fast & Smart:

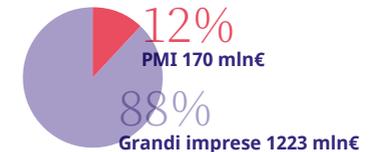
- 1- analisi in real time di dati integrati da diverse fonti informative per rispondere velocemente a esigenze di business;
- 2- ricorso a meccanismi di apprendimento intelligente (Deep Learning e Machine Learning) che consentono la rilevazione di informazioni e pattern nei dati, che si risolvono in un reale vantaggio competitivo.



Il mercato degli Analytics 2018 \*

1393

MILIONI DI €



# Curiosità

## #1

*Un uomo entra in un grande magazzino della catena americana Target, fuori Minneapolis, e chiede di parlare con il direttore. Scopo della visita? Secondo i testimoni, una lamentela piuttosto colorita dovuta al fatto che la figlia, ancora al liceo, avesse ricevuto da Target coupon e volantini su prodotti premaman e per neonati. "State cercando di spingerla a rimanere incinta?". Il manager di Target si è scusato, mortificato al punto che ha chiamato il signore anche il giorno seguente per scusarsi di nuovo. Ma al telefono è il cliente ad essere in difficoltà: "Ho parlato con mia figlia, non lo sapevo, ma è incinta".*

Questo episodio è stato riportato sul New York Times nel 2012.

Come ha spiegato Andrew Pole, responsabile del settore marketing e statistica per

il grande magazzino, questa situazione è il risultato dello studio del comportamento dei clienti sulla base dei Big Data raccolti dal suo team. Osservando le abitudini d'acquisto dei clienti e incrociandole con altre informazioni demografiche, Pole e colleghi sono in grado di anticipare i comportamenti dei clienti e sfruttarli.

Tornando al caso specifico del signore di Minneapolis, Pole e il suo team sono riusciti a correlare gli acquisti alla gravidanza della figlia, perché aveva iniziato ad acquistare una gamma di prodotti solitamente scelta dalle neomamme (un certo tipo di crema, prodotti a base di calcio, magnesio e zinco ed un'altra ventina di prodotti).

Questo aneddoto è l'emblema di come Big Data e Marketing possano consentire di anticipare i bisogni dei consumatori, ma anche di veicolare le abitudini d'acquisto.

## #2

Durante la Seconda Guerra Mondiale, il governo britannico interpellò Abraham Wald, esperto di statistica, per studiare come rendere meno vulnerabili gli aerei.

Wald iniziò a osservare gli aerei di rientro dalle missioni e a creare una mappa degli impatti dei proiettili.

La conclusione del lavoro dell'esperto fu quella di rafforzare le parti non colpite, perché gli aerei che Wald aveva potuto analizzare erano pur sempre tornati alla

base, anche se erano stati colpiti. Quindi, il fatto che quelle zone non fossero state danneggiate e avessero permesso all'aereo di tornare salvo alla base, segnalava l'importanza di preservarle, rafforzandole, per la salvezza del velivolo.

Questa situazione rappresenta un esempio un po' "vintage" di Survival Bias, detto anche Survivorship Bias, che dimostra brillantemente come sia possibile prendere delle decisioni "informate" e data-driven.

# Il legame tra Big Data e Intelligenza Artificiale

“L’espressione “Big Data” è la popolarizzazione delle tecniche di Machine Learning: far lavorare i computer su enormi volumi di dati per scoprire legami e tendenze che gli uomini non possono vedere. E questo spiega tutte le grandi scoperte che stiamo facendo in questi anni”.

Kenneth Cukier – Senior Editor Digital Products dell’Economist

Come sottolineato già in precedenza ogni organizzazione ha ormai accesso a una mole di dati inimmaginabile: informazioni provenienti dall’attività dell’impresa, dal web, dai social, dai sensori IoT e da molte altre fonti che sono linfa vitale da cui estrapolare informazioni utili a guidare la crescita del business. La semplice raccolta o il solo accesso ai dati, senza la corretta analisi, non produce alcun risultato e rimane solo un cumulo enorme di informazioni grezze.

Per rendere questi dati veramente utili è necessario ricorrere ad una sinergia tra tecnologie. La connessione tra Big Data e Intelligenza Artificiale è quindi necessaria per poter trasformare i primi in dati in grado di predire e veicolare verso decisioni più efficaci.

L’automazione delle decisioni deve passare obbligatoriamente per questa convergenza di tecnologie per poter essere agile e veloce come il mercato di oggi richiede.

Fulcro del processo sta nel diventare a tutti gli effetti un’organizzazione data driven, situazione che molto spesso si verifica nella grandi realtà nelle quali i processi decisionali sono maggiormente guidati da dati raffinati. Ciò permette di arrivare a decisioni in tempi rapidi, ma fondate sulla solidità di

valutazioni oggettive e sistematiche.

Fino ad ora le aziende non hanno dedicato abbastanza attenzione alla comprensione dei processi e dei comportamenti dei clienti, e non hanno misurato e sfruttato adeguatamente i dati in loro possesso. Motivo per cui spesso le PMI si sono radicate nella convinzione della poca utilità dei Big Data.

Come trasformare in maniera efficace i dati “grezzi” in qualcosa che fornisca valore aggiunto all’attività? Molte organizzazioni stanno lavorando proprio sulla trasformazione di dati “grezzi” in carburante IA ad alto potenziale. Il reperimento, l’archiviazione e l’elaborazione dei Big Data pone ancora sfide significative per tutte quelle organizzazioni che non possono permettersi di assumere personale interno che se ne occupi.

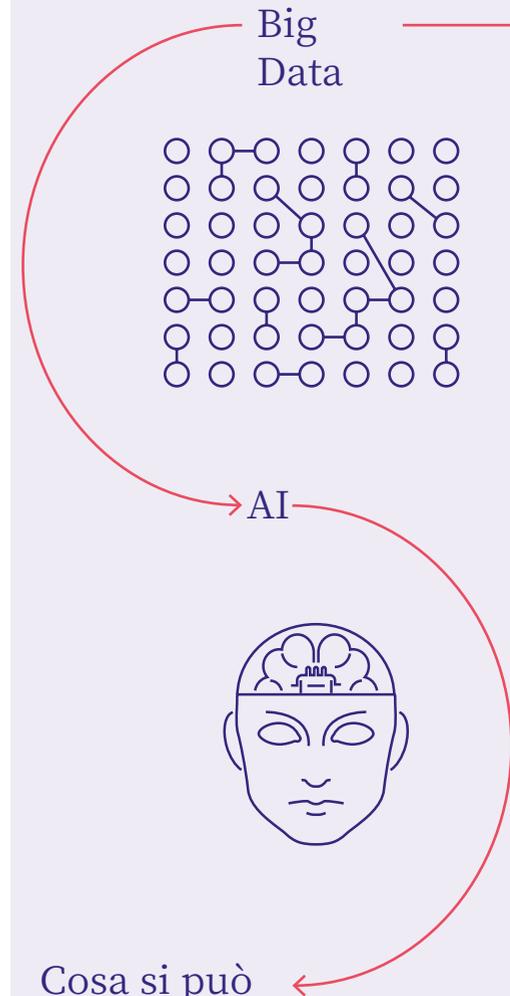
Gli Analytics sui Big Data permettono di ottimizzare efficacemente molti aspetti del business, ad esempio velocizzare e/o facilitare le procedure di interrogazione e visualizzazione di grandi moli di dati, che costituiscono il primo e fondamentale passo verso la valorizzazione dei Big Data.

A questo punto è importante chiarire due concetti che sono comunemente associati: i Big Data sono le materie prime, mentre l’Intelligenza Artificiale (IA) fornisce

gli strumenti per gestirli ed utilizzarli al meglio. Le applicazioni possibili grazie all’interazione Intelligenza Artificiale-Big Data sono molteplici: si può, ad esempio, analizzare il customer sentiment e consentire al marketing di mettere in atto azioni mirate; si può individuare in anticipo la nascita di possibili trend di mercato per guidare l’azienda nelle scelte strategiche di business, si può studiare la concorrenza, e molto altro.

La totalità delle grandi organizzazioni adotta gli Analytics di tipo descrittivo, ma molte di queste stanno sperimentando un’evoluzione verso logiche di predictive, prescriptive e, in alcuni casi, automated Analytics. In breve, si tratta di approcci che utilizzano i Big Data per prevedere o simulare scenari futuri, in modo automatico. La trasformazione passa attraverso tecniche di Machine Learning e Deep Learning che abilitano tipologie di analisi di Real-Time Analytics e Intelligenza Artificiale.

Questo è l’ambito dei contesti in cui opera Interlogica, elaborando piattaforme per la gestione dei Big Data, in grado di accedere a qualsiasi database e data lake, e di analizzare sia dati strutturati che non strutturati. Da qui l’opportunità di esplorare in profondità i dati e definire l’architettura coerente alle esigenze dei clienti.



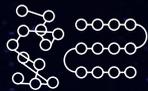
## Cosa si può ricavare

- Customer sentiment
- Prevedere particolari avvenimenti come la rottura di un macchinario
- Occupazione di un determinato spazio
- Nascita di possibili trend di mercato

# Esplora: knowledge discovery platform

Esplora è la suite di visualizzazione/interazione che ti aiuta a prendere decisioni in tutta sicurezza su qualsiasi tipo di dati – Big Data – qualsiasi sia il tuo settore operativo.

## VANTAGGI PRINCIPALI



Riduce la complessità dei progetti di Big Data.



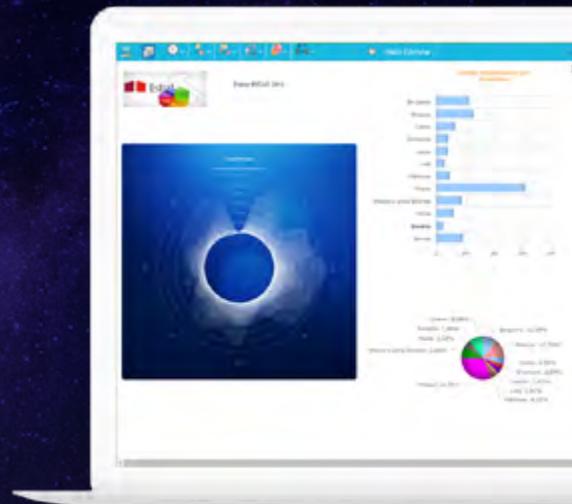
Si collega con facilità a molteplici tipi di database (Oracle, Microsoft, MongoDB, etc )



Integrato con Python, implementa attività di Data Mining.



Aiuta le imprese nel loro processo di trasformazione digitale, facilitando le interazioni tra il dipartimento ICT e le Business Unit.



# Output dei Big Data

L'utilizzo dei dati di cui si è proprietari è la chiave di volta di tutto il processo. I dati in sé possono essere recuperati da qualsiasi fonte e analizzati per risolvere una varietà di problemi ottenendo:



Riduzione dei costi



Diminuzione dei tempi di processo e/o decisione



Sviluppo di nuovi prodotti e ottimizzazione dell'offerta



Decisioni ponderate su dati reali.

L'incontro tra Big Data e Analytics, può avere un forte impatto sulla tua strategia aziendale qualsiasi sia l'ambito di pertinenza.

## AD ESEMPIO SI PUÒ:



Individuare in real time le cause di problemi e malfunzionamenti



Calcolare in maniera automatizzata interi portfolio di rischio



Elaborare soluzioni ed offerte monitorando le abitudini d'acquisto dei clienti



Rilevare comportamenti fraudolenti prima che possano andare a segno mettendo a rischio il tuo business

# Come usare i Big Data

Le aziende che non riusciranno a cogliere nel breve termine le opportunità fornite dai Big Data, avranno difficoltà a stare al passo sui fronti dell'innovazione, della concorrenza sempre più serrata e della produttività.

Per generare un modello di business di successo, è essenziale creare dei flussi funzionali sfruttando diverse fonti di dati, e non è un'operazione semplice da realizzare.

Ci sono molti use case in ogni modello di business che si appoggia ai Big Data.

Il vero valore commerciale di queste fonti di Big Data si rende riconoscibile attraverso specifiche applicazioni, che possono variare notevolmente da settore a settore.

Mentre ci sono interessanti sfide tecniche associate all'integrazione e alla gestione di tutti questi dati, ogni realtà dovrebbe prima prendersi il tempo per identificare il caso d'uso corretto per le proprie esigenze aziendali. Può sembrare all'apparenza una banalità, ma si tratta del primo passo necessario a cogliere le informazioni chiave sul business che possono, una volta approfonditi, generare valore aggiunto.

Ad oggi molte aziende si trovano ancora impreparate e al principio del percorso di esplorazione dei potenziali casi d'uso dei Big Data.

## AUTOMOTIVE

*"Nel prossimo decennio, man mano che le auto si evolveranno in macchine in rete governate dal software, il concetto di marchio automobilistico dovrà affrontare una rivoluzione che vedrà emergere come prioritaria l'esperienza digitale dell'utente*

*all'interno del veicolo, rispetto alle caratteristiche di guida"* Ben Stanley, Global Automotive Research Lead Institute for Business Value, IBM.

L'83% delle persone comprende il valore strategico dei dati che possono alimentare l'efficienza operativa, nuovi modelli di business, esperienze digitali rivolte al consumatore, l'integrazione con altri dispositivi, servizi personalizzati e la capacità di collegarsi ad altri aspetti della vita di una persona. Tuttavia, solo il 18% afferma di operare oggi su una piattaforma di dati digitale. [rif. studio Automotive 2030: Racing toward a digital future]

## CYBER SECURITY

*"Si stima che entro il 2021 i costi complessivi sostenuti per la protezione e le conseguenze dei cyber attacchi supereranno i 1.000 miliardi di dollari. Il mercato dei prodotti e servizi è stimato in 180 miliardi», dice Alessandro Profumo, amministratore delegato di Leonardo, il gruppo nazionale dell'aerospazio e difesa che ha creato una divisione dedicata alla "cyber security"* Il Sole 24 Ore, 24 Settembre 2019.

Watson di IBM memorizza oltre 10 miliardi di elementi (eventi relativi alla sicurezza e dati correlati) e aggiorna la sua conoscenza al ritmo di 4 milioni di elementi in più ogni ora. In un certo senso, Watson per la Cyber Security è come un esperto di sicurezza che legge il Web 24h, non dimentica mai, formula ipotesi su attacchi basati su questa sua conoscenza altamente dinamica e diventa più intelligente nel tempo.

## HR-TECH

*"La professione del HR recruiter sta cambiando molto rapidamente grazie alle nuove tecnologie: Intelligenza Artificiale, Analitiche Predittive, Gamification, sono tutti strumenti che permettono di ridurre il 'time to hire' e migliorare la precisione nella selezione, eliminando pregiudizi o 'unconscious bias' che influenzano negativamente le scelte",* Pietro Novelli, Hr recruitment specialist e Responsabile Italia Oliver James Associates, [Panorama, 17 Ottobre 2018].

Gli HR Analytics sono uno strumento infinitamente utile che consente di monitorare e gestire in maniera data-driven aree aziendali critiche "per definizione", per supportare le scelte dei responsabili della gestione delle risorse umane, valutare obiettivi e performance. Dette anche People Analytics o Talent Analytics permettono l'analisi delle persone e dell'applicazione del processo analitico al capitale umano all'interno dell'organizzazione per migliorare le prestazioni e la fidelizzazione dei dipendenti.

## ENERGY

*Nei prossimi anni avremo a disposizione una grande quantità di dati, che permetteranno di fare previsioni in modo più accurato, e quindi di migliorare la pianificazione, la gestione e la conservazione dell'energia. La produzione e il consumo di energia nel prossimo futuro dovranno diventare più "smart".* [Il Sole 24 Ore, 25 Giugno 2019]

Utilizzando modelli matematici e avendo a disposizione lo storico dei consumi e le previsioni delle temperature su base oraria, è possibile fare buone simulazioni e previsioni su base oraria e giornaliera.

## FINTECH

*Le piattaforme digitali permettono di completare l'automazione dei processi, dalla gestione del cliente prospect all'incasso, di utilizzare i Big Data per acquisire informazioni, di eseguire le transazioni online, di implementare workflow automatici di back office. E possono supportare anche altre fasi del factoring, per esempio la valutazione creditizia e anche l'apertura di canali distributivi innovativi per prospect nuovi, come le aziende di dimensioni molto piccole. Ancora, consentono di risparmiare sui costi operativi.* [Il Sole 24 Ore, 24 Settembre 2019]

I Big Data in ambito Fintech sono ad oggi uno strumento chiave per gli istituti bancari, e, abbinati all'uso dell'IA, si confermano essere uno dei trend che sta destando maggior interesse. Con l'aumentare del numero di record elettronici, i servizi finanziari utilizzano attivamente l'analisi dei Big Data per ricavare approfondimenti aziendali, archiviare i dati e attuare analisi predittive.

## CASE STUDY / BIG TECH

# Amazon

**Fondata da:** Jeff Bezos

**Sede principale:** Seattle

**Dipendenti:** 647.500 (2018)

**Fatturato:** 232,9 miliardi \$ (2018)

Amazon è cresciuta ben oltre la libreria online degli albori, e in gran parte ciò è dovuto all'ampio utilizzo dei Big Data. La sua innovazione più grande è stato il sistema di raccomandazioni personalizzato, basato sui Big Data generati da milioni di transazioni dei suoi clienti. Salendo ad un livello ancora superiore, Amazon Web Services offre elaborazioni di Big Data basate su cloud per applicazioni aziendali.

Amazon ha anche incorporato l'analisi dei Big Data nelle operazioni di assistenza clienti grazie ai loro sofisticati Relationship Management Systems e facendo ampio uso dei dati dei clienti.

Visti i risultati strabilianti su moltissimi fronti, questa fiducia sui Big Data viene continuamente riconfermata portando Amazon a fare da apripista.

## CASE STUDY / BIG TECH

# Google

**Fondata da:** Sergey Brin, Larry Page

**Sede principale:** Mountain View

**Dipendenti:** 98.771 (2018)

**Fatturato:** 136,81 miliardi \$ (2018)

Il gigante delle ricerche Google, non solo ha influenzato in maniera significativa il modo in cui ora possiamo analizzare i Big Data (basti pensare a MapReduce, BigQuery, ecc.), ma è probabilmente il maggior responsabile nel renderli parte integrante della nostra vita quotidiana. Google, infatti, elabora 3.5 miliardi di ricerche al giorno, analizzando per ognuna di esse un database di 20 miliardi di pagine.

Un progetto significativo è BigQuery, un servizio che permette alle aziende di archiviare e analizzare i Big Data sulle proprie piattaforme cloud.

Un altro innovativo progetto di BigG è l'auto a guida autonoma: enormi quantità di dati provenienti da sensori, telecamere e dispositivi di localizzazione accoppiati ai dati di bordo e ai dati raccolti da Google Maps sono analizzati in tempo reale, consentendo alle vetture di Google di essere molto più sicure rispetto alla concorrenza.



# Data scientist e il dream team

Un Data Scientist è un esperto che sa come estrarre significato e interpretare i dati, strutturati o non strutturati, al fine di raggiungere determinati obiettivi di business, trasformando i dati in nuova conoscenza e opportunità.

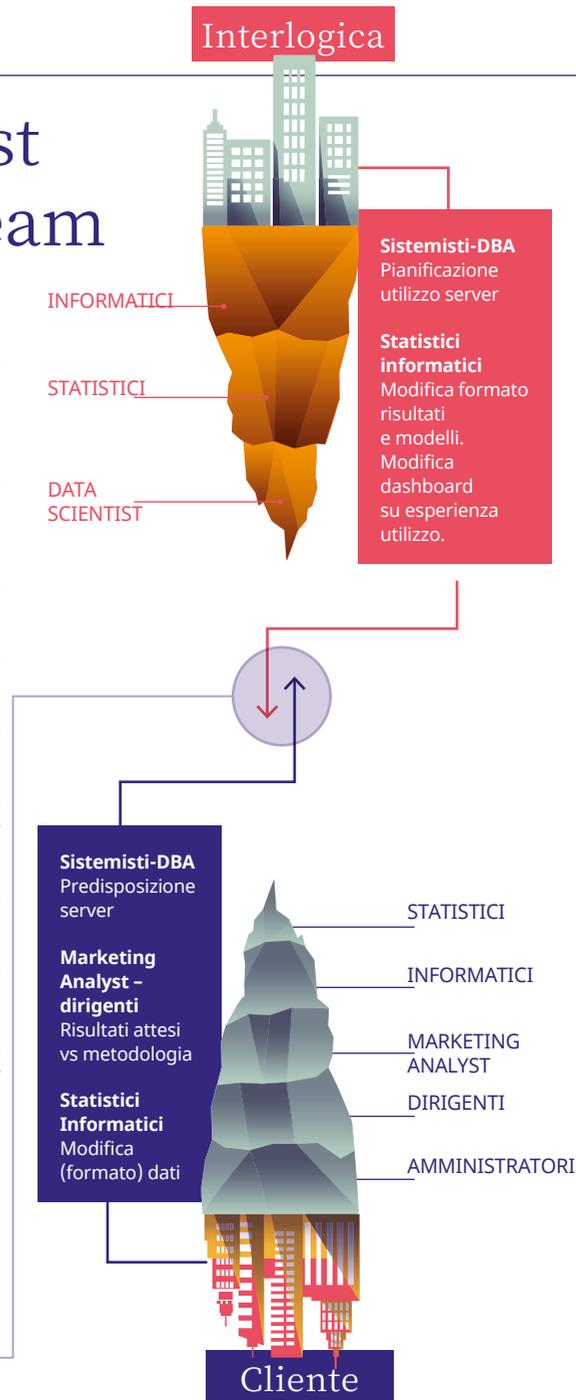
Come professionista deve avere una chiara visione dei tool di lavoro più diffusi e delle piattaforme più adeguate per ogni obiettivo, il che richiede la conoscenza sia di strumenti che di metodi tipici della statistica e del Machine Learning. Gran parte del suo sforzo si concentra nel processo di raccolta, pulizia, organizzazione e formattazione dei dati.

Nel formare il team di Data Science, è necessario prestare attenzione alle peculiarità di ognuno dei suoi componenti. Secondo la nostra esperienza un team multidisciplinare è la soluzione ottimale per poter affrontare progetti eterogenei.

All'interno del nostro Dream Team abbiamo scelto di inserire tre figure diverse: un informatico, che ha competenze in progettazione, ottimizzazione del software e gestione del database; un data scientist che elabora modelli matematici per estrapolare informazioni dai dati e progetta soluzioni per organizzarli in modo fruibile per l'utente; uno statistico che effettua test di significatività sui modelli ed elabora statistiche descrittive utili ad esporre i risultati del suo operato all'utente finale.

### Riunioni periodiche

- Discussione POC
- Pianificazione lavori
- Stato di avanzamento
- Consegna
- Stato dopo consegna



# Tool di Data Science

Il Data Scientist raccoglie i dati da varie fonti - ad esempio applicazioni di business come software ERP e CRM, banche dati, web analytics, social media, documenti elettronici, apparecchi IoT - e li struttura in formati organizzati e accessibili.

Nella sua cassetta degli attrezzi ci sono quindi software e linguaggi di programmazione per interagire con i database, analizzare i dati ed estrarne informazioni utili (Data Mining). Per questi scopi, il data scientist elabora algoritmi o si avvale di metodi statistici e strumenti di Machine Learning, seguendo un approccio scientifico che è al contempo orientato al risultato.

Questo processo consente di individuare all'interno di un gran numero di dati solo le

informazioni di valore per l'azienda, costruendo modelli di correlazione, dimostrando causalità, sviluppando scenari e modelli predittivi.

Punto dolente per molte organizzazioni, e per le PMI in particolare, è il costo di alcuni software proprietari di analisi dei dati e statistici che fanno lievitare notevolmente il budget. Per fortuna sul mercato sono disponibili alternative open source, disponibili per diversi sistemi operativi, in grado di competere con quelli commerciali grazie alle funzionalità e alle community di sviluppatori che hanno alle spalle.

L'Open Source, quindi, non solo è possibile, ma incoraggiato.



INTERLOGICA

# Lesson Learned

In Interlogica elaboriamo soluzioni per la gestione, l'analisi e la visualizzazione dei Big Data (strutturati e non), a partire da qualsiasi database e Data Lake esistente.

Da qui l'opportunità di esplorare in profondità i dati e definire l'architettura ideale per progetti cuciti sulle esigenze di ciascun cliente. A seguire alcuni esempi significativi di come abbiamo aiutato alcuni clienti a estrarre valore dai dataset in loro possesso, attraverso processi, tecnologie e competenze in Analytics e Intelligenza Artificiale.



# Big Data Analysis per il settore energy

## OBBIETTIVO

Piattaforma per la stima dello switch out dei clienti.

—

## CONTESTO

Nel settore dell'energia, una importante multiutility ha deciso di intraprendere un percorso di innovazione con noi basato sui Big Data, attraverso due differenti progetti.

Il primo relativo allo studio del tasso di switch out dei suoi clienti per aumentare la *Customer Retention*.

—

## SFIDA

Comprendere i fattori correlati alle dinamiche di abbandono attraverso l'analisi dei loro database, la ricerca di insight sui clienti e l'attivazione di nuove attività commerciali e di marketing.

## SOLUZIONE

Nello specifico, abbiamo sviluppato uno strumento basato su tecniche di Machine Learning, che permette di analizzare una pletora di informazioni anonimizzate associate ai clienti, come l'anagrafica, lo storico dei contratti e delle offerte sottoscritte, e le interazioni del cliente con la multiutility stessa (reclami, richieste di informazioni o attività di marketing).

Sulla base di questi dati, per ciascun cliente si è stimato il rischio di rescissione del contratto nel prossimo futuro, in modo totalmente automatizzato. La multiutility si è poi occupata di associare le informazioni anonimizzate ai loro reali clienti.

Gli algoritmi di Machine Learning utilizzati sono le Random Forest, che consentono anche di determinare i fattori decisionali distintivi (ad es. l'aumento delle bollette energetiche, etc.) così che si possano proporre offerte mirate per non perdere il cliente, e effettuare delle corrette stime costo/beneficio per determinare il corretto ammontare dell'investimento.



## OBBIETTIVO

Piattaforma per la previsione dei consumi elettrici.

—

## CONTESTO

Big Data e Intelligenza Artificiale si intrecciano anche nel secondo progetto, incentrato sull'efficientamento del processo di previsioni del consumo elettrico.

—

## SFIDA

Ottimizzare il processo di acquisto dell'energia elettrica, riducendo quindi i costi per la multiutility e lo spreco energetico. Ridurre l'errore nella previsione del consumo elettrico dei clienti è punto cruciale in questo processo di ottimizzazione.

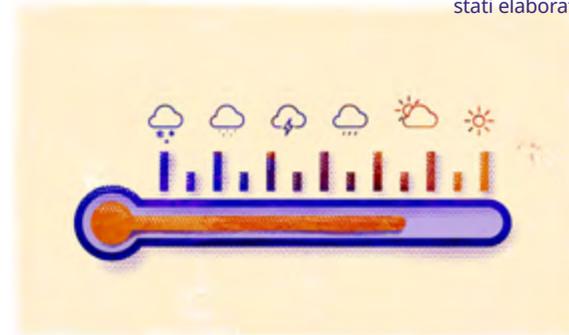
## SOLUZIONE

È stata elaborata una piattaforma che legge in real-time i consumi elettrici e le previsioni meteorologiche, li combina con altre informazioni legate ai consumatori e al calendario, li struttura in un database, e utilizza questa mole di dati per allenare dei modelli predittivi del consumo elettrico.

Si tratta di modelli di decomposizione di serie storiche combinati con metodi di Ensemble Learning (apprendimento d'insieme), selezionati in automatico dalla piattaforma. Quest'ultima è stata infatti dotata di un modello decisionale che consente di selezionare l'algoritmo più adeguato per prevedere il consumo elettrico sulla base della tipologia di consumatore e del periodo dell'anno su cui effettuare la previsione.

Per consentire al cliente di interfacciarsi agilmente con i dati a loro disposizione e visualizzare i consumi previsti dalla piattaforma, abbiamo fornito loro una dashboard interattiva, realizzata in collaborazione con Esplores.

Anche in questo progetto, i modelli sono stati elaborati su dati anonimizzati.



# Big Data per la fatturazione legale

## OBBIETTIVO

Piattaforma di gestione del flusso di fatture elettroniche.

—

## CONTESTO

Uno dei maggiori player della certification authority in Italia ha dato il via a investimenti in progetti d'innovazione con l'uso dei Big Data, così da poter gestire automaticamente il flusso di fatture elettroniche dei loro clienti (il sistema è progettato per reggere picchi di lavoro fino a 500 fatture al secondo).

—

## SFIDA

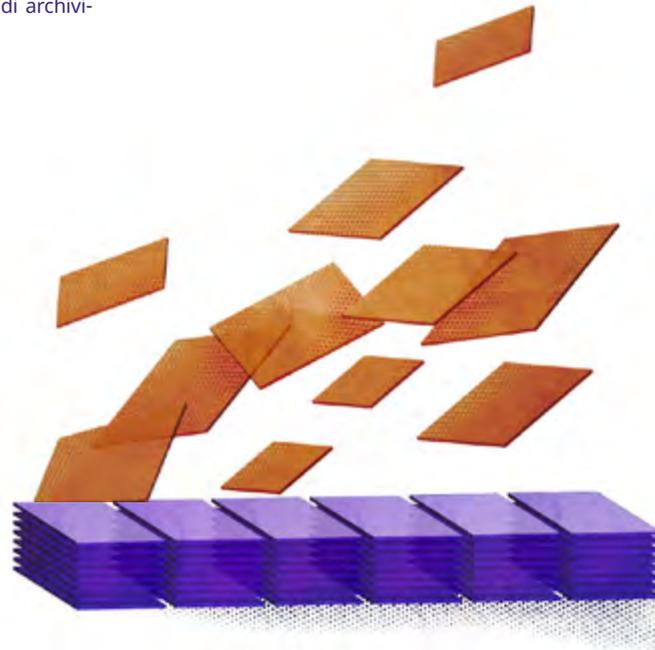
Creare un flusso di gestione automatica della fatturazione che permetta di archivi-

are e analizzare enormi quantità di dati.

## SOLUZIONE

Dopo un'attenta valutazione delle esigenze e delle aspettative del cliente, abbiamo scelto di utilizzare la tecnologia Apache Kafka in combinazione con un cluster Hadoop appositamente progettato per archiviare e analizzare enormi quantità di dati non strutturati in un ambiente di calcolo distribuito.

Tutti questi set di dati sono indicizzati attraverso un database Oracle, che consente un'elaborazione in tempo reale dei relativi metadati. I set di dati resteranno archiviati per tre anni per cui si prevede l'analisi e l'indicizzazione di circa un petabyte di dati in questo lasso di tempo.



# Natural Language Processing applicato al dispatching automatico email

## OBBIETTIVO

Smistamento automatico di email PEC.

—

## CONTESTO

Un quarto progetto ci ha coinvolti nello sviluppo di una piattaforma basata su Intelligenza Artificiale per automatizzare lo smistamento delle email certificate per conto di un Istituto in ambito finanziario.

—

## SFIDA

Migliorare, accelerare e automatizzare le funzioni di analisi del testo per identificare il destinatario delle mail in modo automatico.

—

## SOLUZIONE

Si è scelto di utilizzare il Natural Language Processing (NLP), ossia il trattamento informatico (computer processing) del linguaggio naturale, abbinato a tecniche di Machine Learning.

Per questo progetto il team di Data Science ha analizzato un cluster di email per costruire dei modelli di classificazione del testo.

Questi modelli permettono di classificare le email in entrata in base al contesto di riferimento (mittente, oggetto, corpo), con successivo smistamento automatico al gruppo di lavoro oppure al/ai destinatari.

La piattaforma consente di generare anche delle email di risposta automatica, elaborate sulla base dei contenuti

dell'email stessa.

Il cliente ha, inoltre, la possibilità di correggere eventuali imprecisioni nello smistamento automatico, comunicando alla piattaforma il destinatario corretto; gli algoritmi di IA si ri-modellano in automatico sulle nuove informazioni, incrementando progressivamente la precisione d'uso della piattaforma stessa.

Il risultato è una categorizzazione accurata e affidabile delle email che comporta un risparmio notevole in termini di forza lavoro e riduzione degli errori.

- Utilizza la tecnologia IA di NLP
- Classifica automaticamente le email in ingresso
- Genera una risposta automatica sulla base del contesto di pertinenza dell'email
- Riduce gli errori e determina un risparmio in termini di forza lavoro



# Glossario

## ANALISI PREDITTIVA

Utilizzo di dati, algoritmi statistici e tecniche di Machine Learning per individuare la probabilità di risultati futuri basandosi sui dati storici.

## BIG DATA

Big Data è l'espressione utilizzata per identificare un insieme di dati molto consistente, al punto tale da rendere difficile la gestione con gli abituali strumenti. Generalmente i Big Data rispondono alla regola delle 3 V: volume di dati importante, varietà, velocità con cui arrivano.

## CLUSTER

Gruppo di server che permettono di assicurare una continuità di servizio e di ripartizione di carico di calcolo/rete.

## DATA ANALYSIS

Un processo di ispezione, pulizia, trasformazione e modellazione di dati con il fine di evidenziare informazioni che supportino le decisioni strategiche aziendali.

## DATA ANALYST

Esperto che si occupa di estrapolare il valore dai dati, trasformandoli in una serie di informazioni da comunicare al management aziendale.

## DATABASE

Contenitore informatico che permette di raggruppare le informazioni relative

a un'attività.

## DATA ENGINEER

Esperto che garantisce la disponibilità, la qualità e la fruibilità dei dati a chi li utilizza per metterli al servizio del business identificando e gestendo processi, monitorando fenomeni, individuando opportunità e rischi.

## DATI NON STRUTTURATI

Dati privi di uno schema, e quindi non immediatamente classificabili in modo automatizzato (testi, email, video, messaggi sui social media).

## DATA SCIENCE

Insieme di principi metodologici e tecniche multidisciplinari volto a interpretare ed estrarre conoscenza dai dati.

## DATA SCIENTIST

Esperto capace di raccogliere ed organizzare grandi moli di dati, ed estrapolarne insight, analisi e report.

## DATI STRUTTURATI

Dati organizzati in uno schema e che possono essere ordinati, letti e gestiti automaticamente da un sistema (dati sul traffico di un sito web, coordinate bancarie o coordinate GPS raccolte da uno smartphone).

## DATA WHAREHOUSE

Collezione o aggregazione di dati strutturati, provenienti da sistemi di gestione di basi di dati (DBMS) interni e/o esterni al sistema informativo aziendale.

## EXTRACT TRANSFORM LOAD (ETL)

Processo di estrazione, trasformazione e caricamento dei dati in un sistema di sintesi (data warehouse, data mart, etc).

## HADOOP

Framework di applicazioni distribuite da Apache Software Foundation su licenza Apache, destinato a gestire grandi volumi di dati.

## INTELLIGENZA ARTIFICIALE (IA)

Insieme di metodologie e tecniche per riprodurre o emulare funzioni e ragionamenti tipici della mente umana.

## MACHINE LEARNING (ML)

Settore dell'AI che si occupa di apprendere struttura o comportamenti dai dati, per emularli in modo autonomo e automatico.

## OPEN DATA

Dati che sono liberamente accessibili: possono essere utilizzati da chiunque senza restrizioni di copyright, brevetti o altri meccanismi di controllo.

## PYTHON

Linguaggio di programmazione orientato agli oggetti e utilizzabile per molti tipi di sviluppo software. Distribuito con licenza Open Source.

## R

Linguaggio di programmazione e un ambiente di sviluppo specifico per l'analisi statistica dei dati. Distribuito con licenza Open Source.

## REPLICAZIONE

Processo di copia file in punti diversi di un'architettura per migliorare l'affidabilità e la tolleranza dei guasti.

## SCALA

Linguaggio di programmazione di tipo general-purpose multi-paradigma studiato per integrare le caratteristiche e le funzionalità dei linguaggi orientati agli oggetti e dei linguaggi funzionali.

## STRUCTURED QUERY LANGUAGE (SQL)

Linguaggio utilizzato per interagire con i database, per esempio per estrarne informazioni.

## Contatta i nostri esperti

Ottieni ulteriori informazioni su come Interlogica può aiutarti a sbloccare il valore creato attraverso l'analisi dei Big Data.

### Seguici su

[in company/Interlogica](#)  
[@interlogici](#)  
[@Interlogica](#)

### Maggiori informazioni

[interlogica.it](http://interlogica.it)

### Certificazioni



### Contatti e sedi

#### Venezia – Mestre

Via Miranese, 91/4 - 30174, Mestre (VE)  
+39 041 5354800  
[info@interlogica.it](mailto:info@interlogica.it)

#### Roma

Via G. Coppola di Musitani 34 - 00139 Roma  
[info@interlogica.it](mailto:info@interlogica.it)

Work is not a place  
Puoi trovare noi interlogici  
anche in molti altri luoghi



**I** | CODING  
THE  
FUTURE