



ADOPERARE I
BIG DATA
Alfredo Cartone

Indice

1. BIG DATA E INFORMATION TECHNOLOGY.....	3
2. ALCUNI ESEMPI	5
3. IL DATA SCIENTIST.....	7
RIFERIMENTI BIBLIOGRAFICI.....	10



Attenzione! Questo materiale didattico è per uso personale dello studente ed è coperto da copyright. Ne è severamente vietata la riproduzione o il riutilizzo anche parziale, ai sensi e per gli effetti della legge sul diritto d'autore (L. 22.04.1941/n. 633)

1. BIG DATA E INFORMATION TECHNOLOGY

La grande attenzione rivolta, da sempre, ai temi della p
Nell'utilizzo dei Big Data la qualità dei dati prodotti e, conseguentemente, delle informazioni prodotte, è molto influenzata dalle tecnologie utilizzate. La diffusione dei Big Data ha beneficiato degli avanzamenti nel campo dell'Information Technology.

Grosse quantità di dati richiedono grosse capacità di memorizzazione. Per poter processare grandi quantità di dati è inoltre necessaria un'adeguata capacità di calcolo. Si rivela poi importante creare meccanismi che consentono a più macchine di comunicare tra loro, creando una rete di calcolatori.

I diversi campi che hanno visto un'elevata innovazione funzionale all'ampliamento e alla diffusione dei Big Data sono:

- *Data mining*: l'insieme di tecniche volte all'estrazione di una informazione o di da grandi quantità di dati.
- Base di dati: indica un insieme di dati, omogeneo per contenuti e per formato, memorizzati in un elaboratore e interrogabili via terminale utilizzando chiavi di accesso.
- *Machine learning*: insieme di metodi che fornisce ai computer l'abilità di apprendere senza essere stati esplicitamente programmati.

Attenzione! Questo materiale didattico è per uso personale dello studente ed è coperto da copyright. Ne è severamente vietata la riproduzione o il riutilizzo anche parziale, ai sensi e per gli effetti della legge sul diritto d'autore (L. 22.04.1941/n. 633)

Le innovazioni hanno, quindi, riguardato sia le tecniche che le strategie utilizzate per elaborare i dati. In particolare, ci si è spinti verso la creazione di framework che permettono alle compagnie di modellare enormi quantità di dati con fini previsionali. Tramite tali framework i programmatori che si occupano dello sviluppo possono creare, in base alle loro esigenze, applicativi per l'elaborazione di tali moli di dati.



Attenzione! Questo materiale didattico è per uso personale dello studente ed è coperto da copyright. Ne è severamente vietata la riproduzione o il riutilizzo anche parziale, ai sensi e per gli effetti della legge sul diritto d'autore (L. 22.04.1941/n. 633)

2. ALCUNI ESEMPI

Alcuni esempi possono essere forniti in materia di tecnologie informatiche che hanno aiutato la diffusione nell'uso dei Big Data e la loro odierna popolarità.

1. **Apache Hadoop** è un framework a licenza libera (open-source), ispirato a quello implementato da Google, che permette l'accesso ad un'elevata quantità di dati sotto una licenza libera. Permette alle applicazioni di lavorare con migliaia di nodi e petabyte di dati. Yahoo! è il più grande contributore a questo progetto, e ne fa fortemente uso per migliorare le performance del proprio motore. Apache Hadoop è usato anche da AOL, Ebay, Facebook, IBM.

Gli applicativi di Hadoop permettono, fra le altre prestazioni, di effettuare: marketing analytics, machine learning and/or sophisticated data mining, image processing ed archiviazione e tabulazione di dati.

2. **Oracle Big Data Cloud** è un applicativo che utilizza Apache Hadoop e che permette di archiviare dati per il pronto utilizzo. Tale applicativo a pagamento (è necessario acquisire una licenza) è diffuso principalmente all'interno di imprese.

Questi tipi di applicativi permettono di immagazzinare grandi quantità di dati e di elaborarli in maniera flessibile e veloce, e con

Attenzione! Questo materiale didattico è per uso personale dello studente ed è coperto da copyright. Ne è severamente vietata la riproduzione o il riutilizzo anche parziale, ai sensi e per gli effetti della legge sul diritto d'autore (L. 22.04.1941/n. 633)

costi ridotti. La finalità è quella di ottenere velocemente previsioni affidabili, mediante l'utilizzo di Big Data.



Attenzione! Questo materiale didattico è per uso personale dello studente ed è coperto da copyright. Ne è severamente vietata la riproduzione o il riutilizzo anche parziale, ai sensi e per gli effetti della legge sul diritto d'autore (L. 22.04.1941/n. 633)

3. IL DATA SCIENTIST

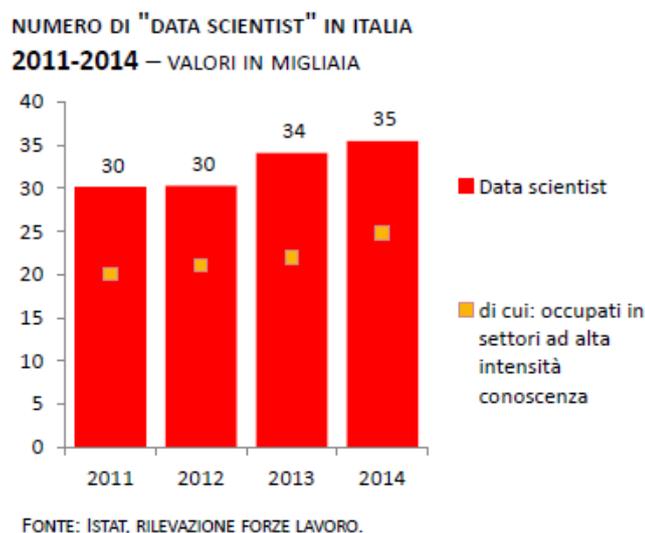


Figura 1. Numero di data scientist tra il 2011 e il 2014. Fonte ISTAT.

Il data scientist è una figura che ha capacità e competenze per adoperare strumenti che permettono l'analisi dei Big Data. Questa è una figura che sta prendendo piede anche in Italia negli ultimi anni. Il Data Scientist non opera esclusivamente nel panorama delle aziende private, ma anche in enti pubblici e di ricerca. I settori in cui sono più richiesti sono, però, la consulenza informatica e la produzione di software.

In figura, il numero di Data Scientist occupati e numero di Data Scientist occupati in settori ad alta conoscenza (dati ISTAT).

Attenzione! Questo materiale didattico è per uso personale dello studente ed è coperto da copyright. Ne è severamente vietata la riproduzione o il riutilizzo anche parziale, ai sensi e per gli effetti della legge sul diritto d'autore (L. 22.04.1941/n. 633)

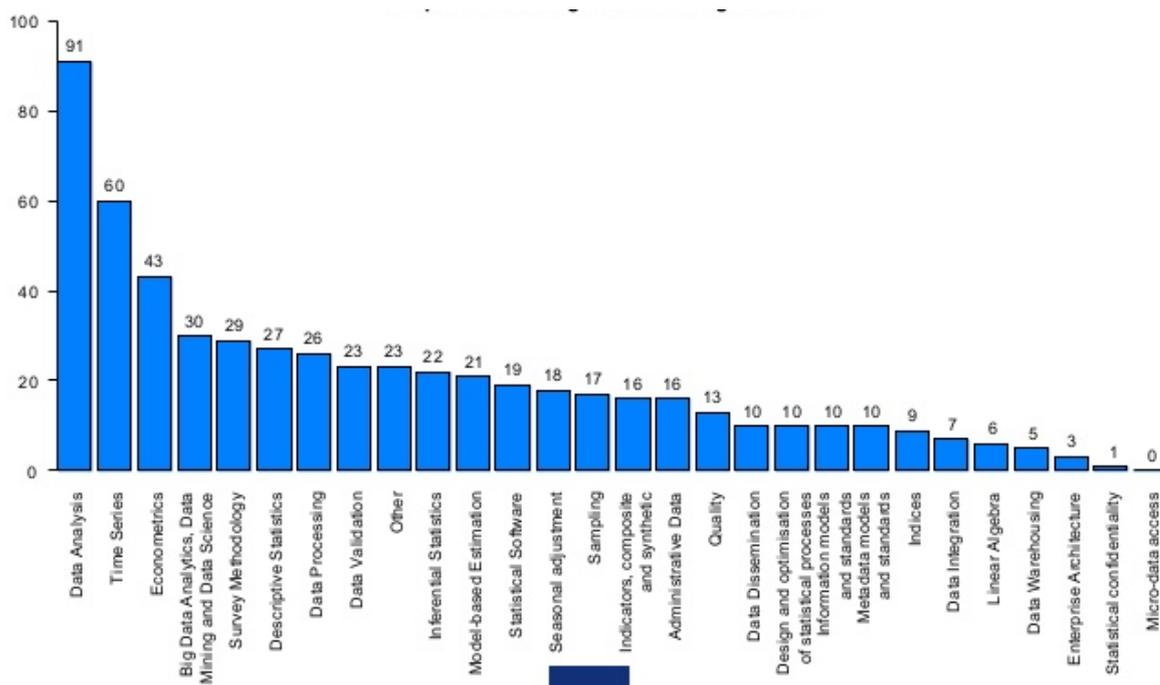


Figura 4. Principali figure nell'ambito degli studi statistici ed econometrici. Come si nota la Data Analysis è al momento al primo posto tra diffuse all'interno delle aree metodologiche (Fonte: Eurostat).

Attenzione! Questo materiale didattico è per uso personale dello studente ed è coperto da copyright. Ne è severamente vietata la riproduzione o il riutilizzo anche parziale, ai sensi e per gli effetti della legge sul diritto d'autore (L. 22.04.1941/n. 633)



Fonte: Istat, Rilevazione sulle forze di lavoro

Figura 3. Impiegati nel settore ICT in valore e in percentuale sull'occupazione (dati ISTAT).

Il *data scientist* è una figura rilevante che può essere utile alla risoluzione di problemi relativi alla raccolta, utilizzo, analisi e disposizione dei dati, ha competenze trasversali che spaziano dall'informatica, alle scienze statistiche ed economiche. Può essere impiegato nel coordinamento nella produzione di informazioni di qualità.

Attenzione! Questo materiale didattico è per uso personale dello studente ed è coperto da copyright. Ne è severamente vietata la riproduzione o il riutilizzo anche parziale, ai sensi e per gli effetti della legge sul diritto d'autore (L. 22.04.1941/n. 633)

RIFERIMENTI BIBLIOGRAFICI

- De Mauro A., Greco M., Michele Grimaldi M (2016). A Formal definition of Big Data based on its essential features. *Library Review*, 65: 3, 122-135.
- Rezzani A. (2013). *Big Data: Architettura, tecnologie e metodi per l'utilizzo di grandi basi di dati*. Maggioli Editore.
- Russom, P. (2011). Big data analytics. *TDWI best practices report, fourth quarter, 19(4)*, 1-34.
- Sahay, B. S., & Ranjan, J. (2008). Real time business intelligence in supply chain analytics. *Information Management & Computer Security*, 16(1), 28-48.

Attenzione! Questo materiale didattico è per uso personale dello studente ed è coperto da copyright. Ne è severamente vietata la riproduzione o il riutilizzo anche parziale, ai sensi e per gli effetti della legge sul diritto d'autore (L. 22.04.1941/n. 633)