Le professioni dell'innovazione

Il Data Scientist

A cura della Struttura "Orientamento al lavoro e digitalizzazione"

Camera di Commercio Roma



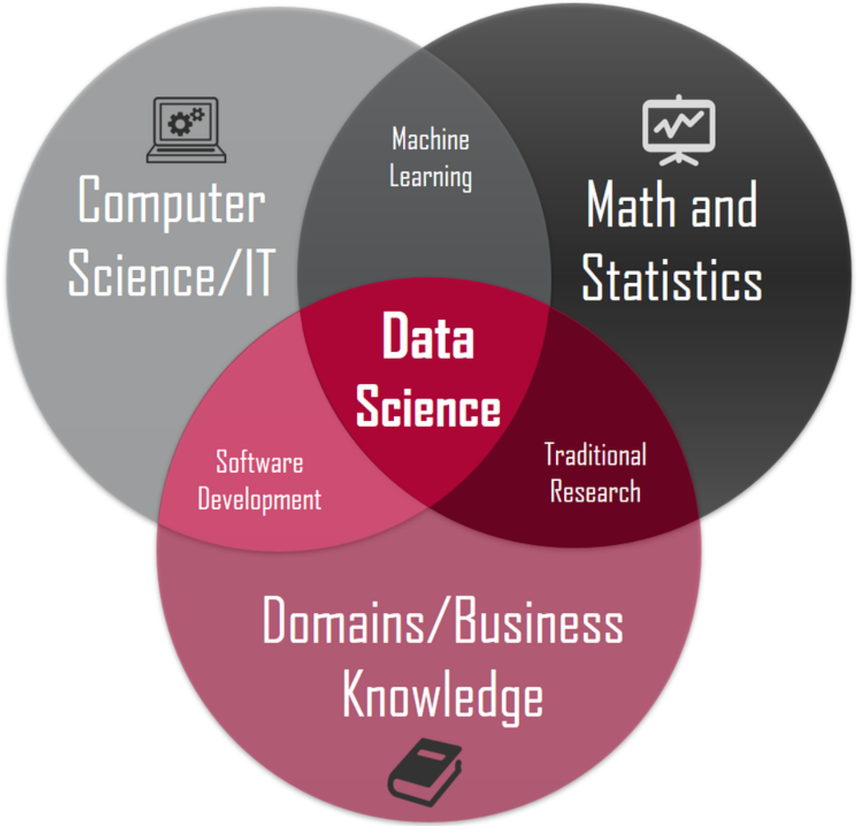
Indice

- Di cosa si occupa la *data science*
- Come la *data science* sta trasformando le imprese
- Chi è e cosa fa il *data scientist*
- *Soft* e *hard skills*
- Percorso formativo del *data scientist* e mercato del lavoro
- Normativa

Di cosa si occupa la *data science* e come sta trasformando le imprese



Di cosa si occupa la scienza dei dati?



La *data science* **è la scienza** che **estrae valore dai dati.** Si basa quindi su metodi scientifici, principalmente statistici, e **tecniche di analisi per trasformare un dato in una informazione utile** per un determinato contesto e determinati obiettivi.

L'**oggetto** della *scienza dei dati* sono i ***big data***: dati dal volume così grande da richiedere tecniche e metodi specifici per poter essere analizzati.

Come la *scienza dei dati* sta trasformando le imprese

Le imprese utilizzano la *scienza dei dati* per trasformarli in un **vantaggio competitivo.**

Ad esempio :

Manutenzione predittiva:
Le **imprese manifatturiere** riescono ad ottimizzare i processi produttivi attraverso la manutenzione predittiva, *cioè utilizzando sistemi di intelligenza artificiale , riescono a prevedere possibili problemi nella catena di produzione anticipando la loro insorgenza.*

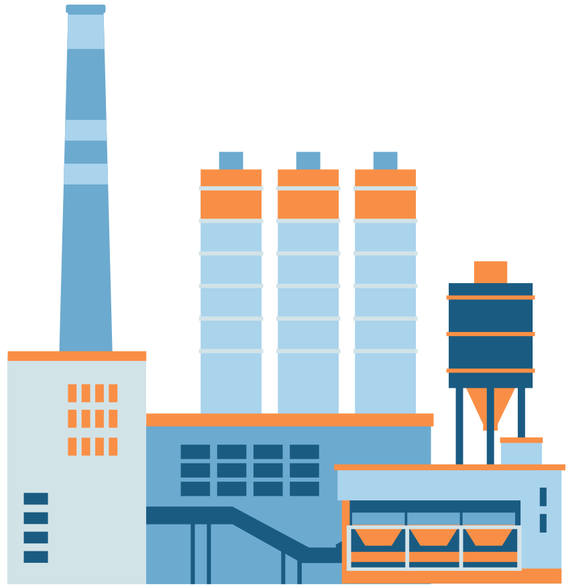


Personalizzazione dell'offerta:
Le imprese, utilizzando i dati, riescono a profilare i propri clienti e personalizzare i propri prodotti in base alle loro esigenze per *migliorare l'esperienza d'acquisto degli stessi.*

In quali settori economici viene utilizzata?



Secondo quanto riportato dall'**Osservatorio Big Data & Business Analytics del Politecnico di Milano,** nel 2020 in Italia le **banche** sono state il primo settore per quota di mercato *analytics* (28%), seguite da **manifattura** (24%), **telco e media** (14%), **servizi** (8%), **Gdo e retail** (7,5%), **assicurazioni** (7%), ***utility*** (6.5%), **PA e sanità** (5%).



Chi è e cosa fa il *data scientist*



Chi è il *data scientist* ?

(Repertorio Regionale delle competenze e dei profili formativi - Regione Lazio)

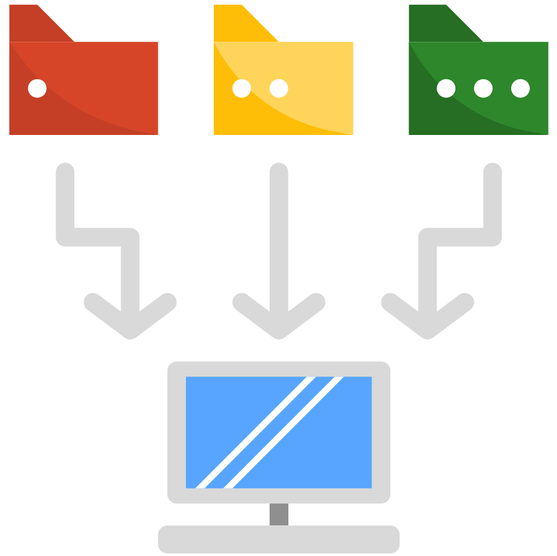


Il *data scientist* è colui che si occupa di **esplorare, raccogliere, visualizzare e analizzare i dati al fine di sviluppare modelli complessi** per raggiungere determinati obiettivi, **utilizzando metodi e strumenti efficaci** come ad esempio **linguaggi di programmazione.**

Cosa fa il *data scientist* ?

(Repertorio Regionale delle competenze e dei profili formativi - Regione Lazio)

Individua, sviluppa e applica metodi e **tecniche di raccolta, organizzazione e standardizzazione di dati, sui quali applica protocolli di *data mining*** (tecniche e metodologie che hanno per oggetto l'estrazione di informazioni utili da grandi quantità di dati), rivolti alla produzione di *analytics* descrittivi, diagnostici, predittivi e prescrittivi.



Cosa fa il *data scientist* ?

(Repertorio Regionale delle competenze e dei profili formativi - Regione Lazio)

Crea modelli per rappresentare problemi complessi a supporto della presa di decisioni, progettando modalità di analisi matematica e statistica, relative a dati a differente strutturazione, fra cui i big data.



Cosa fa il *data scientist* ?

Implementa algoritmi e *software* di analisi e *machine learning*.

Comunica i risultati delle analisi, individuando e **applicando tecniche di *rappresentazione dei dati*** (sviluppa Dashboard e Reportistica).





Camera di Commercio
Roma

Soft skills

1. **Comunicazione:** la più importante tra le *soft skills* in ogni settore, grazie alla quale è possibile sfruttare al meglio tutte le altre competenze.
2. **Lavoro di squadra:** la **capacità di lavorare in armonia** in un gruppo è fondamentale per il *data scientist*. Si tratta di saper accogliere *feedback* e condividere con altri le proprie conoscenze.
3. **Analisi critica:** capacità di analisi di problemi complessi.
4. **Lingua inglese** di settore.
5. **Programmazione:** è una delle capacità più importanti del *data scientist*, poiché gli permette di **crearsi i propri strumenti e personalizzare quelli esistenti.**



Hard skills

(Repertorio Regionale delle competenze e dei profili formativi - Regione Lazio)

1) Raccolta dati e applicazione di metodi e tecniche di *data mining*

Eseguire, sulla base delle esigenze di analisi e delle sorgenti informative disponibili, **operazioni di raccolta dati, loro organizzazione e *data mining*,** individuando e applicando metodi e tecniche matematiche e statistiche.



2) Sviluppo dei processi di data analytics

Pianificare e coordinare i processi di *data analytics*, assicurando il **corretto flusso di informazioni (*feeding*) verso i livelli strategici e decisionali.**



Hard skills

(Repertorio Regionale delle competenze e dei profili formativi - Regione Lazio)

<u>3) Realizzazione di analisi complesse su *big data*</u>
Realizzazione di analisi avanzate di *data set* (anche di notevoli dimensioni e di *big data*), **creando modelli, algoritmi e metodi di raccolta, standardizzazione e analisi dati**, realizzando **reportistica a supporto dei processi decisionali.**



Skill 1: raccolta dati e applicazione di metodi e tecniche di *data mining*

ABILITÀ

- Definire, sulla base delle esigenze aziendali e dei sistemi informativi, i fabbisogni di *data mining*.
- Identificare e caratterizzare le sorgenti informative.
- **Svolgere operazioni di raccolta, trasformazione e configurazione dei dati (*Data Gathering*, *Data Transformation* e *Data Configuration*), sulla base di caratteristiche ed eterogeneità delle fonti grezze, utilizzando il linguaggio SQL e altri strumenti di interazione con basi di dati.**
- Selezionare i dati, in ragione degli obiettivi di analisi.
- Preparare i dati selezionati, eliminando possibili errori e stabilendo dei meccanismi di comportamento in caso di dati mancanti, anche attraverso approcci campionari.
- Applicare tecniche di *data mining*, anche adattando gli algoritmi sulla base delle esigenze di analisi.
- Tracciare le operazioni compiute e valutarne gli esiti, anche in vista del consolidamento del processo.
- Garantire proprietà intellettuale e rispetto della privacy.



Tipologia dati 1/2

Big data: 5V - volume, velocità, varietà, valore e veridicità.

I dati strutturati sono organizzati per categorie e possono essere **ordinati, letti e organizzati automaticamente da un sistema**, come ad esempio i dati sul traffico di un sito web, cifre di vendita, coordinate bancarie o coordinate GPS raccolte da uno *smartphone*.

I dati non strutturati sono più difficili da classificare in modo automatizzato e per questo hanno dei costi maggiori: **si tratta di input di persone,** come per esempio **recensioni dei clienti, email, video, messaggi sui social media e** così via.



Tipologia dati 2/2

Generati dagli utenti (*Human generated)*, ossia **rilasciati da un utente** su una piattaforma social o su qualsiasi altro sito attraverso il login;

Generati dalle macchine (*Machine generated)*, vale a dire **prodotti da sorgenti** quali sensori Gps, centrali di monitoraggio di eventi meteorologici, ecc;

Generati da entrambi (*Business generated)*, cioè un **mix** tra dati *human generated* e dati *machine generated*, al fine di generare format usufruibili.



Skill 2: realizzazione di analisi complesse e su *big data*

ABILITÀ

- Standardizzare e ottimizzare i processi di gestione e aggiornamento dei dati, implementando procedure di cura, controllo della qualità, integrità e interoperabilità del dato.

- **Creare modelli, algoritmi, strumenti per la raccolta e l'analisi di big data, sulla base delle esigenze aziendali, utilizzando linguaggi di programmazione.**

- Sviluppare, sulla base delle esigenze, algoritmi di *machine learning*.

- Utilizzare piattaforme di *big data analytics*.

- Rappresentare gli esiti dell'analisi, utilizzando tecniche e strumenti di rappresentazione dei dati (*data visualization)*.

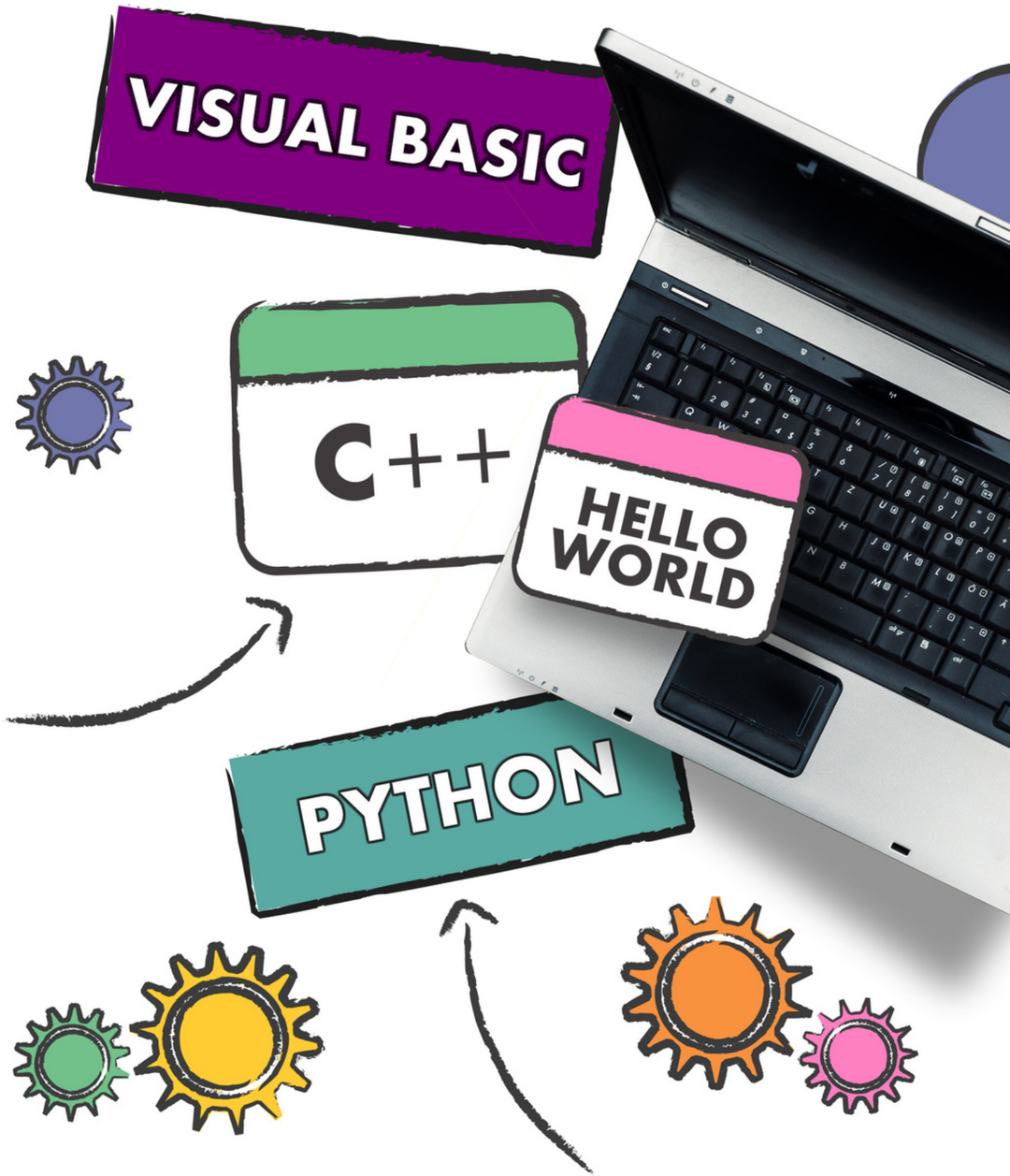


Linguaggi più usati dai *data scientist*

SQL (Structured Query Language) è il linguaggio di interrogazione più diffuso tra quelli usati per **l'interazione con i principali Database Management Systems** (DBMS), soprattutto relazionali.

Python è uno dei linguaggi più popolari: **open source, orientato all'utente** e di facile utilizzo. Esiste dal 1991 ed è dinamico, generico e orientato a oggetti.

R è un linguaggio di programmazione di alto livello tipicamente utilizzato in **ambito statistico**, molto utile per **l'elaborazione statistica** e la **visualizzazione dei dati in forma grafica.**



Skill 3: sviluppo dei processi di *data analytics*

ABILITÀ

Pianificare i processi di *data analytics* di una organizzazione, sulla base delle proprie esigenze strategiche.

Definire e implementare il modello organizzativo dei processi di *data analytics*.

Trasferire gli esiti delle analisi ai livelli aziendali in cui avviene l'elaborazione delle strategie di *business*.

Monitorare i processi di *data analytics* e il loro impatto sul *business*.



Percorso formativo e mercato del lavoro



Come diventare un *data scientist*

- Laurea magistrale in *data science*.
- Laurea specialistica in Matematica, Ingegneria, Fisica, Informatica, Statistica, Economia.
- Master in *data science*.
- Competenza nella programmazione con linguaggi orientati all'analisi statistica dei dati, a partire da R, Python e SQL, e conoscenze di *Analytics* e *Machine Learning*.

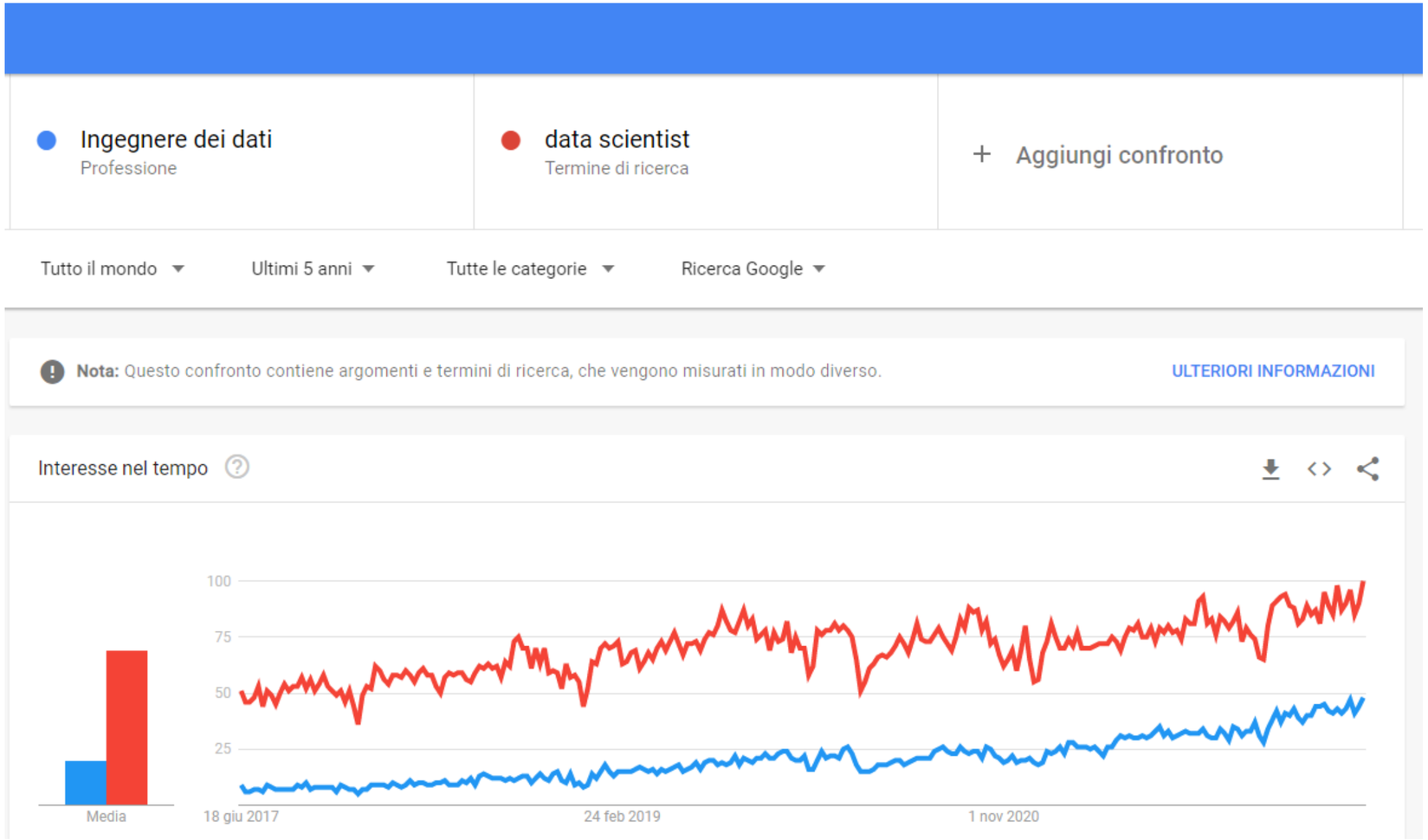


Richiesta sul mercato del lavoro

Nonostante la crescente richiesta sul mercato di figure specializzate e la grande necessità per le aziende di tradurre i dati in un vantaggio competitivo, **esiste un ampio gap tra domanda e offerta**, il quale fa dei *data scientist* delle figure con una **forte capacità contrattuale** in fase di trattativa.



Ricerche su Google trends



Normativa

Legge 14 gennaio 2013 n. 4
Disposizioni in materia di professioni non organizzate.



Profili professionali relativi alla professionalità operanti nel Web (UNI 11621-3) *"Attività professionali non regolamentate – Profili di ruolo professionale per l'ICT – Parte 3: Profili di ruolo professionale relativi alle professionalità operanti nel web".*

Repertorio Regionale delle competenze e dei profili formativi - Regione Lazio
istituito con la Delibera Giunta Regionale n. 452/2012 ed adeguato a quanto previsto dal D.lgs. 13/13 e dal Decreto Interministeriale del 30 giugno 2015, che istituisce il Quadro Nazionale delle Qualificazioni regionali.

LINEE GUIDA - Agid
per la qualità delle competenze digitali nelle professionalità ICT.

Camera di Commercio di Roma

via de' Burrò 147

www.rm.camcom.it

orientamentoedigitalizzazione@rm.camcom.it