



Società
Italiana di
Statistica



Statistica e data science nuove sfide e prospettive future

Corrado Crocetta
Presidente SIS

StatCities 2023

Olbia, Museo Archeologico 15 giugno 2023

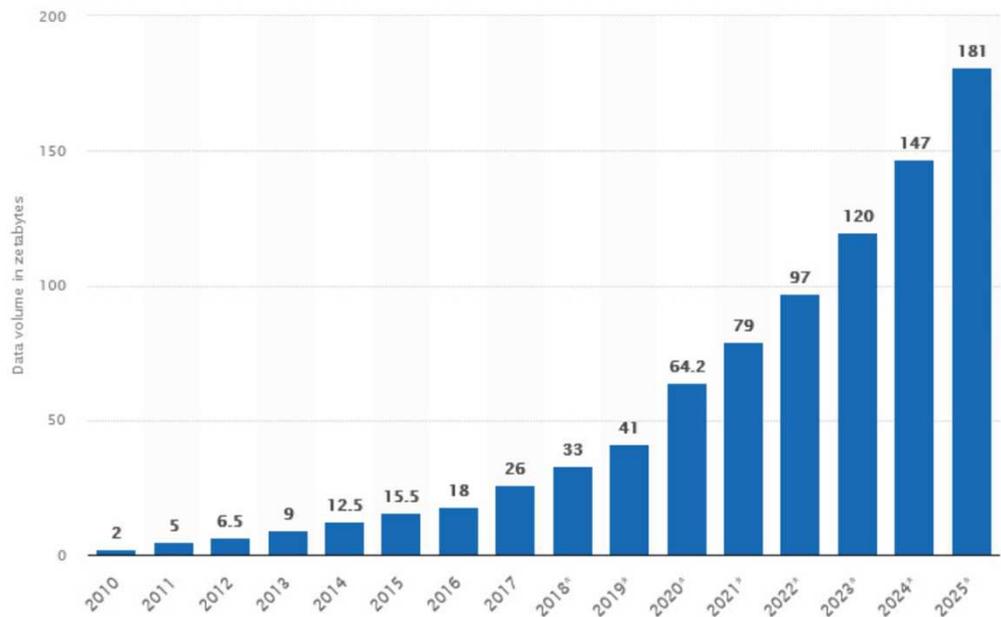


Schema della presentazione

- Data deluge e Big data
- Statistica, Data science e Industria 4.0
- La Data science ai tempi dei big data:
 - rischi
 - opportunità
- Le opportunità di crescita di uno statistico e di un data scientist
- L'accreditamento professionale dello statistico.
- La certificazione delle competenze statistiche: uno sguardo al futuro.

Data deluge e Big Data

Nel 2020 sono stati prodotti, raccolti, archiviati, copiati ed elaborati per diverse finalità più di 64 zettabytes di dati in tutto il pianeta.

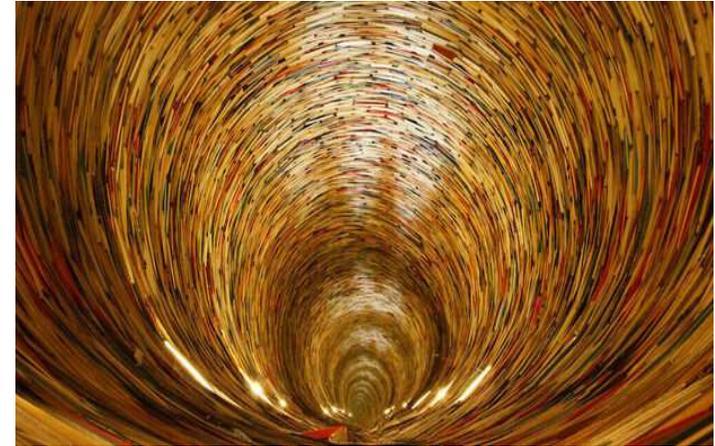


Nome	Simbolo	Multiplo
chilobyte	kB	10^3
megabyte	MB	10^6
gigabyte	GB	10^9
terabyte	TB	10^{12}
petabyte	PB	10^{15}
exabyte	EB	10^{18}
zettabyte	ZB	10^{21}
yottabyte	YB	10^{24}

L'era dei Big Data

«L'universo (che altri chiamano la Biblioteca) si compone d'un numero indefinito, e forse infinito, di gallerie esagonali, con vasti pozzi di ventilazione nel mezzo, bordati di basse ringhiere. Da qualsiasi esagono si vedono i piani superiori e inferiori, interminabilmente»

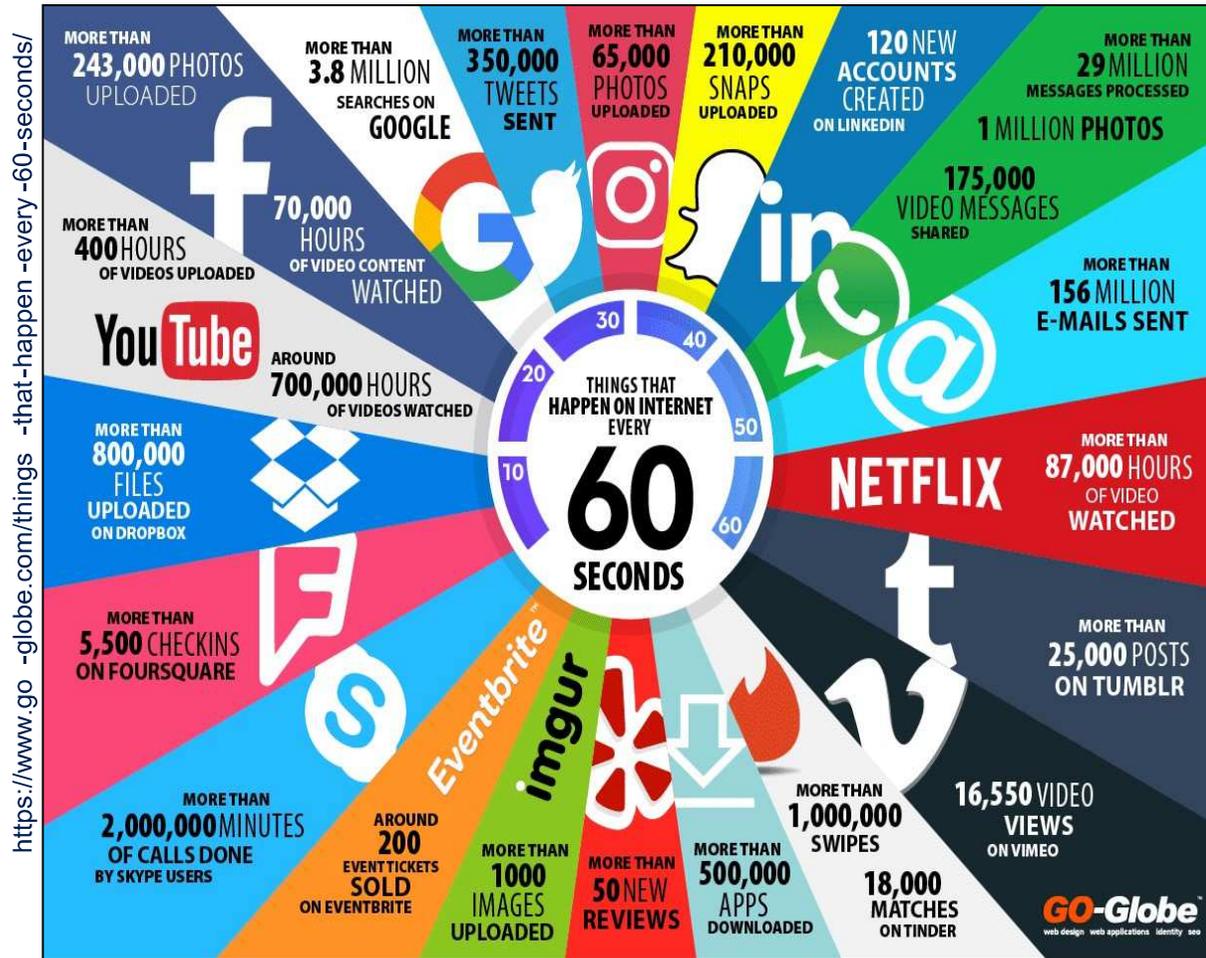
Jorge Luis Borges, 1941, "Il giardino dei sentieri che si biforcano".



Matej Kren. IDIOM Biblioteca Municipale di Praga

Nell'era dei **Big data**, la biblioteca di *Babele* del romanzo, in cui ipotizza di poter avere accesso a tutti i libri che sono stati, sono, saranno pubblicati evoca l'utopia moderna di rincorrere la conoscenza disponendo dell'universo dei dati.

L'era dei Big Data



Caratteristiche:

Volume

Velocità

Varietà

Valore

Veridicità

Variabilità

Cosa sono i Big Data

Modello delle cosiddette “3 V”:

- **Volume** (*data size*),
- **Variety** (*different forms of data sources*),
- **Velocity** (*speed of change*), che si sostanzia nella rapidità con cui i dati si generano, si raccolgono e si elaborano.

A queste tre caratteristiche di base se ne sono aggiunte ulteriori due che sono:

Veracity (*uncertainty of data*) riguarda l'accuratezza o **veridicità** di un *set* di dati (caratteristica importante se deve essere usato come fonte statistica ufficiale).

Value (*obtainable information*) riguarda la possibilità di estrarre da dati così complessi informazioni che **abbiano un significato e siano funzionali**. Per verificare questo aspetto servono tecnologie sofisticate e competenze interdisciplinari.

Vi sono altre V idonee a caratterizzare i *Big data* (Arockia *et al.*, 2017) come:

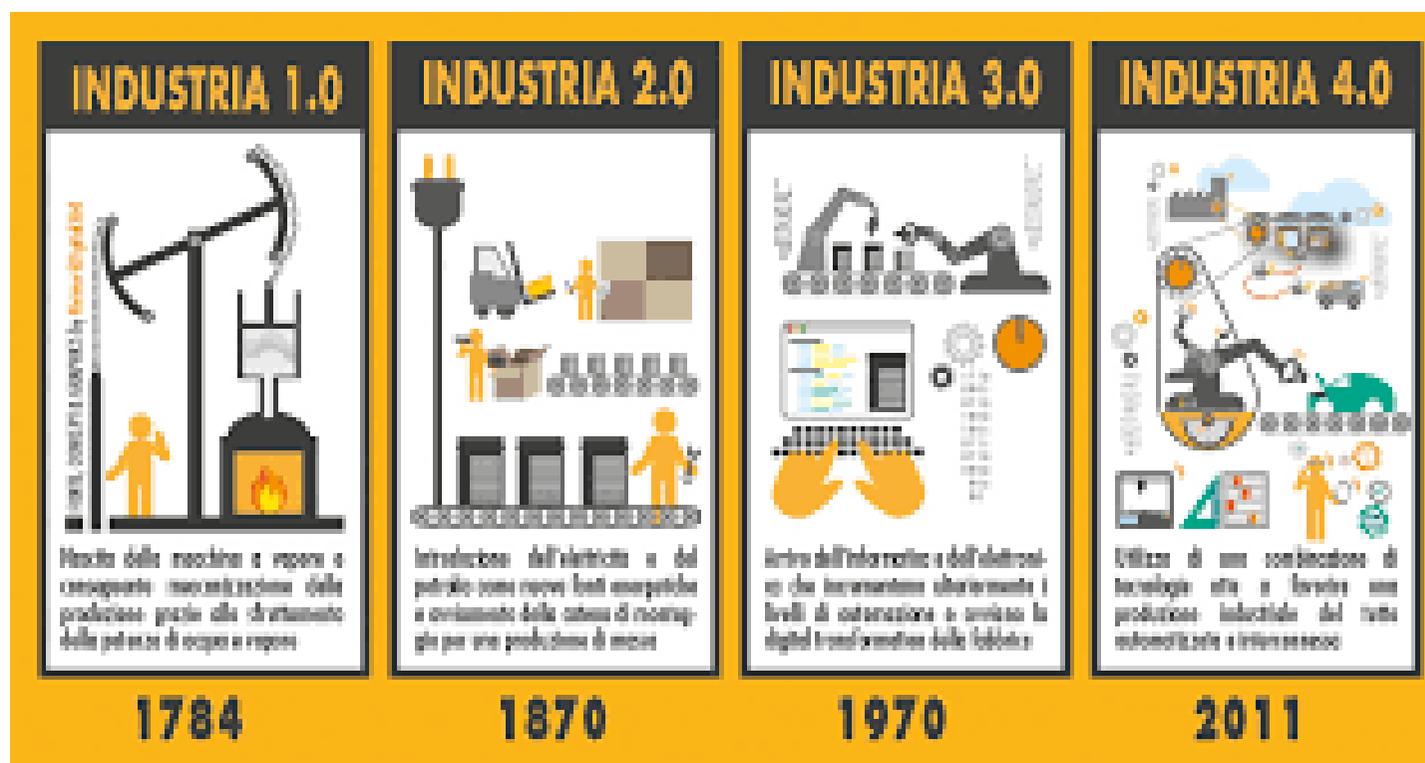
Visualization complessità di visualizzazione dei dati,

Variability impatto che deriva dalla mutabilità dei dati,

Virality velocità di diffusione e di trasmissione dei dati tra gli utenti.



La quarta rivoluzione industriale e i Big Data



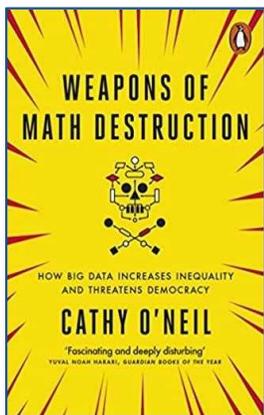
Industria 4.0 si caratterizza per lo sviluppo di: intelligenza artificiale (IA), robotica, Internet delle Cose (IoT), stampa 3D, ingegneria genetica, computer quantistici e altre tecnologie.

**Le tecniche data-driven
possono apportare grandi
benefici per:**

- **Cittadini**
- **Imprese**
- **Pubblica amministrazione**



Ma anche grandi rischi

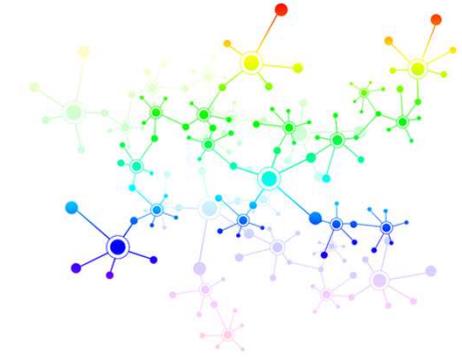


Ma possono esserci dei rischi relativi a:

- Diritti delle persone
- Sicurezza e salute delle persone
- Democrazia della società

Attualità	Descrizione
Micro-targeting comportamentale: caso Cambridge Analytica (USA)	Profiling psicologico effettuato per veicolare informazioni tali da condizionare l'opinione degli elettori americani durante la campagna politica
Giustizia predittiva: caso COMPAS	Piattaforma di supporto ai giudici americani nel calcolo della probabilità di recidiva da parte del reo. I neri sono visti, come probabilità, più pericolosi di quanto non siano, e i bianchi meno pericolosi.
Marketing: il mercato dei dati con i cookie terze parti	Circa il 90% dei cookie terze parti viene utilizzato per il web tracking e allo scopo di proporre pubblicità mirata. Oltre il 70% dei cookie nascondono altro codice software per finalità non dichiarate
Processi decisionali automatizzati nella PA: caso «buona scuola» (Italia)	Circa 10mila insegnanti sono stati trasferiti ingiustamente, rispetto ai criteri di merito, da una procedura automatizzata

La statistica al tempo dei big data



La **statistical literacy** prevede l'utilizzo critico e consapevole dati provenienti da fonti affidabili e di metodi quantitativi e strumenti grafici. E' uno strumento fondamentale per assumere decisioni razionali in condizioni di incertezza

La statistica è la sorella maggiore della matematica: è troppo saggia per dare tutto per certo. [WSD 2010](#)

Se torturi i dati abbastanza, alla fine confesseranno quello che vuoi [Darrell Huff](#)

You Don't Need Big Data — You Need the Right Data [Maxwell Wessel](#)

Futures studies, costruzione di scenari con testimoni privilegiati, sfruttare le conoscenze a priori dei fenomeni

DATA SCIENTIST: Chi era costui ?

La parola Data Scientist è stata coniata nel 2008 da DJ Patil and Jeff Hammerbacher (founder and Chief Scientist of Cloudera)

A new professional figure capable of carrying out analyses on large masses of data in order to extract relevant information which can bring value to the company in which he/she works.

Deve avere le competenze necessarie per processare big data o data warehouse, utilizzando modelli statistici e tecniche di machine learning e software specifici ma deve anche possedere solide conoscenze nello specifico dominio di applicazione.

Corsi universitari di statistica e data science

Nell'a.a 2022-23 presso le università italiane vi sono:

21 corsi L-41 Statistica

15 corsi LM- 82 Scienze statistiche

10 corsi LM -83 Scienze statistiche e attuariali

11 corsi LM data Data science



Il programma di accreditamento per statistici e biostatistici

La Società Italiana di Statistica con la SISMEC ha realizzato un programma di **accreditamento** per valorizzare la professionalità statistica.

L'accreditamento è una forma di **riconoscimento delle capacità professionali**.

E' su base **volontaria** e mira a valorizzare le **competenze statistiche** maturate in ambito professionale e lavorativo anche se non si possiedono titoli di studio specifici.

La **SIS** che è la più importante società scientifica in ambito statistico in Italia attesta che il candidato possiede i **requisiti necessari** per essere accreditato.

Chi si rivolge ad un statistico accreditato sa che **possiede le competenze** statistiche dichiarato.

L'accreditamento **vale per 5 anni**. Può essere rinnovato.



Come funziona l'accreditamento?

SIS e SISMEC, società scientifiche di riferimento per la Statistica in Italia, attestano, sulla base di criteri trasparenti, che la persona che ha chiesto di essere accreditata possiede:

- una solida preparazione accademica,
- una pluriennale esperienza lavorativa,
- le necessarie capacità comunicative.

Si impegna a:

- svolgere attività di formazione continua
- rispettare il codice etico della SIS



L'accreditamento SIS-SISMEC e riconosciuto a livello internazionale

L'accreditamento come forma di riconoscimento e valorizzazione è molto diffuso nel mondo ed è promosso anche **dall'International Statistical Institute**

La SIS è *accreditation partner* **dell'American Statistical Association** e collabora attivamente al programma di accreditamento europeo, promosso dalla **Federation of National Statistical Societies (FENStatS)**.

Cosa avverrà in futuro?

Seguendo l'esempio dei sistemi di certificazione delle competenze linguistiche (TOEFL, IELTS, esami Cambridge English) e informatiche (esempio la [certificazione Microsoft Office Specialist, MOS](#)), mediante test standardizzati si vuole misurare la conoscenza delle competenze di statistica e data science di una persona, rilasciando un open badge.

La SIS ha avviato una collaborazione strutturata con AUSED.

Vi è una domanda importante di certificazione di competenze statistiche da parte di biostatistici chiamati a far parte di comitati etici, di uffici di statistica di enti pubblici, di esperti di aziende private.





Grazie per l'attenzione

corrado.crocetta@uniba.it