

STATISTICA - LEZIONE 1

Società dell'informazione la STATISTICA: fornisce gli elementi utili per raccogliere, analizzare, verificare e interpretare i dati disponibili

Lo Statistico

raccoglie le informazioni di interesse su un particolare fenomeno
valuta l'affidabilità

la sintetizza

evidenzia gli aspetti salienti

ricerca le RELAZIONI con altri fenomeni di interesse

interpreta le informazioni e interpreta le relazioni individuate

al fine di prevedere comportamenti futuri e prendere decisioni opportune

(NB) La statistica rappresenta un approccio quantitativo alla conoscenza

- Non è una scienza esatta (leggi di probabilità)
- ≠ dalla matematica, ma usa molti strumenti
- Si interfaccia con altre discipline per:
 - studio dei fenomeni e caratterizzazione di relazioni tra fenomeni nei vari ambiti di ricerca
 - utilizzo di strumentazione informatica per la raccolta dati e l'analisi

Non è una disciplina a sé stante
è TRASVERSALE alle altre.

L'insieme di dati di enorme dimensione provengono da

- DATA WAREHOUSES (in ambito commerciale, bancario, finanziario, ...)
= "MAGAZZINO DATI" (immagazzinamento di dati) > es. variazioni della borsa
- DATI DI RACCOLTA AUTOMATICA di RICERCHE SCIENTIFICHE (genetiche, meteorologiche, astronomiche, ...)

importante è la disponibilità di STRUMENTAZIONE INFORMATICA, avanzata e veloce, in grado di sostenere procedure intensive di calcolo e processi in parallelo

Le tecniche moderne di analisi hanno quindi diversi punti di contatto con altre (più recenti) discipline, quali: Data Base Management, Data Base Technology, Intelligenza artificiale, Machine Learning, Data Visualization, Pattern Recognition

WALMART: fondata nel 1962 (multinazionale statunitense), + grande catena al mondo nelle grandi distribuzioni organizzate con:

- + 275 milioni di clienti che visitano ogni settimana gli oltre 11.600 negozi in 28 paesi
- e-commerce websites in 41 paesi
- guadagnare di 514,4 trilioni di \$ nel 2019
- Circa 2,2 milioni di collaboratori nel mondo, di cui 1,5 in USA
- Circa 2,5 petabytes di dati ottenuti da 1 milione di clienti ogni h!

↓
unità di misura dell'informazione o della quantità di dati

Le strategie di analisi statistica hanno lo scopo di:

- individuare strategie di offerta migliori negli store, online, via app
- valutare le strategie di stoccaggio dei prodotti (= metodo di conservazione di un prodotto in un deposito)
- prevedere le variazioni di acquisto in base alle condizioni meteorologiche
- valutare come eventi in singole zone possono alterare la domanda di prodotti
- proporre offerte definite per ogni cliente web sulla base della sua storia di acquisti online
- individuare nuovi prodotti da inserire negli store sulla base delle info circolanti su social media

Valutazione del rischio di credito si basa sulla storia del cliente nella banca. Se è buono o insolvente. Si costruisce una previsione sul comportamento futuro del nuovo cliente e decide se assegnare o meno il prestito

Software R programma usato dagli statistici

- sviluppato a fine anni '90 in Nuova Zelanda perché non c'erano software gratuiti
- è caratterizzato da molti aspetti positivi per l'utilizzo

Perché R?

- software più utilizzato per le analisi
- fornisce numerosissimi strumenti per l'analisi statistica dei dati aggiornati
- gestisce i modelli statistici anche molto complessi
- open source: a sorgente aperta
- multiplatforma
- interfacciabile con altri linguaggi di programmazione (es "C")
- chi utilizza R? Google, New York Times, Facebook, Twitter, Airbnb

www.r-project.org DISPENSA su MOODLE + manuali in R

"Download" si deve scaricare! (versione base)