

Scuola Superiore di Catania

CORSO DI LABORATORIO a.a. 2017/2018

Big data: nuovi paradigmi per l'analisi dei dati

Responsabile: Giovanni Gallo

Motivazioni

La digitalizzazione delle comunicazioni e dei servizi, la disponibilità a costi sostenibili di estese reti di sensori per il monitoraggio di quasi ogni aspetto dei fenomeni fisici e sociali ha reso disponibili per lo studio e l'analisi scientifica nuove tipologie di dati. Le caratteristiche principali di queste nuove sorgenti di informazioni sono:

- a) il formato digitale;
- b) l'elevatissimo numero di record raccolti;
- c) il superamento del concetto di "campionamento statistico" a favore della raccolta ed analisi sistematica e completa di "tutti" i dati;
- d) ricerca euristica di correlazioni tra fenomeni per la costruzione di modelli predittivi.

Obiettivi

Il corso si pone come obiettivi:

- a) fornire conoscenze di base nell'analisi "Big data";
- b) familiarizzare con gli strumenti di calcolo tipici del settore;
- c) far acquisire la capacità di utilizzare l'analisi "Big data" in modo critico e competente;
- d) fornire degli spunti per una riflessione metodologica, epistemologica e delle conseguenze politiche e sociali di questo nuovo paradigma per l'analisi dei dati;
- e) conoscere direttamente alcune delle esperienze pilota e delle applicazioni dei "Big data";
- f) svolgere attività pratiche di laboratorio/progetto di analisi di "Big data".

Destinatari

Il corso è orientato agli allievi della Classe Scientifica impegnati nella formazione Magistrale (secondo livello). Esso potrà anche rivolgersi ad allievi di primo livello ben motivati ed agli allievi della Classe Umanistica, specialmente se interessati agli studi sociali. Auguriamo anche che esso possa prevedere come uditori la presenza di un gruppetto selezionato di laureandi e dottorandi della Università di Catania che ne facciano richiesta, o come accaduto nella precedente edizione da uditori industriali.

Pre-requisiti

Le nozioni propedeutiche per una proficua frequenza del corso includono:

- a) Statistica elementare e esperienza, anche elementare, nella analisi dei dati;
- b) Capacità di base nella programmazione con strumenti di alto livello (MATLAB, R, C#);

Articolazione di massima e valutazione

Il corso prevede trenta ore di didattica frontale e laboratorio affidato a docenti ed esperti provenienti in massima parte dalla Università di Catania ed articolati in moduli dettagliati di seguito. A completamento della parte di formazione di base seguiranno interventi mirati di esperti internazionali invitati. La valutazione di profitto sarà fatta valutando la qualità dei risultati progettuali ottenuti dagli allievi.

Struttura modulare del corso

1. Introduzione sul tema Big Data, con esempi di applicazioni reali e strategie per attuare un progetto di big data (4 ore)
2. Algoritmi per i big data (4 ore)

3. Workshop di introduzione al software R (6 ore)
4. Workshop Hadoop e altri strumenti con sviluppo di progetto "analisi log di sito web" mirato anche all'utilizzo dei dati del progetto di weblog della SSC (12 ore)
5. Casi di studio dei big data per il monitoraggio della salute umana (4 ore)
6. Casi di studio dei big data per il law enforcement (2 ore)
7. Casi di studio dei big data per lo studio dei sistemi complessi (2 ore)
8. Casi di studio dei big data per la manifattura industriale di precisione (4 ore)
9. Contributi di esperti internazionali invitati: max tre conferenze di ore 2, con esperti nel settore.