UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Data Analytics

2122-2-F1801Q104

Obiettivi

I principali obiettivi del corso riguardano l'acquisizione di conoscenze relative alle principali tecniche di analisi di dati, sia strutturati che non strutturati, maturando competenze specifiche di data, networks e social media analytics. Gli studenti saranno inoltre in grado di progettare e realizzare applicativi per lo sviluppo di specifiche funzionalità di analytics (ad esempio, script in python ed R).

Contenuti sintetici

Introduzione al data analytics, con particolare riferimento a diverse tipologie di dati e diverse tecniche di analisi. Approfondimenti in specifici contesti di data analytics quali network analytics e social media analytics.

Programma esteso

1 Il processo di data analytics Tipi di dati: dati strutturati, semistrutturati e non strutturati Tipi di analisi: predittiva, prescrittiva, descrittiva

2 Introduzione alla Network Analysis: definizioni base e domini applicativi

3 Reti statiche e dinamiche: proprietà statistiche per l'analisi strutturale

4 Misure di centralità

5 Clustering su grafi: algoritmi ed applicazioni

6 Modelli di influenza e contagio nelle social reti

7 Introduzione al Social Media Analytics

8 Subjectivity Detection, Sentiment Mining, Irony Detection: lessici, modelli probabilistici del linguaggio, ensemble learning

9 Named-Entity Recognition and Linking: Conditional Random Fields, Probabilistic Topic Models, Word Sense Disambiguation

10 Social Media Tagging e Summarization: Tag Generation, Tag analysis, Summarization

11 Problematiche e tecniche di visualizzazione

Prerequisiti

Nessun prerequisito essenziale. Sono utili conoscenze di base di algebra lineare e teoria dei grafi.

Modalità didattica

Lezioni ed esercitazioni in aula. Attività di laboratorio. Svolgimento di esercitazioni e laboratori individuali ed in gruppo.

Il corso sarà erogato in Italiano.

Nel periodo di emergenza Covid-19 le lezioni si svolgeranno completamente da remoto in modalità sincrona e asincrona. Tutte le lezioni saranno registrate e rese disponibili agli studenti.

Materiale didattico

Albert-László BARABÁSI. **Network science**. Cambridge university press (2016).

D. Easley, J. Kleinberg. **Networks, Crowds, and Markets: Reasoning About a Highly Connected World**. Cambridge University Press (2010).

Charu C. Aggarwal . **Social Network Data Analytics**, Springer (2012).

N. Danneman, R. Heimann. Socia Media Mining with R. (2014).

Periodo di erogazione dell'insegnamento

Secondo semestre.

Modalità di verifica del profitto e valutazione

Progetto e orale. Sono assenti prove in itinere intermedie.

Il progetto consisterà nello sviluppo di uno strumento di analytics basato su metodi e modelli presentati a lezione. Il progetto prevede una valutazione espressa in un range 0-24. Per ciascun progetto verrà valutato:

- Metodologia adottata (modelli e metodi) : 7 punti

- Dimostratore e tecniche di visualizzazione adottate: 5 punti

- Analisi dei risultati sperimentali: 7 punti

- Presentazione: 5 punti

L'orale prevede 4 domande di teoria tra gli argomenti del corso elencati nel programma dettagliato. Per ciascuna domanda verrà data una valutazione compresa pari a -2, per una risposta errata o mancata risposta, e +2 punti per una risposta corretta.

Nel periodo di emergenza Covid-19 gli esami si svolgeranno completamente da remoto.

Orario di ricevimento

Su appuntamento.

Nel periodo di emergenza Covid-19 il ricevimento studenti si svolgerà completamente da remoto.