UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Data Analytics

2526-2-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 text analytics. Gli studenti saranno inoltre in grado di progettare e realizzare applicativi per lo sviluppo di specifiche funzionalità di analytics (Python).

Gli obiettivi specifici rispetto ai Descrittori di Dublino (DdD) sono i seguenti:
1. Conoscenza e capacità di comprensione
Lo studente acquisirà:

- Conoscenze teoriche sulle principali tecniche di analisi dei dati, strutturati e non strutturati;
- Comprensione dei fondamenti della Teoria dei Grafi e Natural Language Processing (NLP);
- Conoscenze degli strumenti di visualizzazione.

2. Conoscenza e capacità di comprensione applicate
Lo studente sarà in grado di:

- progettare e sviluppare applicazioni di analytics in Python;
- applicare metodi e modelli appresi per risolvere problemi concreti di data analytics;
- integrare tecniche di visualizzazione per rappresentare risultati in modo efficace.

3. Autonomia di giudizio
Lo studente svilupperà la capacità di:

- valutare criticamente l’efficacia delle tecniche di analytics adottate in funzione del tipo di dato e contesto applicativo;
- scegliere consapevolmente modelli e strumenti adeguati all’obiettivo analitico;
- interpretare i risultati sperimentali, analizzando implicazioni e limiti.

4. Abilità comunicative
Lo studente sarà in grado di:

- presentare in modo chiaro e strutturato i risultati delle proprie analisi tecniche;
- esporre la propria soluzione progettuale durante l'orale e giustificare le scelte metodologiche adottate;
- lavorare in gruppo nella realizzazione del progetto, contribuendo alla documentazione e alla presentazione.

5. Capacità di apprendere

- Lo studente sarà in grado di sviluppare capacità autonome di studio attraverso il completamento di assignment facoltativi derivanti da attività di laboratorio focalizzate all'uso di modelli e strumenti di network analytics e Natural Language Processing.

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 text analytics.

Programma esteso

Teoria dei Grafi

1. Reti statiche e dinamiche: proprietà statistiche per l'analisi strutturale
2. Misure di centralità
3. Clustering su grafi: algoritmi ed applicazioni (community detection)

Natural Language Processing

1. Fondamenti dell'elaborazione del linguaggio naturale
2. Modelli linguistici statistici e neurali
3. Word e Sentence Embedding
4. Transformers e meccanismi di attenzione
5. Large Language Models (ELMO, BERT, GPT e LLAMA)
6. Metodi di explainability per i modelli del linguaggio

Prerequisiti

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

Modalità didattica

- 12 lezioni da 2 ore svolte in modalità erogativa in presenza;
- 12 attività di laboratorio da 2 ore svolte in modalità erogativa in presenza;

Il corso sarà erogato in Italiano.

Materiale didattico

Albert-László BARABÁSI. Network science. Cambridge University Press.
Cristopher MANNING and Hinrich SCHÜTZE. Foundations of Statistical Natural Language Processing. MIT Press.
Daniel Jurafsky and James Martin, "Speech and Language Processing, 2nd Edition", Prentice Hall, 2008.

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): max 7 punti
- Dimostratore e tecniche di visualizzazione adottate: max 5 punti
- Analisi dei risultati sperimentali: max 7 punti
- Presentazione: max 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.

Orario di ricevimento

Su appuntamento.

Sustainable Development Goals