

## SYLLABUS DEL CORSO

### Data Analytics

2425-2-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 (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 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**
4. Basics of Natural Language Processing
5. Affective Computing: lessici e modelli neurali del linguaggio (word embeddings, sentence embeddings)
6. Named-Entity Recognition: Conditional Random Fields ed estensioni neurali
7. Topic Extraction: Latent Dirichlet Allocation, Neural Topic Models
8. Tecniche e strumenti di visualizzazione

## Prerequisiti

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

## Modalità didattica

- 8 lezioni da 2 ore svolte in modalità erogativa in presenza;
- 4 lezioni da 2 ore svolte in modalità interattiva da remoto (lezioni asincrone);
- 8 attività di laboratorio da 2 ore svolte in modalità erogativa in presenza;
- 4 attività di laboratorio da 2 ore svolta in modalità interattiva da remoto (lezioni sincrone).

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.

Fabio TAMBURINI. Neural Models for the Automatic Processing of Italian. Patron Editore.

## 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.

## **Orario di ricevimento**

Su appuntamento.

## **Sustainable Development Goals**

---