

# MODULO DI APPROFONDIMENTO



## TECNICHE DI TEXT PROCESSING PER L'ANALISI DEI DATI

25, 26 e 27 novembre 2020

### ► Text Processing in Python

- Pre-processing di testo: pulizia di testo con Pandas, stopwords, stemming
- Introduzione a espressioni regolari
- Matrice TF-IDF, similarità e classificazione
- Word embedding (Word2Vec)
- Librerie Python per manipolazione di testo: textblob, NLTK, spacy e gensim
- Hands-on ed esercizi

### ► Opinion Mining e Sentiment Analysis

- Introduzione al problema e soluzioni allo stato dell'arte
- Derivare opinioni, sentimento e soggettività da testi
- Hands-on ed esercizi

### ► Labeling automatico e categorizzazione di testo

- Introduzione al problema e soluzioni allo stato dell'arte
- Tecniche di machine/deep learning classificazione di testo (BERT)
- Latent dirichlet allocation (LDA)
- Hands-on ed esercizi