



UNIVERSITÀ  
DEGLI STUDI DI MILANO-BICOCCA

## SYLLABUS DEL CORSO

### Data Processing and Analysis

2223-3-E4102B090

---

#### Obiettivi

Il corso intende fornire le competenze (metodologiche e tecnici) e gli strumenti per la comprensione e la realizzazione di soluzioni per il processamento di Big Data (strutturati e non), attraverso l'uso degli algoritmi e tool di AI per l'estrazione e rappresentazione della conoscenza da dati reali. Inoltre, il corso intende fornire gli strumenti tecnici per la modellazione e realizzazione di data model in accordo con il paradigma NoSQL, focalizzando principalmente sui graph-database e databaseNoSQL. Infine, si forniranno competenze tecniche e metodologiche relativamente ad algoritmi di explainable AI per la comprensione di algoritmi di machine learning black box

#### Contenuti sintetici

Introduction to AI and Big Data Analytics

Getting knowledge from data

Modelling and Querying the Resulting knowledge

#### Programma esteso

- **Introduction to AI and Big Data Analytics**

1. Goal and rationale of AI. The relation between Big Data and AI
2. The value of knowledge – digital economy and data-driven decision making

- **Getting knowledge from data**

1. Word Embedding (Word2Vec, Doc2Vec, GLOVE, FastText, StarSpace)
2. Evaluate word embedding models (intrinsic vs extrinsic evaluation)
3. Topic Modelling through Python

**- Modelling and Querying the Resulting knowledge through NoSQL**

1. introduction to NoSQL data stores
2. graph-databases and graph-traversal query languages (Cypher)
3. Document Databases

- **Explainable AI (global and local interpretation models)**

1. Introduction to XAI, local/global interpretation models. model agnostic-specific algorithms
2. XAI techniques as in the state of the art (eg. LIME, SHAP, etc)

## **Prerequisiti**

Nessuno

## **Modalità didattica**

Lezioni frontali, seminari monotematici, esercitazioni, assegnamenti da svolgere a casa.

## **Materiale didattico**

Lezioni con l'ausilio di slide, laboratorio e casi applicativi. Articoli scientifici di riferimento saranno forniti dal docente. Il Software utilizzato sarà open-source

## **Periodo di erogazione dell'insegnamento**

1 marzo - 15 aprile 2021

## **Modalità di verifica del profitto e valutazione**

La modalità di verifica si basa su una prova scritta ed una eventuale prova orale.

La prova scritta si svolge al computer ed è composta da domande aperte e chiuse e risposta multipla su tutti gli argomenti del corso.

In sede di valutazione viene considerata la capacità dello studente di rispondere a quesiti specifici facendo riferimento agli aspetti teorici e pratici (mediante esempi) connessi all'argomento richiesto.

La prova scritta è comune sia per gli studenti frequentanti sia per i non frequentanti.

La prova orale è mirata ad accertare la conoscenza teorica dello studente sugli argomenti del corso. Saranno quindi valutate le capacità di ragionare e approfondire le tematiche proposte in sede di esame e il rigore metodologico del loro sviluppo.

## **Orario di ricevimento**

su Appuntamento

## **Sustainable Development Goals**

---