UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

# SYLLABUS DEL CORSO

## Data Science Lab

**2526-1-FDS02Q003**

### Obiettivi formativi

Rendere gli studenti in grado di affrontare l'impostazione, la conduzione e la gestione di progetti di data science ed estrazione di informazioni da sistemi complessi di dati.

Il corso si concentra in particolare sull'impostazione e conduzione di progetti di estrazione dell'informazione in ambiti "aperti", cioè in assenza di obiettivi analitici precisi, dove i flussi dati costituiscono la base per l'ideazione, il design e l'implementazione di nuovi servizi, richiedendo al "data scientist" capacità creative, abilità organizzativa e comunicativa e solidità metodologica.

Sono previste anche lezioni teoriche di approfondimento su algoritmi e metodologie per il trattamento di strutture dati di particolare interesse per la pratica della data science (es. dati temporali o dati parzialmente ordinati).

OBIETTIVI SPECIFICI DI APPRENDIMENTO
Conoscenza e capacità di comprensione: dovendo affrontare un problema "aperto", gli studenti svilupperanno la capacità di comprensione e sintesi di fenomeni complessi.

Applicazione di conoscenze: attraverso lo sviluppo e la stesura del progetto, gli studenti impreranno ad applicare le conoscenze acquisite durante i propri studi.

Autonomia di giudizio: gli studenti dovranno valutare autonomamente il valore della propria domanda di ricerca e similmente individuare il percorso statistico più appropriato per rispondervi.

Abilità comunicative: attraverso lo sviluppo del progetto, gli studenti impareranno ad assumere la prospettiva del destinatario della ricerca.

Capacità di apprendere: poiché le metodologie statistiche richieste nello sviluppo del progetto dipendono dal tipo di domanda di ricerca e non sono preliminarmente stabilite, gli studenti impareranno ad acquisire nuove competenze, in itinere.

## Contenuti sintetici

Il corso illustra e introduce alla tipologia di attività che il data scientist è chiamato a svolgere nel contesto attuale, affronta la tematica della natura e della gestione di progetti di data science in ambito socio-economico e aziendale e si concreta nell'assegnazione di progetti di elaborazione dei dati (presi da ambiti reali) e generazione di contenuti informativi, che gli studenti, suddivisi in gruppi, devono condurre a termine nel periodo del corso, con la supervisione del docente.

## Programma esteso

Il corso si suddivide in due parti, Nella prima, più breve, si svolgeranno alcuni seminari per illustrare le caratteristiche tipiche dei progetti di data science nei contesti socio-economici e di business; nella seconda, più ampia, gli studenti verranno suddivisi in gruppi (di 3-4 unità), ai quali saranno assegnati progetti basati su dati reali, da sviluppare e completare durante il corso.

PARTE I

1. Il contesto tecnologico ed economico attuale: complessità dei processi socio-economici e nuove esigenza conoscitive e di servizi a valore aggiunto.

2. Complessità dei dati e nuove fonti: web, e-commerce, Internet of Things, Smartphone...

3. Processi basati sui dati in ambito aziendale: marketing, crm, operations...

4. Esempi di progetti di data science in ambiti aziendali e istituzionali.

5. Elementi di project management e caratteristiche/criticità dei progetti di data science: obiettivi, qualità dei dati, scelte tecnologiche.

6. Computabilità e limiti del machine learning.

7. La sintesi statistica di fenomeni complessi e la costruzione di indicatori sintetici.

8. Analisi dei dati parzialmente ordinati.

PARTE IIIdentificazione e assegnazione di progetti di data science (i contenuti dei progetti saranno definiti durante il corso).

## Prerequisiti

Non sono previsti prerequisiti formali, ma è necessaria una conoscenza delle tecniche di base della statistica inferenziale, dell'analisi multivariata e di data mining, nonché la conoscenza di base della programmazione R.

## Metodi didattici

Lezioni frontali e supervisione periodica dei progetti di elaborazione dei dati.

Tutte le attività didattiche si svolgeranno in modalità EROGATIVA, tranne quelle relative alla discussione sullo sviluppo delle attività di progetto, che saranno in modalità INTERATTIVA.

ORE DI DATTICA EROGATIVA: 35
ORE DI DIDATTICA INTERATTIVA: 13

## Modalità di verifica dell'apprendimento

L'apprendimento verrà verificato attraverso l'elaborazione di un progetto finale, senza prove intermedie (PROJECT WORK).

Il punteggio formale è determinato valutando: (i)la rilevanza della domanda di ricerca, (ii) l'appropriatezza della metodologia statistica utilizzata, (iii) la qualità dell'argomentazione e del discorso e (iv) la qualità editoriale del rapporto di progetto finale.

Questa modalità di verifica dell'apprendimento è motivata dall'obiettivo di mettere gli studenti nelle condizioni operative tipiche dell'attività lavorativa e di farne in particolare emergere le abilità soft (organizzative, comunicative, creative...).

## Testi di riferimento

Computers Ltd. What they really can't do, Harel D. Oxford University Press, 2000

Documenti e slide fornite dal docente, disponibili online.

## Periodo di erogazione dell'insegnamento

II semestre

## Lingua di insegnamento

English

## Sustainable Development Goals

SCONFIGGERE LA POVERTÀ | SALUTE E BENESSERE | RIDURRE LE DISUGUAGLIANZE | PACE, GIUSTIZIA E ISTITUZIONI SOLIDE