



UNIVERSITÀ  
DEGLI STUDI DI MILANO-BICOCCA

## SYLLABUS DEL CORSO

### Data Mining M

2122-2-F8204B014

---

#### Obiettivi formativi

Il corso si pone come obiettivo l'introduzione di tecniche avanzate per l'apprendimento supervisionato (*deep learning*) e il perfezionamento delle abilità di modellizzazione con finalità previsiva (*predictive modelling*), con relative implementazioni nell'ambiente di programmazione R. Il corso si divide in due parti:

1. **Introduction to Deep Learning** (3CFU, Prof. Borrotti);
2. **Applied Predictive Modelling** (3CFU, Prof. Solari).

#### Contenuti sintetici

##### **Introduction to Deep Learning**

Il modulo si pone l'obiettivo di introdurre i concetti fondamentali del *deep learning* permettendo la loro applicazione consapevole a problemi applicativi.

##### **Applied Predictive Modelling**

Il modulo si pone l'obiettivo di approfondire il processo di modellizzazione con finalità previsiva. Questo processo include la pre-elaborazione dei dati, la suddivisione in dati di training e di test, la messa a punto dei modelli e la stima dell'errore di previsione.

#### Programma esteso

##### **Introduction to Deep Learning**

Introduzione al Deep Learning  
Reti Neurali

Reti Neurali Shallow e Deep  
Recurrent Neural Network  
Convolutional Neural Network

### **Applied Predictive Modelling**

Principi fondamentali: sovra-adattamento, compromesso tra distorsione e varianza, ottimismo

Metodi di *ensemble*

Il modello e il processo di modellizzazione

### **Prerequisiti**

Si consiglia la conoscenza degli argomenti trattati nei corsi *Probabilità e Statistica Computazionale M* e *Statistica Avanzata M*.

### **Metodi didattici**

Le lezioni si svolgono sia in aula che in laboratorio, integrando aspetti di carattere teorico con quelli pratico-applicativi di analisi dei dati e di programmazione in R.

### **Modalità di verifica dell'apprendimento**

#### **Introduction to Deep Learning**

La modalità di verifica consiste nella presentazione di un progetto applicativo concordato con il docente. Il progetto si compone di

- Report
- File Rmarkdown contenente tutto il codice utilizzato per ottenere i risultati
- Slides della presentazione

La presentazione del progetto ha come obiettivo la verifica delle abilità di analisi e di comunicazione dei risultati. Agli studenti frequentanti viene data l'opportunità di svolgere il progetto in gruppi di massimo 4 persone. Gli studenti non frequentanti devono contattare il docente per l'assegnazione del progetto almeno 3 settimane prima della data dell'esame.

#### **Applied Predictive Modelling**

La modalità di verifica consiste nell'analisi di un dataset per il quale bisogna prevedere le osservazioni dei dati di test sulla base dei dati di training. Oltre alle previsioni, bisognerà produrre una relazione contenente la descrizione dell'analisi e il codice utilizzato da consegnare entro la scadenza prevista (almeno una settimana prima dell'appello d'esame). Sarà possibile consegnare le previsioni una volta sola per A.A. Qualora lo studente (oppure il docente) richieda la prova orale, il voto finale è una media dei voti dell'analisi dei dati e della prova orale. L'analisi del dataset ha come obiettivo la verifica delle abilità di modellizzazione dei dati a fini previsivi.

## Testi di riferimento

### Introduction to Deep Learning

- Efron, Hastie (2016) *Computer-Age Statistical Inference: Algorithms, Evidence, and Data Science*. Cambridge University Press.
- Bishop (2009) *Pattern Recognition And Machine Learning*. Springer.
- Goodfellow, Bengio, Courville (2017) *Deep Learning*. MIT.
- Chollet, Allaire (2018) *Deep Learning with R*. Manning
  
- Altro materiale fornito dal docente

### Applied Predictive Modelling

- Azzalini, Scarpa (2004). *Analisi dei dati e data mining*. Springer-Verlag Italia
- Kuhn, Johnson (2019). *Feature Engineering and Selection*. Chapman and Hall/CRC
- Kuhn, Silge (2021+). *Tidy Modeling with R*. (in progress)
- Lewis, Kane, Arnold (2019) *A Computational Approach to Statistical Learning*. Chapman And Hall/Crc.

## Periodo di erogazione dell'insegnamento

Primo semestre, secondo periodo.

## Lingua di insegnamento

Le lezioni si svolgono in italiano, tuttavia i libri di testo sono in lingua inglese.

---