

## SYLLABUS DEL CORSO

### Machine Learning M

2324-1-F8204B006

---

#### Obiettivi

Lo studente apprenderà le tecniche di Machine Learning più efficaci, comprendendo i fondamenti teorici di ogni tecnica e acquisendo il *know-how* per poterle applicare con successo alla risoluzione di problemi pratici. Sarà inoltre fornita una panoramica sulle più innovative soluzioni per l'identificazione del miglior algoritmo di Machine Learning e della sua configurazione ottimale (Automated Machine Learning – AutoML), dato un dataset. Lo strumento di riferimento per il corso sarà R, ma verranno anche presentate alcune soluzioni equivalenti in Python (ad esempio scikit-learn) e Java (ad esempio WEKA, KNIME).

#### Contenuti sintetici

Concetti basi del Machine Learning: tipologie di dati, *istanze*, *features*, *tasks* e *scenarios*, *parametri* e *iper-parametri*, misure di performance

Tecniche di apprendimento non-supervisionato

Tecniche di apprendimento supervisionato: classificazione e regressione

Modellare non-linearietà nei dati: tecniche basate sul concetto di *kernel*

Automated Machine Learning: configurazione automatica di un modello di Machine Learning

#### Programma esteso

##### Introduzione

- Machine Learning scenarios & tasks, notazioni utili
- Tipi di dati e problemi: tabular, streams, text, time-series, sequences, spatial, graph, web, social

## **Unsupervised Learning**

- Concetti di similarità e distanza
- Clustering
- Outlier detection

## **Supervised Learning**

- Generalità: classificazione e regressione, metriche, tecniche di validazione (hold-out,  $k$  fold-cross, leave-one-out)
- Approcci model-free/instance-based, un semplice algoritmo: k-nearest neighbors (KNN)
- Approcci model-based: Support Vector Machine (lineari)

## **Supervised Learning per dati non-lineari**

- Non-linearità, VC dimensions, kernel-trick
- Decision Tree e Random Forest
- Kernel-based learning: kernel-SVM e Gaussian Processes per classificazione e regressione
- Dimensionality reduction: Principal Component Analysis (PCA) e kernel-based PCA (kPCA)

## **L'approccio connessista**

- Artificial Neural Networks: paradigma di apprendimento
- Deep Learning: “*a fraction of the connectionist tribe*”

## **Una panoramica su Automated Machine Learning (AutoML)**

## **Esercizi ed esempi pratici**

## **Prerequisiti**

Si consiglia la conoscenza di elementi di base di informatica, matematica applicata, probabilità e statistica

## **Modalità didattica**

L'intera attività formativa viene svolta attraverso lezioni in presenza, video-registrate al fine di essere rese disponibili in formato digitale. Le lezioni riguarderanno sia aspetti teorici che applicazioni pratiche, specificatamente l'utilizzo di librerie software e dati open.

## **Materiale didattico**

- Testo di riferimento: Mehryar Mohri, Afshin Rostamizadeh and Ameet Talwalkar (2018). Foundations of Machine Learning.

- Slides e materiale didattico fornito dal docente

#### **Altri tesi suggeriti:**

- Charu C. Aggarwal (2015). Data Mining – the Textbook
- Carl Edward Rasmussen and Christopher K. I., Williams (2006). Gaussian Processes for Machine Learning.
- Robert B. Gramacy (2020). Surrogates – Gaussian Processes Modeling, Design, and Optimization for the Applied Statistics.

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

Second semester

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

La modalità di verifica prevede le seguenti 2 prove:

- lo svolgimento di un progetto con associata redazione di un rapporto tecnico, stile articolo scientifico,
- un esame orale (individuale e obbligatorio) finalizzato a verificare il grado di comprensione degli argomenti trattati.

Il progetto può essere svolto in *team* (max 3 studenti per gruppo) ed i dataset oggetto delle attività saranno concordati con il docente a partire da piattaforme open quali OpenML, Kaggle o UCI Repository.

La qualità del progetto è stabilita sulla base del corretto utilizzo degli algoritmi di ML e all'analisi critiva dei risultati.  
L'esame orale è finalizzato alla verifica della comprensione di aspetti teorici e metodologici del ML.

Il progetto contribuisce al 60% della valutazione finale, la prova orale al restante 40%.

Non sono previste prove intermedie.

#### **Orario di ricevimento**

Su appuntamento

#### **Sustainable Development Goals**

ISTRUZIONE DI QUALITÁ

---