



# UNIVERSITÀ DI PISA

---

## DATA MINING AND MACHINE LEARNING

**FRANCESCO MARCELLONI**

Anno accademico **2023/24**  
CdS **ARTIFICIAL INTELLIGENCE AND  
DATA ENGINEERING**  
Codice **878II**  
CFU **12**

Moduli	Settore/i	Tipo	Ore	Docente/i
ADVANCED DATA MINING AND MACHINE LEARNING	ING-INF/05	LEZIONI	60	FRANCESCO MARCELLONI ALESSANDRO RENDA
FUNDAMENTALS OF DATA MINING AND MACHINE LEARNING	ING-INF/05	LEZIONI	60	FRANCESCO MARCELLONI ALESSANDRO RENDA

### Obiettivi di apprendimento

#### *Conoscenze*

Gli studenti che completeranno con successo l'insegnamento avranno una solida conoscenza delle principali tecniche per pre-processare i dati, frequent pattern mining, frequent sequential pattern mining, graph mining, classificazione, predizione, clustering, outlier detection. Questa conoscenza permetterà loro di affrontare ogni tipo di problema inerente il data mining e di identificare la tecnica più adatta per risolverlo.

#### *Modalità di verifica delle conoscenze*

Durante la verifica delle conoscenze, gli studenti devono dimostrare di aver appreso le diverse tecniche insegnate durante lo svolgimento del corso e devono essere capaci di identificare la soluzione più adatta per problemi di data mining specifici.

I metodi sono:

- esame orale
- report e presentazione del progetto

Ulteriori informazioni: allo studente è richiesto di sviluppare un progetto in cui vengono utilizzate tecniche di data mining. I risultati del progetto vengono discussi durante una presentazione.

#### *Capacità*

Al termine del corso,

- lo studente saprà affrontare i più comuni problemi di data mining, trovando le soluzioni più idonee per risolverli
- lo studente saprà valutare e confrontare più soluzioni e scegliere la più adatta

#### *Modalità di verifica delle capacità*

Lo studente dovrà preparare e presentare una relazione scritta che riporti i risultati dell'attività di progetto

#### *Comportamenti*

Lo studente potrà acquisire un metodo per affrontare problemi di data mining e per selezionare le migliori soluzioni da adottare

#### *Modalità di verifica dei comportamenti*

Durante le sessioni di laboratorio saranno valutati il grado di accuratezza e precisione delle attività svolte dallo studente  
Durante lo sviluppo del progetto saranno verificate le modalità di gestione e organizzazione delle fasi progettuali

#### *Prerequisiti (conoscenze iniziali)*

Conoscenze di base di matematica

Conoscenze di linguaggi di programmazione



## UNIVERSITÀ DI PISA

---

### Indicazioni metodologiche

Le lezioni verranno svolte frontalmente con l'ausilio di slide

Le esercitazioni verranno svolte in laboratorio con l'ausilio di slide e esempi di programmazione.

Durante il corso, verrà sviluppato dallo studente un progetto che costituirà parte della valutazione finale

L'intero corso è tenuto in Inglese

### Programma (contenuti dell'insegnamento)

Data Preprocessing: data cleaning, integration, reduction, transformation and discretization.

Frequent pattern mining: basic concepts, A-priori algorithm, Pattern-Growth approach, vertical data format, pattern evaluation methods, constraint-based frequent pattern mining, colossal pattern.

Classification: basic concepts, decision tree induction, Bayes classification methods, rule-based classification, lazy learners, techniques for improving accuracy, model evaluation and selection.

Clustering: basic concepts, partitioning methods, hierarchical methods, density-based methods, grid-based methods, model evaluation and selection, clustering with constraints.

Outlier detection: statistical, proximity-based, clustering-based and classification-based approaches.

Sequential Pattern Mining: basic concepts, AprioriAll, AprioriSome, AprioriDynamicSome

Graph Mining: basic concept, geodesic distance, SimRank, Density-based approaches to graph clustering.

Time series: basic concepts and machine learning for forecasting.

Text mining: basic concepts, text classification and case studies.

Trustworthy AI: basic concepts and current solutions.

Distributed frameworks: basic concepts, Hadoop, MapReduce paradigm, Spark, some examples of data mining algorithms implemented by using MapReduce

Python programming language: Introduction, Jupyter Notebook environment and Machine Learning libraries

### Bibliografia e materiale didattico

Slides del corso

Libro: J. Han and M. Kamber. Data Mining: Concepts and Techniques. Morgan Kaufmann, 4th ed., 2022

Articoli sui differenti algoritmi descritti durante le lezioni forniti dal docente

### Modalità d'esame

L'esame è composto dalla discussione del progetto e una prova orale.

La discussione del progetto viene tipicamente tenuta qualche giorno prima dell'esame orale. Il candidato deve presentare come il progetto è stato sviluppato, motivare le sue scelte progettuali e discutere i risultati ottenuti. Il progetto viene valutato positivamente se il candidato mostra di aver seguito un approccio corretto e di aver valutato in modo critico le possibili soluzioni, scegliendo la più appropriata

La prova orale consiste in un colloquio tra il candidato e il docente su alcune domande che potrebbero anche assegnate in forma scritta al candidato.

La prova orale è superata se il candidato mostra padronanza degli argomenti trattati, si esprime in modo chiaro e con terminologia corretta, mostra capacità di analisi e sintesi.

*Ultimo aggiornamento 18/09/2023 17:24*