# Laboratorio informatico per le decisioni aziendali (Data Mining)

## I Mod.: Prof. Gabriele Cantaluppi; II Mod.: Prof. Stefano Gatti

***OBIETTIVO DEL CORSO E RISULTATI DI APPRENDIMENTO ATTESI***

Il corso si propone di conferire agli studenti la conoscenza dei concetti, termini e metodi dello statistical machine learning e la capacità di applicarli correttamente a problemi reali nell’ambito dell’economia e della gestione aziendale. Il corso si propone inoltre di sviluppare competenze fondamentali per le analisi quantitative svolte negli altri insegnamenti del curriculum e richieste dalle professioni coinvolte nella gestione dei dati.

Particolare attenzione viene posta agli aspetti metodologici e realizzativi, mediante illustrazione di casi reali ed esercitazioni in aula con l’ausilio di strumenti informatici largamente utilizzati nel contesto accademico e nel mondo aziendale.

Al termine del corso gli studenti saranno in grado di padroneggiare le seguenti abilità e conoscenze:

1. Conoscenza dei concetti, dei termini e dei metodi dello statistical machine learning e delle principali funzioni dei software Microsoft Excel e Python con Jupyter Notebook (DD1- Knowledge and understanding);
2. Capacità di applicare correttamente i metodi di statistical e machine learning a problemi reali in ambito economico e nella gestione aziendale (DD2- Applying knowledge and understanding);
3. Sviluppo di pensiero critico che consentirà loro di formulare giudizi indipendenti, fondati su solide basi scientifiche (DD3-Making judgements).
4. Capacità di costruire, leggere e interpretare gli output delle procedure di statistical machine learning trattate nel corso e di comunicare i relativi risultati, attraverso l’estrazione di informazioni qualitative dai dati (DD4-Communication).
5. Sviluppo di competenze utili per le analisi quantitative che verranno svolte negli insegnamenti del curriculum e per quelle richieste dalle posizioni professionali coinvolte nella gestione dei dati, capacità di ragionamento rigoroso e assunzione di decisioni data-driven (DD5- Lifelong learning skills).

***PROGRAMMA DEL CORSO***

Il corso prevede 60 ore di lezione e 8 di esercitazione in aula informatica sui software utilizzati durante il corso (equamente ripartite tra i due moduli).

Nel primo modulo (statistico, 30 ore), che consente l’approfondimento degli aspetti riguardanti l’applicazione delle metodologie statistiche per l’analisi dei dati aziendali, viene posta particolare attenzione alle procedure presenti in Excel, tra le quali figurano l’analisi della regressione multipla, la simulazione delle distribuzioni di probabilità e l’analisi delle serie storiche (medie mobili, destagionalizzazione, livellamento esponenziale, previsione di possibili scenari).

Il secondo modulo (informatico, 30 ore) è dedicato all’approfondimento pratico dell’analisi dei dati aziendali e all’applicazione di metodi di *data mining* (in particolare classificazione, *clustering*, regole associative) volti ad affrontare problematiche aziendali tipiche (profilazione, analisi di propensione, segmentazione, *market basket analysis*).

***BIBLIOGRAFIA[[1]](#footnote-1)***

Per la parte statistica

G. Boari-G. Cantaluppi, *Raccolta di temi ed esercizi per il corso di Laboratorio Informatico per le Decisioni Aziendali,* 2° modulo (statistico), EDUCatt, Milano, 2017.

Testi consigliati per approfondimenti su inferenza statistica e metodi di data mining

G. Cicchitelli-P. D’Urso-M. Minozzo  *Statistica: principi e metodi,* Pearson, Milano, 2018. [*Acquista da VP*](https://librerie.unicatt.it/scheda-libro/giuseppe-cicchitelli-pierpaolo-durso-marco-minozzo/statistica-principi-e-metodi-ediz-mylab-9788891915382-713889.html)

G. James-D. Witten-T. Hastie-R. Tibshirani, *An Introduction to Statistical Learning,* Springer, New York, 2021, https://www.statlearning.com/ (capp. 1, 2, 3, 5).

Per la parte informatica

Jupyter Team, *Jupyter Project Documentation,* https://docs.jupyter.org/en/latest/index.html.

J. VanderPlas, *Python Datascience handbook,* O’Reilly Media Inc, USA, 2016, https://jakevdp.github.io/PythonDataScienceHandbook/.

Testi consigliati per approfondimenti per la parte teorica di modelli e sistemi per le decisioni

M. Bramer, *Principles of Data Mining,* Springer, 2020, 4th Edition.

C. Vercellis, *Business Intelligence: modelli matematici e sistemi per le decisioni,* McGraw-Hill, Milano, 2006. [*Acquista da VP*](https://librerie.unicatt.it/scheda-libro/vercellis-carlo/business-intelligence-9788838663468-173991.html)

***DIDATTICA DEL CORSO***

Lezioni teoriche (equamente ripartite tra inquadramento teorico dei temi trattati e sviluppo di abilità informatiche) e tutorial da svolgere in auto apprendimento.

***METODO E CRITERI DI VALUTAZIONE***

Prova scritta di valutazione globale sul programma al termine dell’intero corso. È prevista una prova intermedia al termine del primo modulo (statistico).

Per il primo modulo (statistico), la prova prevede tre esercizi, comprensivi di domande con risposta aperta. Due esercizi hanno carattere numerico, uno carattere teorico.

Per il secondo modulo (informatico), la prova prevede domande con risposta aperta, con risposte chiuse e un esercizio pratico di analisi dei dati in cui lo studente dovrà scrivere un piccolo programma in Python utilizzando Jupyter Notebook.

Le prove dei due moduli hanno eguale peso con riferimento alla valutazione finale. Ai fini del superamento dell’esame lo studente dovrà conseguire una votazione positiva in entrambi i moduli.

L’esame è volto a valutare innanzitutto capacità di ragionamento e rigore analitico sui temi oggetto del corso, nonché proprietà di linguaggio e abilità comunicative.

***AVVERTENZE E PREREQUISITI***

In relazione alla modalità di svolgimento delle lezioni, a forte carattere applicativo, si raccomanda agli studenti di partecipare muniti del proprio computer portatile.

I software che verranno utilizzati durante il corso sono:

– Microsoft Excel.

– Jupyter Notebook, open source, disponibile il download sul sito: *https://jupyter.org/*.

– Python, open source, disponibile per il download sul sito: *https://www.python.org/*.

Il corso presuppone la conoscenza delle nozioni di base di statistica (analisi dei dati, probabilità e inferenza) a livello degli insegnamenti sia di ‘Statistica I’/‘Statistica (Analisi dei dati e probabilità)’/ Statistics sia di ‘Statistica applicata’ impartiti nei Corsi di Studio triennali della Facoltà di Economia della nostra università. Le lezioni introduttive sono comunque dedicate al consolidamento di alcune delle nozioni presentate nell’insegnamento di ‘Statistica applicata’.

*Orario e luogo di ricevimento degli studenti*

Il giovedì dalle 10.30 alle 11.00, presso il Dipartimento di Scienze statistiche, previa comunicazione via mail.

1. I testi indicati nella bibliografia sono acquistabili presso le librerie di Ateneo; è possibile acquistarli anche presso altri rivenditori. [↑](#footnote-ref-1)