

Primi passi di Data Science: La Statistica come strumento per la classificazione

Dalla teoria alla pratica: classificazione e previsione

Come possono i dati clinici anticipare una diagnosi? Come decidono le banche a chi prestare denaro? Come si determina se un'email è spam basandosi su mittente, oggetto e contenuto? Come si determinano specie di piante in base a determinate caratteristiche?

Il laboratorio introduce metodi di classificazione statistica per individuare gruppi omogenei di unità statistiche e prevedere fenomeni economici, sociali e ambientali. Gli studenti acquisiranno competenze teoriche sulle tecniche previsive e abilità pratiche con il software R, imparando quando e come applicare questi strumenti e interpretarne correttamente i risultati.

Il laboratorio si articola in due incontri: uno in presenza presso l'università per la formazione teorico-pratica e uno facoltativo online per la presentazione di un progetto a scelta dei partecipanti consistente in un'analisi di dati che replichi quanto fatto nella giornata in università.

Alla fine delle attività gli studenti riceveranno un attestato di partecipazione che certificherà le ore frequentate (6 senza progetto, 15 con progetto).

Risultati di apprendimento atteso:

Alla fine del laboratorio PLS-Statistica gli studenti saranno in grado di: i) spiegare e applicare gli strumenti statistici di base per risolvere un problema di classificazione attraverso gli alberi decisionali; ii) implementare l'analisi con il software R; iii) presentare i risultati ottenuti.

Programma dei Moduli

Giornata in università: 2 o 13 febbraio 2026

Mattina: Aula U6-08, Edificio U6

9.30 - 9.45: Benvenuto (Viviana Amati)

9.45 - 12.30: Introduzione ai problemi di classificazione e previsione in ambito statistico. Presentazione della teoria degli alberi di classificazione con particolare attenzione al loro utilizzo in ambito previsorio (Andrea Marletta).

12.30 - 14.00: Pranzo

Pomeriggio: Lab 908, 909, 910, 911, Edificio U9

14.00 - 17.00: Introduzione all'uso del software R. Esempi pratici di creazione di alberi di classificazione con rappresentazione grafica dei modelli creati (Alice Giampino, Daniele Pirotta, Valentina Zangirolami, Viviana Amati)

Data da definire - Ulteriore incontro facoltativo (online)

Il laboratorio si conclude con un progetto facoltativo da svolgere entro aprile, in cui i partecipanti in gruppi replicheranno un'analisi dati con R applicando le tecniche apprese. I dataset saranno forniti da noi o concordati in base agli interessi dei partecipanti. Ogni gruppo avrà un tutor dedicato che li supporterà nel periodo tra la giornata in università e l'incontro online finale con la presentazione del progetto.

Modalità: primo incontro in presenza in università nei giorni 2 o 13 febbraio 2026,
secondo incontro online in data da definire

Iscrizioni e Informazioni: Viviana Amati (viviana.amati@unimib.it)
Alice Giampino (alice.giampino@unimib.it)