



UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Data Mining M

2223-2-F8204B014

Obiettivi formativi

Il corso si pone come obiettivo l'approfondimento di tecniche per l'analisi dei dati e di *data mining* e il perfezionamento delle abilità di modellizzazione con finalità previsiva, con relative implementazioni nell'ambiente di programmazione R.

Contenuti sintetici

Approfondimento dell'insieme delle metodologie di *data mining* e dell'apprendimento supervisionato e dei suoi principi generali.

Verrà fornito un quadro del processo di modellizzazione dei dati con finalità previsiva fruibile attraverso R. 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 delle dell'errore di previsione.

Programma esteso

- Principi fondamentali: sovra-adattamento, compromesso tra distorsione e varianza, ottimismo
- *Splines* di regressione e modelli addittivi
- Selezione delle variabili
- Riduzione della dimensionalità
- 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

La modalità di verifica consiste nell'analisi di un dataset e in una prova orale.

L'analisi dei dati ha come obiettivo la verifica delle abilità di modellizzazione dei dati a fini previsivi. La prova orale riguarderà sia l'esposizione dell'analisi dei dati sia la verifica dello studio degli argomenti trattati a lezione.

Per l'analisi dei dati, 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.

Testi di riferimento

- Azzalini, Scarpa (2004). *Analisi dei dati e data mining*. Springer-Verlag Italia
- Hastie, Tibshirani, Friedman (2009). *The Elements of Statistical Learning*. Springer

Approfondimenti :

- Kuhn, Silge (2022). *Tidy Modeling with R*. O'Reilly Media, Inc.
- Lewis, Kane, Arnold (2019) *A Computational Approach to Statistical Learning*. Chapman And Hall/Crc.
- Kuhn, Johnson (2019). *Feature Engineering and Selection*. 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.

Sustainable Development Goals
