



**UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA**

SYLLABUS DEL CORSO

Machine Learning and Ai-Applications for Data Management

2627-2-F7603Q028-F7603Q02802

Obiettivi

Al termine di questo modulo, gli studenti acquisiranno le basi teoriche e gli strumenti operativi dei principali approcci di machine learning per il data management, applicati all'analisi di macro-indicatori sintetici per la sostenibilità. In particolare, saranno in grado di:

- padroneggiare gli elementi fondamentali dell'analisi multivariata e del processo decisionale multicriterio per analizzare sistemi di dati complessi definiti da macro-indicatori di sostenibilità (conoscenza e capacità di comprensione)
- conoscere i principi e le condizioni operative delle principali tecniche e scegliere e gestire gli approcci di indagine più adatti alle finalità dell'analisi (capacità di applicare conoscenza e comprensione)
- valutare le caratteristiche, i campi di applicazione e i vantaggi e gli svantaggi delle singole tecniche di machine learning (autonomia di giudizio)
- proporre e giustificare la scelta dell'approccio ritenuto più adatto a uno specifico problema, comunicando i risultati a un pubblico che potrebbe non avere familiarità con questi metodi (abilità comunicative)
- apprendere in modo autonomo ed esplorare ulteriori estensioni metodologiche del machine learning per il data management (capacità di apprendimento).

Contenuti sintetici

- Struttura e pre-trattamento dei dati nell'ambito del machine learning; covarianza e correlazione
- Strategie per l'esplorazione dei dati e la razionalizzazione di problemi complessi: Analisi delle Componenti Principali
- Misure di distanza per la similarità e la diversità
- Modellazione supervisionata per l'analisi degli indicatori di sostenibilità
- Strategie per il processo decisionale multicriterio
- Esperienze pratiche e casi di studio

Programma esteso

La struttura dei dati multivariati:

- Covarianza e correlazione
- Scalatura e pre-trattamento dei dati: centratura, auto-scaling, range scaling, variance scaling

Strategie per l'esplorazione dei dati e la razionalizzazione di problemi complessi:

- Analisi della struttura ed esplorazione dei dati relativi a sistemi complessi nell'ambito dei macro-indicatori sintetici per la sostenibilità
- Analisi delle Componenti Principali (PCA): scores e loadings; autovalori e definizione delle componenti significative (analisi del rango)
- Esempi di PCA e casi di studio

Analisi della similarità e della diversità nei sistemi complessi:

- I concetti di analogia, similarità, dissimilarità e distanza
- Misure di distanza e similarità per dati quantitativi e binari
- Cluster Analysis: esempi e casi di studio

Modellazione supervisionata per l'analisi degli indicatori di sostenibilità:

- Il concetto di bias e i metodi di validazione; tecniche di validazione dei modelli supervisionati
- Metodi di regressione e classificazione multivariata: esempi e casi di studio

Strategie per il processo decisionale multicriterio:

- Funzioni di desiderabilità e di utilità per l'ordinamento razionale di dati multivariati; fronte di Pareto
- Esempi e casi di studio

Prerequisiti

Conoscenze di base di statistica univariata

Modalità didattica

Lezioni frontali e attività pratiche di laboratorio:

10 ore, lezioni in presenza, didattica erogativa

10 ore, lezioni in presenza, didattica interattiva

10 ore, lezioni in presenza, didattica interattiva per la realizzazione di una relazione su un caso di studio, da analizzare a gruppi da parte degli student

Materiale didattico

Slides

Articoli scientifici messi a disposizione sul sito di e-learning del corso

Dati e codice per le sessioni pratiche messi a disposizione sul sito di e-learning del corso

Periodo di erogazione dell'insegnamento

Primo semestre

Modalità di verifica del profitto e valutazione

Per il metodo di verifica del profitto e valutazione, si veda il syllabus del corso *Laboratory of Data management and synthetic indicators for sustainability*

Orario di ricevimento

Su appuntamento via email

Sustainable Development Goals

ISTRUZIONE DI QUALITÀ
