UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Data Science Lab in Public Policies and Services

2526-2-FDS01Q043

Obiettivi

L'insegnamento vuole far apprendere a studenti e studentesse come analizzare dati medici (specialmente quelli di cartelle cliniche elettroniche) attraverso tecniche di statistica computazionale, di apprendimento automatico, e di analisi di sopravvivenza per scoprire nuova conoscenza sulle condizioni dei pazienti.
Questo insegnamento si propone di fornire i concetti di base dell'epidemiologia che sono alla base di un corretto approccio metodologico a un progetto di ricerca in sanità pubblica. Lo studente sarà in grado di trattare i dati in sanità pubblica in particolare concentrandosi su diversi aspetti tra cui il disegno dello studio, la gestione e l'analisi dei dati. Lo studente sarà in grado di implementare e di calcolare indicatori di qualità/performance.

Contenuti sintetici

Dataset search and retrieval
Data preparation and data cleaning
Exploratory data analysis
Unsupervised machine learning
Supervised machine learning
Feature ranking
Result understanding and validation
R and Python programming languages
Survival analysis
Epidemiologia della popolazione
Disegni di studio
Metodi statistici con applicazione ai registri e ai dati sanitari amministrativi

Programma esteso

Dataset search and retrieval
Data preparation and data cleaning
Exploratory data analysis
Unsupervised machine learning
Supervised machine learning
Feature ranking
Result understanding and validation
R and Python programming languages
Basics in population epidemiology.
Study designs: advanced designs to combine data from different sources
(registry data, biomarkers, biobanks, surveys).
Survival analysis: survival estimate and Cox model regression.
Record linkage approaches and statistical methods with application to registries and administrative health data.
Examples of Quality/performance indicators, outcome research with administrative data, system of indicators to evaluate the appropriateness of clinical pathways in chronic diseases.

Prerequisiti

Statistica di base e basi dell'apprendimento automatico
Conoscenza di base di R o Python

Modalità didattica

Lezioni in presenza ed esercitazioni in presenza
3 lezioni di 2 ore condotte in modalità remota (asincrona)

Materiale didattico

Slides presentate a lezione ed articoli scientifici segnalati a lezione

Articoli scientifici:
Davide Chicco, Vasco Coelho (2025) "A teaching proposal for a short course on biomedical data science", PLOS Computational Biology 21(4): e1012946. https://doi.org/10.1371/journal.pcbi.1012946

Libri di testo:
Kenneth J. Rothman Sander Greenland, Timothy L. Lash . Modern Epidemiology. Lippincott Williams & Wilkins; 3rd ed.
Eric Vittinghoff, David V. Glidden, Stephen C. Shiboski, Charles E. McCulloch. Regression Methods in Biostatistics Linear, Logistic, Survival, and Repeated Measures Models. Statistics for Biology and Health book series. Springer; 2nd edition (March 6, 2012)
Marie Reilly "Beyond classic epidemiological designs" https://www.routledge.com/Controlled-Epidemiological-Studies/Reilly/p/book/9780367186784 Chapman & Hall/CRC Biostatistics Series 2023

Periodo di erogazione dell'insegnamento

Secondo semestre

Modalità di verifica del profitto e valutazione

Lavoro personale su un progetto scientifico comprendente entrambe le unità didattiche per verificare le capacità dello studente nell'applicazione della metodologia di ricerca in sanità pubblica. Consegna di una relazione e presentazione orale del lavoro svolto per l'unità didattica Big Data in Public and Social Services.
Questionario a risposta chiusa per valutare la preparazione sul programma dell'unità didattica Big Data in Public Health.

Orario di ricevimento

Da concordare via email scrivendo a davide.chicco(AT)unimib.it o paola.rebora(AT)unimib.it

Sustainable Development Goals

SALUTE E BENESSERE