UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Data Science Lab in Public Policies and Services

2526-2-FDS01Q043

Aims

This module aims at teaching students how to analyze medical data (especially, data of electronic health records) through computational statistics and machine learning techniques to infer new knowledge about the conditions of patients.
This course aims to provide the basic concepts of epidemiology that are at the basis of a proper methodological approach to a research project in public health. The student will be able to deal with data in public health particularly focusing on several aspects including study design, data managment and analysis. The student will be able to implement design strategies on registries and administrative health data and able to calculate quality/performance indicators

Contents

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
Population epidemiology
Study designs
Statistical methods with application to registries and administrative health data

Detailed program

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.

Prerequisites

Basic statistics and basic machine learning
Basic knowledge of R o Python

Teaching form

In-person theory classes and practice exercise classes
3 2-hour lectures conducted in a remote (asynchronous) delivery mode

Textbook and teaching resource

Classes slides and scientific papers mentioned during classes

Articles:
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

Textbooks:
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

Semester

Second semester

Assessment method

Personal work on a scientific project including both teaching units to test the ability of the student in the application of research methodology in public health. Delivery of a report, and oral presentation of the work done, for the Data in Public and Social Services unit.
Questionnaire with closed answer to evaluate the preparation on the program of the teaching unit Big Data in Public Health.

Office hours

To define via email by writing to davide.chicco(AT)unimib.it or paola.rebora(AT)unimib.it

Sustainable Development Goals

GOOD HEALTH AND WELL-BEING