HEALTH DATA SCIENCE AND DATA ANALYTICS IN HEALTHCARE

Key Information

Module code: ISSU0089
Taught during: Session Two: Monday 22 July - Friday 9 August 2019
Module workload: 45 teaching hours plus approximately 100 study hours
Module leader: Dr Holger Kunz, Dr Spiros Denaxas
Department: Institute of Health Informatics
Credit: 15 UCL credits, 7.5 ECTS, 4 US Level: Level 2, second year Undergraduate
Pre-requisites: Successful completion of a first year undergraduate level module either in biomedicine or in a quantitative subject
Assessment: Practical assessments (50%)
Final test (50%)

Module Overview
Health Data Science is an exciting new area that combines scientific inquiry, statistical knowledge, substantive expertise, and computer programming in the area of healthcare and biomedicine. One of the main challenges for businesses, research institutes, and policy makers when using big health data is to find people with the appropriate skills. Students taking this module will be introduced to the most fundamental data analytic tools and techniques, and learn how to use specialised software to analyse real-world health data.

Week One
- Principles of research design in health data science;
- Foundations of health informatics and health data analytics;
- Programming skills in Python.

Week Two
- Machine Learning / Artificial Intelligence
- Classification models;
- Resampling methods;
- Model selection.

Week Three
- Regression Models;
- Unsupervised learning;
- Unstructured data analysis;

Please note that this module description is indicative and may be subject to change.
Module Aims
This course will be delivered by researchers and university teachers specialised in different areas of health data science. The course aims to provide an introduction to the health data science approach to the quantitative analysis of healthcare data using the methods of statistical learning, an approach blending classical methods with recent advances in artificial intelligence and machine learning. The course will cover the main analytical methods from this field with hands-on applications in Python using example datasets in different areas of medicine, so that students gain experience with and confidence in using the methods we cover. It also covers data preparation and processing, including working with structured data, CSV-files, and unstructured textual data.

Teaching Methods
Lectures and seminars (computer labs), student presentations, classroom debates, private reading and outlines/assignments. Reading lists will be available online via the UCL library site and the readings will be regularly revised and updated. Student support will be provided via seminars/tutorials (computer lab sessions) and office hours. Relevant materials and forums will also be housed on the Moodle system.

Learning Outcomes
Upon successful completion of this module, students will:

• Have a sound understanding of the field of health data science and develop the ability to analyse real-world data using some of its main methods;
• Applying classification to predict class labels for new feature vectors;
• Become comfortable applying regression models for continuous and limited outcome variables;
• Explore more complex models such as ensembles;
• Develop familiarity with descriptive and predictive analytics, and their application to big health data problems;
• Explore methods of unsupervised learning and clustering;
• Have received a solid foundation for more advanced or more specialised study and research.

Assessment Methods
• Practical assessments (50%)
• Final test (50%)

Key Texts
The primary texts are:

• Trevor Hastie, Robert Tibshirani, Jerome Friedman, (2009), The Elements of Statistical Learning, Springer.
• Ramona Nelson, Nancy Staggers (2016), Health Informatics: An interprofessional Approach, Elsevier
• Sebastian Raschka (2016), Python Machine Learning, Packt Open Source

Module Leaders
Holger Kunz is a computer scientist and Teaching Fellow at the Institute of Health Informatics at UCL. He has conducted research in applied machine learning for medical imaging and the treatment of eye tumours. He has also conducted data science research for clinical indicator systems and quality management/dashboards.
in a hospital setting and in the field of eHealth for health and wellbeing and disease prevention. He has presented his research at international conferences in Vancouver, Sydney, Portland, Lyon and Glasgow. He is passionate about health informatics and to improve the health of patients with data-driven methods.

Spiros Denaxas is an Associate Professor in Biomedical Informatics based in the Institute of Health Informatics at UCL. His background is in computer science, information systems engineering and bioinformatics. His research lab (http://denaxaslab.org) operates at the intersection between health research and computer science and focuses on creating and evaluating data-driven methods for transforming electronic health records into research-ready datasets and answering clinically meaningful questions.