LONDON
SCHOOL *of*
HYGIENE
&TROPICAL
MEDICINE

MODULE SPECIFICATION

Academic Year (student cohort covered by specification)	2023-24
Module Code	2490
Module Title	Machine Learning
Module Organiser(s)	Alex Lewin and Pierre Masselot
Faculty	Epidemiology & Population Health
FHEQ Level	Level 7
Credit Value	CATS: 15 ECTS: 7.5
HECoS Code	101030
Term of Delivery	Term 2
Mode of Delivery	For 2023-24 this module will be delivered face-to-face. Teaching will comprise a combination of live and interactive activities (synchronous learning) as well as recorded or self-directed study (asynchronous learning).
Mode of Study	Full-time
Language of Study	English
Pre-Requisites	Students are required to have completed the module Statistics for Health Data Science, or equivalent. Students must be able to demonstrate basic grounding in probability, hypothesis testing, least squares, maximum likelihood and familiarity with linear regression models and generalised linear models (GLMs). Students must also be able to demonstrate familiarity with R software and be able to perform data manipulation in that software and to run packages.
Accreditation by Professional Statutory and Regulatory Body	None
Module Cap (indicative number of students)	25 students.
Target Audience	This is a compulsory module for the programme MSc Health Data Science. It is intended for students who wish to be able to apply modern machine learning methods in health data research. The focus will be on critical understanding and applied analysis.

Module Description	This module provides an introduction to statistical and machine learning, with application to health data science. The module will cover a range of widely-used machine learning techniques used in health data research. The principles of learning algorithms will be covered, and the data analysis tasks will use existing packages in the freely-available R software.
Duration	5 weeks at 2 days per week
Timetabling slot	Slot C1
Last Revised (e.g. year changes approved)	September 2021

Programme(s) This module is linked to the following programme(s)	**Status**
MSc Health Data Science	Compulsory

Module Aim and Intended Learning Outcomes

Overall aim of the module
The overall module aim is to: • introduce the concept of statistical and machine learning, and cover a range of supervised learning methods.

Module Intended Learning Outcomes
Upon successful completion of the module a student will be able to: 1. contrast the principles behind a range of statistical and machine learning methods; 2. examine the application of a range of machine learning techniques to address health data science questions; 3. critically evaluate strengths and limitations of statistical and machine learning methods in health data science projects; 4. critically assess the application of different techniques and choose an appropriate algorithm to answer a specific health data science question.

Indicative Syllabus

Session Content

The module is expected to cover the following topics:

- Regression with the generalised linear model (GLM)
- Support vector machines
- K-nearest neighbours
- Decision trees and random forests
- Regularised/penalised methods for feature selection (e.g. LASSO)
- Ensemble methods (boosting and Super Learner)
- Principles of machine learning algorithms
- Cross-validation and prediction

Teaching and Learning

Notional Learning Hours

Type of Learning Time	Number of Hours	Expressed as Percentage (%)
Contact time	40	27
Directed self-study	50	33
Self-directed learning	30	20
Assessment, review and revision	30	20
Total	**150**	**100**

Student contact time refers to the tutor-mediated time allocated to teaching, provision of guidance and feedback to students. This time includes activities that take place in face-to-face contexts such as lectures, seminars, demonstrations, tutorials, supervised laboratory workshops, practical classes, project supervision as well as where tutors are available for one-to-one discussions and interaction by email. Student contact time also includes tutor-mediated activities that take place in online environments, which may be synchronous (using real-time digital tools such as Zoom or Blackboard Collaborate Ultra) or asynchronous (using digital tools such as tutor-moderated discussion forums or blogs often delivered through the School’s virtual learning environment, Moodle).

The division of notional learning hours listed above is indicative and is designed to inform students as to the relative split between interactive (online or on-campus) and self-directed study.

Teaching and Learning Strategy

Flipped classroom

- Prior to the introduction of each topic, students will be expected to do some background reading first in order to engage with a formative quiz at the beginning of, and group discussion during, each teaching session.

Interactive lectorials

- The teaching sessions will be run as lectorials, an interactive format that alternates between lecture-based delivery of material and hands-on practical work (optionally in small groups)
- Students will be expected to bring laptop computers to class to engage with the material in the appropriate software environment

Adaptive release

- Practical worksheets will only become available to students after completing the quiz based on the background reading to ensure that students are engaging critically with the material.

Assessment

Assessment Strategy

The module will employ a number of different formative and summative assessment strategies. To ensure students critically engage with the pre-reading material, formative quizzes will be used during interactive contact sessions.

The summative assessment will be a data analysis project, assessed by oral presentation and the submission of a reproducible data analysis report (commented code in Rmarkdown) (individual work) on which the presentation is based.

Summative Assessment

Assessment Type	**Assessment Length (i.e. Word Count, Length of presentation in minutes)**	**Weighting** (%)	**Intended Module Learning Outcomes Tested**
Individual Presentation	10-minute pre-recorded presentation and 5-minute Q&A session. and Accompanying R-markdown document containing clearly commented code.	100	1- 4

Resitting assessment
Resits will accord with the LSHTM's Resits Policy

Resources

Indicative reading list

James, G., D. Witten, T. Hastie, and R. Tibshirani. An Introduction to Statistical Learning: With Applications in R. Springer Texts in Statistics. Springer New York, 2014. http://faculty.marshall.usc.edu/gareth-james/.

Kuhn, M., and Johnson, K. Applied Predictive Modeling. SpringerLink : Bücher. Springer New York, 2013.

Kenett, R. and Redman, T. The Real Work of Data Science. Wiley, 2019.

Other resources

Module information, including timetables, lecture notes, practical instructions and key literature for each session will be made available via the Virtual Learning Environment (Moodle).

Teaching for Disabilities and Learning Differences

The module-specific site on Moodle gives students access to lecture notes and copies of the slides used during the lecture. Where appropriate, lectures are recorded and made available on Moodle. All materials posted on Moodle, including computer-based sessions, have been made accessible where possible.

LSHTM Moodle is accessible to the widest possible audience, regardless of specific needs or disabilities. More detail can be found in the Moodle Accessibility Statement which can also be found within the footer of the Moodle pages. All students have access to “SensusAccess” software which allows conversion of files into alternative formats.

Student Support Services can arrange learning or assessment adjustments for students where needed. Details and how to request support can be found on the LSHTM Disability Support pages.