



## MODULE SPECIFICATION

<b>Academic Year (student cohort covered by specification)</b>	2022-23
<b>Module Code</b>	2490
<b>Module Title</b>	Machine Learning
<b>Module Organiser(s)</b>	Alex Lewin and Pierre Masselot
<b>Faculty</b>	Epidemiology & Population Health
<b>FHEQ Level</b>	Level 7
<b>Credit Value</b>	CATS: <b>15</b> ECTS: <b>7.5</b>
<b>HECoS Code</b>	101030
<b>Term of Delivery</b>	Term 2
<b>Mode of Delivery</b>	For 2022-23 this module will be delivered by predominantly face-to-face teaching modes..  Where specific teaching methods (lectures, seminars, discussion groups) are noted in this module specification these will be delivered by predominantly face-to-face sessions. There will be a combination of live and interactive activities (synchronous learning) as well as recorded or self-directed study (asynchronous learning)
<b>Mode of Study</b>	Full-time
<b>Language of Study</b>	English
<b>Pre-Requisites</b>	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.
<b>Accreditation by Professional Statutory and Regulatory Body</b>	None
<b>Module Cap (indicative number of students)</b>	25 students.

<b>Target Audience</b>	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.
<b>Module Description</b>	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.
<b>Duration</b>	5 weeks at 2 days per week
<b>Timetabling slot</b>	Slot C1
<b>Last Revised (e.g. year changes approved)</b>	August 2022

<b>Programme(s)</b>	<b>Status</b>
This module is linked to the following programme(s)	
MSc Health Data Science	Compulsory

## Module Aim and Intended Learning Outcomes

<b>Overall aim of the module</b>
The overall module aim is to: <ul style="list-style-type: none"> <li>introduce the concept of statistical and machine learning, and cover a range of supervised learning methods.</li> </ul>

<b>Module Intended Learning Outcomes</b>
Upon successful completion of the module a student will be able to: <ol style="list-style-type: none"> <li>contrast the principles behind a range of statistical and machine learning methods;</li> <li>examine the application of a range of machine learning techniques to address health data science questions;</li> <li>critically evaluate strengths and limitations of statistical and machine learning methods in health data science projects;</li> <li>critically assess the application of different techniques and choose an appropriate algorithm to answer a specific health data science question.</li> </ol>



## 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
<b>Total</b>	<b>150</b>	<b>100</b>

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.

The division of notional learning hours listed above is indicative and is designed to inform students as to the relative split between interactive 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](#).