Imperial College London

Module Specification (Curriculum Review)

Basic details

		Earliest cohort	Latest cohort
UID		Cohorts covered	2024-25

Long title: Data Science and Machine Learning for Physics

New code: PHYS60022

New short title:

Brief description of module *(approx. 600 chars.)*:

This module will provide an introduction to the techniques required to analyse large data sets through hands on experience. The course will be taught in the Python computing language and will use standard packages such as numpy, scipy, matplotlib, pandas and pytorch. The module assumes full competence in Python programming skills from Years 1 and 2. You will be introduced to the implementation of different techniques required to analyse data (statistical techniques and machine learning techniques) through working through examples and then analysing different data sets.

576 characters

Available as a standalone module/ short course? N

Statutory details

	ECTS	CATS	Non-credit
Credit value	7.5	15	N

HECOS codes:

FHEQ level: Level 6

Allocation of study hours

	Hours	
Lectures	0	
Group teaching	10	*Incl. seminars, tutorials, problem classes.*
Lab/ practical	10	
Other scheduled	10	*Incl. project supervision, fieldwork, external visits.*
Independent study	157.5	*Incl. wider reading/ practice, follow-up work, completion of assessments, revisions.*
Placement	0	*Incl. work-based learning and study that occurs overseas.*
Total hours	187.5	
ECTS ratio	25.00	

Project/placement activity

Is placement activity allowed? No

Module delivery

Delivery mode: Taught/ Campus

Delivery term: Term 2

Other:

Other: Assessmennt in Term 3

Ownership

Primary department	Physics
Additional teaching departments	None
Delivery campus	South Kensington

Collaborative delivery

Collaborative delivery?	N
External institution	N/A
External department	N/A
External campus	N/A

Associated staff

Role	CID	Given name	Surname
Module Leader		Patrick	Dunne
Lecturer		Tim	Evans
Lecturer		David	Colling
Lecturer		Nicholas	Wardle

Learning and teaching

Module description

Learning outcomes	By the end of the course you will have: 1) a basic understanding of how to use Jupyter notebooks to record and explain the analysis of a data set using machine learning; 2) a basic understanding of and experience of the handling and manipulation of large datasets in Python; 3) you will be able to choose from the introductory range of machine learning tools you have been introduced to and will be able to apply your skills with them to the classes of problems covered on this course; 4) you will be able to justify which of these tools are appropriate for which classes of problem by assessing both their applicability and accessibility within commonly used software frameworks.
Module content	The module will introduce you to packages commonly used in the physical sciences (such as numpy, scipy, matplotlib, pandas, Tensorflow etc) and have you use them to analyse different dataset and to understand different classes of problems. The module will also include an introduction to topics specific to Big Data applications such as data compression and data curation/processing architectures.
Learning and Teaching Approach	The course will be delivered through ten notebooks containing practical examples of the analysis of different datasets. Each notebook contains all the course material for one week. The work on each notebook is supported by a one hour lecture and a three hour lab session in the Physics computing suite. For the lab session, students work in small groups or independently on the material in the notebook, including exercises, for that week. In addition, some support from staff or demonstrators will be available for part of this time during the lab sessions.
Assessment Strategy	Summative assessment of all the learning outcomes will be done through a practical test in which students will carry out data science tasks using a jupyter notebook. There will be a formative-only half-day practical test in a similar format to the final assessment partway through the course.

Feedback	Feedback on the formative assessment half-day practical test will be via a short one to one session with the demonstrator who marked it.
Reading list	"The Hundred-Page Machine Learning Book" by Andriy Burkov (short and free, online). "Hands on machine learning with Scikit-Learn, Keras and TensorFlow 3rd Edition" by Aurélien Géron. "Introduction to Machine Learning with python" by Andreas C. Müller, Sarah Guido.

Quality assurance

Date of first approval	
Date of last revision	
Date of this approval	
Module leader	Patrick Dunne

Office use only

QA Lead	
Department staff	
Date of collection	
Date exported	
Date imported	

Notes/ comments	

Template version 16/06/2017

Programme structure
Associated modules

UID	Legacy code	Module title	Requisite type

UID	Legacy code	Module title	Requisite type

Assessment details

		Pass mark
Grading method	Numeric	40%

Assessments

Assessment type	Assessment description	Weighting	Pass mark	Must pass?
Practical	In person practical test	100%	40%	Y
		100%		