



Machine Learning and Data Mining I Course Descriptor

Course Title	Machine Learning and Data Mining I	Faculty	EDGE Innovation Unit (London)
Course code	NCHNAP563	Course Leader	Professor Scott Wildman (interim)
Credit points	15	Teaching Period	This course will typically be delivered over a 6-week period.
FHEQ level	5	Date approved	June 2020
Compulsory/Optional	Compulsory		
Pre-requisites	None		
Co-requisites	None		

COURSE SUMMARY

This course introduces the learner to three of the most well used machine learning techniques for data mining and predictive modelling: regression, decision trees, clustering and principal component analysis (PCA). Learners will explore the difference between supervised, unsupervised and machine learning, and study how to build and analyse robust predictive models using tools such as Python and R. It uses tools and libraries to analyze data sets, build predictive models, and evaluate the fit of the models. Covers common learning algorithms, including dimensionality reduction, classification, principal-component analysis, k-NN, k-means clustering, gradient descent, regression, logistic regression, regularization, multiclass data and algorithms, boosting, and decision trees. This course will also examine data bias and accurate model building: assessing the appropriateness of training and test sets, evaluation and deployment.

COURSE AIMS

- Train learners in supervised and unsupervised machine learning and data mining techniques for data science application.
- Give learners the tools and knowledge to evaluate different machine learning techniques and make recommendations for particular data science problems.
- For learners to explore how to perform machine learning and develop predictive models.

LEARNING OUTCOMES

On successful completion of the course, learners will be able to:

KNOWLEDGE AND UNDERSTANDING

- K1b Have knowledge and critical understanding of the underlying mathematical concepts and principles behind machine learning algorithms.
- K2b Understand the difference between machine learning algorithms, their pros and cons and their application in data science.
- K3b Have a critical understanding of data variance, data bias, data (un)correlation and predictive model evaluation metrics.

SUBJECT SPECIFIC SKILLS

- S1b Apply and evaluate regression, decision trees and clustering for problem solving within data science.
- S2b Use industry standard machine learning and data mining tools.

TRANSFERABLE AND PROFESSIONAL SKILLS

- T1b Critically evaluate different approaches to problem solving.
- T2b Effectively communicate arguments, analyses and conclusions.
- T3bi Develop logical analyses and conceptual thinking.
- T3bii Demonstrate an effective technical proficiency of written English that uses a wide range of literacy skills and vocabulary selected appropriately to communicate to specialist and non-specialist audiences.

TEACHING AND LEARNING

This is an e-learning course, taught throughout the year.

This course can be offered as a standalone short course.

Teaching and learning strategies for this course will include:

- On-line learning
- On-line Discussion Groups
- On-line assessment

Course information and supplementary materials will be available on the College’s Virtual Learning Environment (VLE).

Learners are required to attend and participate in all the formal and timetabled sessions for this course. Learners are also expected to manage their self-directed learning and independent study in support of the course.

The course learning and teaching hours will be structured as follows:

- Off-the-job learning and teaching (6 days x 7 hours) = 42 hours
- On-the-job learning (12 days x 7 hours) = 84 hours (e.g. 2 days per week for 6 weeks)
- Private study (4 hours per week) = 24 hours

Total = 150 hours

Workplace assignments (see below) will be completed as part of on-the-job learning.

ASSESSMENT

FORMATIVE

Learners will be formatively assessed during the course by means of set assignments. These will not count towards the final degree but will provide learners with developmental feedback.

SUMMATIVE

Assessment will be in two forms:

AE	Assessment Type	Weighting	Online submission	Duration	Length
1	Written assignment	60%	Yes	-	2,000 words +/- 10%, excluding data tables
2	Set exercise using workplace datasets	40%	Yes	Requiring on average 10-20 hours to complete	-

FEEDBACK

Learners will receive formal feedback in a variety of ways: written (via email correspondence); oral and indirectly through discussion during group tutorials. Learners will

also attend a formal meeting with their Academic Mentor and Employer. These tri-partite reviews will monitor and evaluate the learner's progress.

Feedback is provided on summatively assessed assignments and through generic internal examiners' reports, both of which are posted on the VLE.

INDICATIVE READING

Note: Comprehensive and current reading lists for courses are produced annually in the Course Syllabus or other documentation provided to learners; the indicative reading list provided below is used as part of the approval/modification process only.

BOOKS

- Alpaydin, E., (2014), *Introduction to machine learning*, Cambridge, Massachusetts: MIT Press
- Olive, D.J., (2017), *Linear Regression*, Cham: Springer International Publishing: Imprint: Springer
- Jolliffe, I.T., (2002), *Principal Component Analysis*, New York, NY: Springer

JOURNALS

Learners are encouraged to consult relevant journals on machine learning and data mining.

ELECTRONIC RESOURCES

Learners are encouraged to consult relevant electronic resources on machine learning and data mining.

INDICATIVE TOPICS

- Regression, decision trees, clustering and PCA
 - Supervised and unsupervised machine learning
 - Predictive model building and deployment
-

Title: NCHNAP563 Machine Learning and Data Mining I					
Approved by: Academic Board					
Location: Academic Handbook/Programme specifications and Handbooks/ Undergraduate Apprenticeship Programmes/BSc (Hons) Data Science Programme Specification/Course Descriptors					
Version number	Date approved	Date published	Owner	Proposed next review date	Modification (As per AQF4) & category number
2.1	May 2022	May 2022	Scott Wildman	September 2026	Category 1: Corrections/clarifications to documents which do not change approved content.
2.0	January 2022	April 2022	Scott Wildman	September 2025	Category 3: Changes to Learning Outcomes
1.0	June 2020	June 2020	Scott Wildman	June 2025	