6.1 Machine Learning for Data Analysis

6.3.1 Headline information about the module

Module title	Machine Learning for Data Analysis
Module NFQ level (only if an NFQ level can be demonstrated)	9
Module number/reference	M3
Parent programme(s) the plural arises if there are embedded programmes to be validated.	Master of Science in Data Analytics
Stage of parent programme	AWARD
Semester (semester1/semester2 if applicable)	1
Module credit units (FET/HET/ECTS)	ECTS
Module credit number of units	10
List the teaching and learning modes	Full time & Part time blended learning
Entry requirements (statement of knowledge, skill and competence)	See Section 4
Pre-requisite module titles	none
Co-requisite module titles	none
Is this a capstone module? (Yes or No)	No
Specification of the qualifications (academic, pedagogical and professional/occupational) and experience required of staff (staff includes workplace personnel who are responsible for learners such as apprentices, trainees and learners in clinical placements)	Academic and Professional: PhD desirable and a minimum of an MSc is required. However, in exceptional cases, NFQ Level 8 in Data Analytics, Computer Science, Software Development, Software Engineering or equivalent may be acceptable when combined with significant industrial experience. Pedagogical: Teaching experience is desired. Completion of postgraduate CPD/Certificate in Teaching and Learning or similar preferred. Experience in blended learning delivery required. In absence of experience, training will be mandatory and will be provided.
Maximum number of learners per centre (or instance of the module)	120
Duration of the module	1 semester
Average (over the duration of the module) of the contact hours per week (see * below)	2.5 Hours
Module-specific physical resources and support required per centre (or instance of the module)	Physical resource requirements are 1 laptop or PC/workstation per student. On campus and online resources as per programme specification.
Analysis of required learning effort	
*Effort while in contact with staff	

On campus Lecture / Classroom demonstrations		Mentoring and small-group tutoring		Other (Reflective development, directed reading/group work)		Online classes & Directed e-learning (hours)	Independent learning (hours)	Other hours (specify)	Work-based learning hours of learning effort	Total effort (hours)
Hours	Minimum ratio teacher/learner	Hours	Minimum ratio teacher/learner	Hours	Minimum ratio teacher/learner					
10	1:16	10	1:5	10	1:16	30	190	N/A	N/A	250
Allocation of marks (within the module)										
				Continuous assessment		Supervised project	Proctored practical examination	Proctored written examination	Total	
Percentage contribution				100%					100%	

6.1.2 Module aims and objectives

The learning aims of this module:

1. The different categories of machine learning techniques.
2. The different stages of the Knowledge Discovery life cycle.
3. The major Supervised, Unsupervised and Semi-Supervised learning techniques
4. The application, optimisation and validation of various machine learning techniques

6.1.3 Minimum intended module learning outcomes

On successful completion of this module the learner will be able to:

1. Modify and implement Machine Learning Algorithms to solve analytical problems.
   (Linked to PLO 1, PLO 2, PLO 5)
2. Determine whether a given data analysis problem requires the use of supervised, semi-supervised or unsupervised learning methods. Develop and implement the chosen learning method.
   (Linked to PLO 1, PLO 2, PLO 4)
3. Develop a machine learning strategy for a given domain and communicate effectively to team members, peers and project stakeholders the insight to be gained from the interpreted results
   (Linked to PLO 1, PLO 4, PLO 6)
4. Implement a range of classification and regression techniques and detail / document their suitability for a variety of problem domains.
   (Linked to PLO 5)
5. Critically evaluate the performance of Machine Learning models, propose strategies to optimise performance.
   (Linked to PLO 3)

6.1.4 Rationale for inclusion of the module in the programme and its contribution to the overall MIPLOs

Machine learning is the method that automates data analysis through analytical model building. This module will equip the learner with a wide range of machine learning skills and techniques necessary to understand and analyse large data sets. This module will also serve as the basis for more advanced data analytics introduced in later modules.

Within this module learners will also implement the concepts introduced in Statistics for Data Analysis.

6.1.5 Information provided to learners about the module

A copy of the Module Descriptor will be provided to learners at the start of the module via the College LMS (Moodle)

6.1.6 Module content, organisation and structure

The following indicative syllabus contains a learner reflective component as outlined in the teaching and learning strategy for this programme.

Content
Syllabus rationale : Implementation of theoretical concepts Introduction • Supervised, semi-supervised and unsupervised learning (conceptual) • Classification and Regression (integrated concept : Statistics for Data Analysis) (conceptual) • Machine Learning, Deep Learning and Reinforcement Learning (conceptual) • CRISP-DM, KDD and SEMMA (practical)
Supervised Learning (integrated concept : Statistics for Data Analysis / Data Preparation and Visualisation) • Linear Regression (practical) • Nearest Neighbour (practical) • Gaussian Naive Bayes (practical) • Decision Trees (practical) • Support Vector Machine (SVM) (practical) • Random Forest (practical)
Unsupervised Learning (integrated concept : Data Preparation and Visualisation) • Clustering (practical) • Association (practical) • Anomaly Detection (practical) • Dimensionality Reduction (practical)
Semi-Supervised Learning (integrated concept : Data Preparation and Visualisation) • Natural Language Processing (practical)

Case Studies (practical examples)

- Supervised Learning
- Unsupervised Learning
- Semi-Supervised Learning
- Reinforcement Learning
- Deep Learning

Validation and Optimisation (integrated concept : Statistics for Data Analysis)

- Validation (Re-substitution, Hold-out, K-fold cross-validation, LOOCV, Random subsampling, Bootstrapping…) (practical)
- Optimisation (loss functions/cost functions, Gradient Descent, Momentum, AdaGrad, RMSProp, Adam…) (practical)

The above topics (where applicable) will be performed within a testing environment (programming I.D.E / spreadsheet to allow practical integration of theoretical knowledge)

Programming skills are continuously developed throughout this module through practical implementation of module content. Conceptual / theoretical topics will also be practically demonstrated through the use of appropriate tool sets, to emphasise the synergy between theory and programmatic demonstration.

6.1.7 Module teaching and learning (including formative assessment) strategy

To provide the learner with a strong foundation in the core topics covered during the lectures, practical sessions will reinforce lecture content and provide supervised time to complete some assessment tasks. Sessions will be interactive, with instructor-led example exercises highlighting important topics discussed in lectures.

To provide formative assessment for this module the learner will:

- Be provided an opportunity at the beginning of each week to engage in group discussion on the material covered the previous week, thereby allowing reflection and ensuring their competency.
- Complete student-suggested tasks in a peer learning environment to encourage collaboration and allow learners to self-evaluate their current knowledge while gaining new knowledge and insights (this strategy links directly to PLO 7, PLO 8)
- Join additional discussions covering any lab-based exercises which have been provided to the learner.

Course specific online and on campus learning activities according to learning type.

Learning Type	Online activities	On campus activities
Knowledge Acquisition	• Pre-recorded presentations / demonstrations • Multi-media text-based materials • Videos • Guest speakers • Open ed resources	• Face to face lecturers • Practical Demonstrations
Collaboration	• Group projects • Discussion forums	• Group projects

	• Virtual classroom peer learning • Team virtual lab activities • Group presentations • Mentoring	• Team based lab activities / practical workshops • Group presentations
Discussion	• Discussion forums (synchronous and asynchronous) • Zoom breakout room discussions • Online tutorials • Project supervision • Webinars (industry experts) • Reflective activities	• Class discussion • Tutorials • Project supervision • Face to face lab / practical activities
Investigation	• Open ed resources • Lab observations • Project research • Information and data sourcing, analysis and evaluation • Flipped Learning	• data sourcing and analysis
Practice	• Virtual lab • Simulations • Case studies • Analysis of data sets • Presentations • Online quizzes / MCQs	• Labs • Practical workshops • Group work
Production	• E-portfolio • Reflective journal • Assessment outputs • Quiz / MCQs • Case studies • GitHub records	• Exam • Case studies • Student demonstration

6.1.8 Work-based learning and practice-placement

Not Applicable

6.1.9 E-learning

Collaborative blended learning strategies will be utilised for this module to ensure peer learning can be experienced not only through face-to-face traditional learning but also through online approaches. The online element will be achieved through a variety of interactive methods, including discussion forums, collaborative blogs and wikis, virtual labs and classrooms, group online supervision, interactive demonstrations and e-portfolios. This integrated learning approach ensures learning can be both reflective and collaborative while developing the efficacy of the individual student as the module progresses.

6.1.10 Module physical resource requirements

Physical resource requirements are 1 laptop or PC/workstation per student. On campus and online resources as per programme specification.

6.1.11 Reading lists and other information resources

Recommended:

Aurélien Géron, 2019, 2nd Edition, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems, O'Reilly Media [ISBN: 978-1492032649]

Andriy Burkov, 2019, The Hundred-Page Machine Learning Book, Andriy Burkov, [ISBN: 978-1999579500]

6.1.12 Module summative assessment strategy

Module specific online and on campus assessment activities according to assessment type.

Assessment	Online Assessment	On Campus Assessment
Knowledge Acquisition	• Pre-recorded presentations / demonstrations	• Live presentations / demonstrations
Collaboration	• Group projects • Team virtual lab activities • Group pre-recorded presentations / demonstrations	• Team lab activities • Group presentations / demonstrations
Formative	• Discussion forums (asynchronous) • Project (individual & group) supervision • Reflective activities	• Project (individual & group) supervision
Investigation	• Theoretical and practical project research • Problem Based Learning	• Problem Based Learning
Practical	• Virtual lab • Analysis of data sets • Online quizzes / MCQs • Technical tasks (individual and group)	• Lab • Analysis of data sets • Technical tasks (individual and group)
Production	• Artefacts • Student code repository records	• Artefacts • Student code repository records

Indicative Full Time Assessment Schedule

	Week*	1	2	3	4	5	6	7	8	9	10	11	12	13
Semester 1	Statistics for Data Analysis				CA (10%)		IA (20%)	Reading Week		CA (35%)				IA (35%)
	Data Preparation & Visualisation						IA (20%)					CA (30%)		IA (50%)
	Machine Learning for Data Analytics			IA (10%)			IA (30%)				CA (20%)			IA (40%)
	Research & Professional Ethics			IA (10%)					CA (20%)					P (20%)

Indicative Part Time Assessment Schedule

	Week*	1	2	3	4	5	6	7	8	9	10	11	12	13
Sem 2	Machine Learning for Data Analysis			IA (10%)			CA (30%)				CA (20%)			IA (40%)
	Research & Professional Ethics			IA (10%)					CA (20%)					P (20%)
	Big Data Storage & Processing			IA (10%)		CA (15%)				CA (35%)				IA (40%)

Indicative Assessment Schedule for Certificate in Machine Learning for Data Analysis

Students undertaking the module Machine Learning for Data Analysis as part of the Certificate in Machine Learning for Data Analysis, will be assessed by 100% continuous assessment comprising of 4 distinct assessments. In accordance with QQI Assessments and Standards, students will be required to attain a pass mark in the module overall in order to successfully pass the module. Where a student does not attain a pass mark in the module overall, they will be afforded a maximum of three repeat assessment opportunities.

Assessment schedule:

week	1	2	3	4	5	6	7	8	9	10	11	12	13
Machine Learning for Data Analysis				CA (10%)		CA (20%)			CA (35%)				CA (35%)

CA = Continuous Assessment (MCQ, Class tasks, mini projects, lab assessments etc.) **P** = Presentation **IA** = Integrated Assessment (may be individual or group)

Note: Please also note the above grid does not include non-graded formative assessment which builds to graded forms of assessment.

Mapping of summative and formative assessment: MIMLO to MIPLO

A ✓ indicates that the PLO has been formatively assessed as per the teaching and learning strategy for the programme	Module Learning Outcomes					Programme Learning Outcomes							
	1	2	3	4	5	1	2	3	4	5	6	7	8
Summative Assessment	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓