

H9DMML1: Data Mining and Machine Learning I

Module Details	
Module Code:	H9DMML1
Long Title	Data Mining and Machine Learning I APPROVED
Title	Data Mining and Machine Learning I
Module Level:	LEVEL 9
EQF Level:	7
Valid From:	2019 (July 2019)
Language of Instruction:	English
Credits:	5
Field of Study:	0610 - Information and Communication Technologies (ICTs) not further defined or elsewhe
Module Delivered in	no programmes
Module Coordinator:	MICHAEL BRADFORD
Module Author:	Margarete Silva
Authors	Silva Margarete (12 June 2019 to ---)
Departments:	School of Computing
Exemption Note:	
Specifications of the qualifications and experience required of staff	MSc/PhD in a computing or cognate discipline. May have industry experience also.
Is this a capstone module?	No
Learning Outcomes	
On successful completion of this module the learner will be able to:	
#	Learning Outcome Description
LO1	Critically analyse fundamental data mining and knowledge discovery methodologies in order to assess best practice guidance when applied to data mining problems in specific contexts
LO2	Extract, transform, explore, and clean data in preparation for data mining and machine learning.
LO3	Build and evaluate data mining and machine learning models on various datasets and problem domains.
LO4	Extract, interpret and evaluate information and knowledge from various datasets.
LO5	Critically review current data mining research and assess research methods applied in the field
Programme Module Dependencies	
No requisites	
Dependencies	
Module Recommendations	
No recommendations listed	
Co-requisite Modules	
No Co-requisite modules listed	
Requirements	
No requirements listed	
Entry requirements	A level 8 degree or its equivalent in any discipline

H9DMML1: Data Mining and Machine Learning I

Module Content & Assessment			
Indicative Content			
1. Overview of Data Mining and Machine Learning			
1.1) History and Evolution. Revision of data science methodologies: KDD, CRISP-DM. Data Security. Taxonomy and overview of data mining and machine learning techniques			
2. General data pre-processing and transformation strategies			
2.1) Intro to prediction. Identifying and Handling Missing Values. Looking for Outliers. Transformations for Single/Multiple Predictors. Adding/removing predictors. Binning . Feature Selection			
3. Prediction models evaluation			
3.1) Data Splitting and Sampling Methods (Holdout, Cross-fold Validation, Stratification, etc.). Model Tuning and Overfitting. Determining the best model			
4. Regression Models			
4.1) Quantitative Methods of Performance. The Variance/Bias Trade-off. Linear Regression			
5. Regression Models			
5.1) Partial Least Squares Regression. K-Nearest Neighbours Regression			
6. Regression Models			
6.1) Regression Trees. Model-based Regression Trees			
7. Regression Models			
7.1) Rule-based Models. Model Tuning via LASSO, ElastiNet, and similar. Computing Considerations			
8. Classification Models			
8.1) Logistic Regression. Linear Discriminant Analysis			
9. Classification Models			
9.1) K-Nearest Neighbours. Naïve Bayes			
10. Classification Models			
10.1) Decision Trees (e.g., C5.0, Random Forests, etc.)			
11. Unsupervised Machine Learning			
11.1) Notions of distance and similarity. Euclidian vs. non-Euclidian spaces. Clustering: k-means, k-medoids			
12. Unsupervised Machine Learning			
12.1) Clustering for outlier detection. Plotting and understanding clusters. Cluster evaluation measures: DBIndex, WSSSE, scree plots			
Assessment Breakdown			%
Coursework			100.00%
Special Regulation			
_Mod_Spec_Regs			
Assessments			
Full Time			
Coursework			
Assessment Type:	Formative Assessment	% of total:	Non-Marked
Assessment Date:	n/a	Outcome addressed:	1,2,3,4,5
Non-Marked:	Yes		
Assessment Description: Formative assessment will be provided on the in-class individual or group activities. Feedback will be provided in written or oral format, or on-line through Moodle. In addition, in class discussions will be undertaken as part of the practical approach to learning.			
Assessment Type:	Project	% of total:	100
Assessment Date:	n/a	Outcome addressed:	1,2,3,4,5
Non-Marked:	No		
Assessment Description: Produce a portfolio of studies that critically compare the performance of different machine learning methods applied to at least 3 related large datasets.			
No End of Module Assessment			
No Workplace Assessment			
Reassessment Requirement			
Coursework Only <i>This module is reassessed solely on the basis of re-submitted coursework. There is no repeat written examination.</i>			
Reassessment Description The repeat strategy for this module is to repeat the project, learners may build upon previous submissions.			

H9DMML1: Data Mining and Machine Learning I

Module Workload				
Module Target Workload Hours 50 Hours				
Workload: Full Time				
Workload Type	Workload Description	Hours	Frequency	Average Weekly Learner Workload
Lecture	Classroom & Demonstrations (hours)	24	Every Week	24.00
Tutorial	Other hours (Practical/Tutorial)	24	Every Week	24.00
Independent Learning	Independent learning (hours)	77	Every Week	77.00
Total Hours				125.00
Total Weekly Learner Workload				125.00
Total Weekly Contact Hours				48.00
This module has no Apprenticeship workload.				
This module has no Blended workload.				
This module has no Block workload.				
This module has no Distance/Correspondance workload.				
This module has no Evening workload.				
This module has no Online workload.				
This module has no Part Time workload.				
MLO Competency Mapping				
Learning Outcomes	Competencies			
LO1: Critically analyse fundamental data mining and knowledge discovery methodologies in order to assess best practice guidance when applied to data mining problems in specific contexts				
LO2: Extract, transform, explore, and clean data in preparation for data mining and machine learning.				
LO3: Build and evaluate data mining and machine learning models on various datasets and problem domains.				
LO4: Extract, interpret and evaluate information and knowledge from various datasets.				
LO5: Critically review current data mining research and assess research methods applied in the field				

Module Resources
<i>Recommended Book Resources</i>
<p>Witten, I. H., Frank, E., Hall, M. A. & Pal, C. J.. (2016), Data Mining: Practical machine learning tools and techniques (4th ed), Morgan Kaufmann.</p> <p>Lantz, B.. (2015), Machine learning with R (2nd ed), Packt Publishing Ltd.</p> <p>Kelleher, J. D., Mac Namee, B., & D'Arcy, A.. (2015), Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies, MIT Press.</p>
<i>Supplementary Book Resources</i>
<p>Mueller, A. C.. (2016), Introduction to machine learning with Python, O'Reilly.</p> <p>Hofmann, M., & Klinkenberg, R.. (2013), RapidMiner: Data Mining Use Cases and Business Analytics Applications, CRC Press.</p> <p>Han, J., Pei, J., & Kamber, M.. (2011), Data mining: concepts and techniques (3rd ed), Elsevier.</p> <p>Berthold, M., & Hand, D. J.. (2003), Intelligent data analysis: an introduction, Springer Science & Business Media.</p>
<i>This module does not have any article/paper resources</i>
<i>Other Resources</i>
<p>[website], UC Irvine Machine Learning Reposi, http://archive.ics.uci.edu/ml/</p> <p>[website], Kaggle platform for predictive modelling competitions, https://www.kaggle.com/</p> <p>[website], Datasets for Data Mining and Data Science, http://www.kdnuggets.com/datasets/index.html</p> <p>[website], Datacamp, http://www.datacamp.com</p> <p>[website], Bloomberg, https://www.bloomberg.com/europe</p> <p>[website], Yahoo! Finance, https://uk.finance.yahoo.com</p> <p>[website], Google Finance, https://www.google.com/finance</p> <p>[website], Central Statistics Office, http://www.cso.ie</p> <p>[website], Eurostat, http://ec.europa.eu/eurostat</p> <p>[website], Data.gov, https://www.data.gov</p> <p>[website], Amazon Web Services Public Datasets, https://aws.amazon.com/datasets</p> <p>[website], DataMarket, https://datamarket.com</p> <p>[website], The Pew Research Centre, http://www.pewresearch.org/data</p> <p>[website], The Fama-French Data Library, http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html</p> <p>[website], Federal Reserve Economic Data (FRED), https://fred.stlouisfed.org</p>

Review

Module Teaching and Learning (including formative assessment) Strategy

Learning will take place in a classroom environment and will feature both lectures and practical sessions for each topic. Materials and resources will be placed on Moodle, the College's virtual learning environment. Such materials may include: lecture notes, lab/tutorial descriptions, assessments, datasets, discussion groups, videos (e.g., created by lecturers or from external sources such as YouTube or Datacamp.com). The learners will also be provided with software tools (e.g., Python Scikit-Learn, R/RStudio, RapidMiner, Weka), for conducting the practical data analysis, mining and machine learning tutorials. Students may be required to work in groups during the labs/tutorials. Moodle may also be used to both interact with students and in some instances to provide formative assessments and feedback. Learners will have access to library resources, both physical and electronic, outside of the classroom where required. Students will also have access to learning, computing and maths support services.

Assessment Type	Assessment Description	Outcome Addresses	% of Total	Assessment Date
Formative Assessment	Formative assessment will be provided on the in-class individual or group activities. Feedback will be provided in written or oral format, or on-line through Moodle. In addition, in class discussions will be undertaken as part of the practical approach to learning.	All Learning Outcomes.	N/A	Ongoing
Project	Produce a portfolio of studies that critically compare the performance of different machine learning methods applied to at least 3 related large datasets.	LO1-5	100%	Week 7 (20%) and Week 13 (80%)

Work-based learning and practice placement

N/A

E-Learning

While the teaching and learning strategy for this module is classroom-centred, Moodle will be used to manage all the learning activities, resources and materials. Such resources may include videos created by the lecturer or from external sources such as YouTube or datacamp.com. The students will also be encouraged to do additional practical tutorials available on the Internet with links to these being provided through the Moodle course page.

Module Physical resource requirements

This module will require the following physical resources:

Access to high speed Wi-Fi network
BYOD laptop/notebook with a substantial hardware configuration. Its minimal suitable configuration is 8GB of RAM (16GB are recommended); a modern 64-bit x86 multi-core processor (Intel i5 or superior); 250+ GB of available space in hard disk; Wi-Fi card; and a recent version of Ubuntu, macOS, or Windows.
Laptop charging points
Library books (See further detail below)

Classes and Tutorials will take place either in a state-of-the-art lecture computer laboratory or in a standard lecture theatre. This includes projectors, whiteboards, audio-visual equipment, and portable document cameras (e.g., Ladibug).
In the main, printed material will not be provided as full use of virtual environments will be made.
The NCI Norma Smurfit Library holds copies of the recommended books reading list.

Specifications for module staffing requirements

This module requires a lecturer holding a MSc/PhD in a computer science or cognate discipline.

Sample assessment materials

Timetabling, learner effort and credit

Data Mining and Machine Learning I is a 5-credit module that will take place in semester 1 for full-time students. The module consists of 48 hours of contact classes.

Module assessment rationale

Sample assessments are included in Appendix 5.

Max number of learners per centre					
Registration Type				Number	
Student Quota Type 1				60	
Discussion Note:					
Affiliated Entities					
Entity Code	Entity Title	Entity Version	Entity Type	Via	Outcome
MSCAI	MSc in Artificial Intelligence	1.0	Programme		Yes
MSCDA	MSc in Data Analytics	2.0	Programme		Yes
PGDDA	Post Graduate Diploma in Science in Data Analytics	2.0	Programme		Yes

Documents

No Documents

statuslog

<i>Initial Status</i>	<i>End Status</i>	<i>Available</i>	<i>Date</i>	<i>Comment</i>
Vice Dean (Postgraduate)	Approved	McCarthy David	13/Aug/2019 16:00	Content reviewed and approved by Vice Dean of Postgraduate Studies.
Awaiting Module Coordinator	Vice Dean (Postgraduate)	McCarthy David	13/Aug/2019 16:00	Content reviewed and approved by Vice Dean of Postgraduate Studies.
Draft	Awaiting Module Coordinator	Grecos Christos	15/Jul/2019 12:19	YES
Awaiting Module Coordinator	Draft	McCarthy David	11/Jul/2019 15:32	Request from Christos Grecos to make minor corrections before approval.
Draft	Awaiting Module Coordinator	Silva Margarete	04/Jul/2019 15:52	yes
	Draft	Silva Margarete	12/Jun/2019 10:23	New Module Created