

H9DMML1: Data Mining and Machine Learning I

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
EHEA Level:	Second Cycle
Credits:	5
Module Coordinator:	MICHAEL BRADFORD
Module Author:	Margarete Silva
Departments:	School of Computing
Specifications of the qualifications and experience required of staff	MSc/PhD in a computing or cognate discipline. May have industry experience also.
Learning Outcomes	
<i>On successful completion of this module the learner will be able to:</i>	
#	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
Dependencies	
Module Recommendations	
No recommendations listed	
Co-requisite Modules	
No Co-requisite modules 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			
Overview of Data Mining and Machine Learning History and Evolution. Revision of data science methodologies: KDD, CRISP-DM. Data Security. Taxonomy and overview of data mining and machine learning techniques			
General data pre-processing and transformation strategies Intro to prediction. Identifying and Handling Missing Values. Looking for Outliers. Transformations for Single/Multiple Predictors. Adding/removing predictors. Binning . Feature Selection			
Prediction models evaluation Data Splitting and Sampling Methods (Holdout, Cross-fold Validation, Stratification, etc.). Model Tuning and Overfitting. Determining the best model			
Regression Models Quantitative Methods of Performance. The Variance/Bias Trade-off. Linear Regression			
Regression Models Partial Least Squares Regression. K-Nearest Neighbours Regression			
Regression Models Regression Trees. Model-based Regression Trees			
Regression Models Rule-based Models. Model Tuning via LASSO, ElastiNet, and similar. Computing Considerations			
Classification Models Logistic Regression. Linear Discriminant Analysis			
Classification Models K-Nearest Neighbours. Naïve Bayes			
Classification Models Decision Trees (e.g., C5.0, Random Forests, etc.)			
Unsupervised Machine Learning Notions of distance and similarity. Euclidian vs. non-Euclidian spaces. Clustering: k-means, k-medoids			
Unsupervised Machine Learning Clustering for outlier detection. Plotting and understanding clusters. Cluster evaluation measures: DBIndex, WSSSE, scree plots			
Assessment Breakdown			%
Coursework			100.00%
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 0 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 Weekly Contact Hours				48.00

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>	
Discussion Note:	