

## H8MACL: Machine Learning

Module Code:	H8MACL
Long Title	Machine Learning <b>APPROVED</b>
Title	Machine Learning
Module Level:	LEVEL 8
EQF Level:	6
EHEA Level:	First Cycle
Credits:	10
Module Coordinator:	Sophie Flanagan
Module Author:	ORLA LAHART
Departments:	School of Computing
Specifications of the qualifications and experience required of staff	Lecturer must have MSc or PhD degree in computer science or cognate discipline. Experience in lecturing machine learning and coding in Python. May also have industry experience. Lab Assistants are required for tutorials and they should have experience in Python coding and reasonable knowledge of machine learning techniques.
<b>Learning Outcomes</b>	
<i>On successful completion of this module the learner will be able to:</i>	
#	<b>Learning Outcome Description</b>
LO1	Comprehend, compare and contrast fundamental machine learning concepts and techniques
LO2	Comprehend and assess potential ethical implications of machine learning.
LO3	Extract, transform, explore and clean data in preparation for machine learning
LO4	Build and evaluate machine learning models on various problem domains
LO5	Summarise, critique and present results from machine learning for decision-making
<b>Dependencies</b>	
<b>Module Recommendations</b>	
No recommendations listed	
<b>Co-requisite Modules</b>	
No Co-requisite modules listed	
<b>Entry requirements</b>	

# H8MACL: Machine Learning

Module Content & Assessment			
Indicative Content			
<b>Introduction and Ethics in Machine Learning</b> Forms of learning (Supervised, Unsupervised, Reinforcement) Ethics in data sourcing & handling Review of regulatory & privacy components (including the Data Protection Act) Ethical implications of Machine Learning Methodologies (e.g., KDD, SEMMA & CRISP-DM) Review of basic data exploration statistics			
<b>Data Preprocessing</b> Data cleaning (i.e., handling missing values, outliers, noise data) Data integration (i.e., entity integration problem, and handling of redundant, correlated, duplicated, and conflicting data) Data transformation (i.e., normalization, binning, log transformation, scaling) Data reduction (i.e., dimensionality reduction like PCA and MCA, attribute subset selection, sampling)			
<b>Regression</b> Review of linear and multiple linear regression Assessing the model's accuracy Model selection (i.e., AIC and BIC) Measuring predictors' importance Subset selection Shrinkage methods			
<b>Classification</b> Introduction to classification Review of logistic regression Review of k-nearest neighbours Classification performance measures (e.g., Confusion matrix, precision and recall, ROC curve)			
<b>Model Evaluation and Selection</b> Bias-Variance trade-off Curse of dimensionality Evaluation methods (i.e., split validation, cross-validation, and bootstrap methods) Understanding, detecting and handling (massive) class imbalance			
<b>Unsupervised Learning</b> Introduction to unsupervised learning Notions of distance and similarity Partitioning methods (e.g., k-Means, k-Medoids) Plotting and understanding clusters Cluster evaluation metrics (i.e., DBIndex, silhouette coefficient)			
<b>Tree-Based Models</b> Decision Trees Regression and classification trees Node purity Pruning			
<b>Ensemble Models</b> Bagging Random Forest Boosting Stacking			
<b>Naïve Bayes Classification</b> Introduction to Naïve Bayes Bayes theorem Maximum a posteriori hypothesis Class conditional independence Naïve Bayes classifier			
<b>Introduction to Artificial Neural Networks</b> Feedforward neural network architecture Sigmoid activation function Backpropagation			
<b>Introduction to Deep Learning</b> Introduction to deep learning Deep feedforward networks Recurrent and recursive neural network Evaluation of deep learning			
<b>Text Analysis</b> Text tokenization Text normalization Feature extraction (e.g., Bag of words model, TF-IDF model) Sentiment analysis			
Assessment Breakdown			%
Coursework			100.00%
Assessments			
Full Time			
Coursework			
<b>Assessment Type:</b>	Formative Assessment	<b>% of total:</b>	Non-Marked
<b>Assessment Date:</b>	n/a	<b>Outcome addressed:</b>	1,2,3,4,5
<b>Non-Marked:</b>	Yes		
<b>Assessment Description:</b> 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.			
<b>Assessment Type:</b>	Assignment	<b>% of total:</b>	25
<b>Assessment Date:</b>	Week 4	<b>Outcome addressed:</b>	2,3
<b>Non-Marked:</b>	No		
<b>Assessment Description:</b> Learners may be provided with one or more datasets and will be required to apply suitable data cleaning, pre-processing and transformation operations on different attributes of the datasets. In addition, learners will be required to identify and discuss ethical implications of handling and applying machine learning to these datasets.			
<b>Assessment Type:</b>	Project	<b>% of total:</b>	75
<b>Assessment Date:</b>	Week 12	<b>Outcome addressed:</b>	1,3,4,5
<b>Non-Marked:</b>	No		
<b>Assessment Description:</b> This assessment will evaluate learner's comprehension of fundamental machine learning theory and concepts, their applicability and limitations to different problems. Learners will have to (1) identify a topic of interest and one relevant research or business question in that topic; (3) select at least two datasets useful to answer the question; (3) apply data pre-processing and transformation techniques to prepare the datasets for machine learning analysis; (4) perform exploratory analysis in these datasets; (5) apply, evaluate and optimize suitable machine learning techniques to extract knowledge from the selected datasets useful for a decision-making process in the topic of choice; (6) report and interpret the findings to answer the question of interest, and (7) elaborate a video presentation highlighting the project's main objectives, methodology, main findings, challenges faced.			
No End of Module Assessment			
No Workplace Assessment			
Reassessment Requirement			
<b>Coursework Only</b> This module is reassessed solely on the basis of re-submitted coursework. There is no repeat written examination.			
<b>Reassessment Description</b> This module is reassessed solely on the basis of re-submitted coursework. There is no repeat written examination.			

## H8MACL: Machine Learning

Module Workload				
Module Target Workload Hours 0 Hours				
Workload: Full Time				
Workload Type	Workload Description	Hours	Frequency	Average Weekly Learner Workload
Lecture	On-line/Classroom activities	24	Per Semester	2.00
Tutorial	Practical & Tutorial activities	36	Per Semester	3.00
Independent Learning	Independent Learning activities	190	Per Semester	15.83
Total Weekly Contact Hours				5.00

Module Resources	
<i>Recommended Book Resources</i>	
<p>Ethem Alpaydin. (2020), Introduction to Machine Learning, 4th ed.. MIT Press, Cambridge, MA, p.712, [ISBN: 978-0262043793].</p> <p>Shai Shalev-Shwartz, Shai Ben-David. (2015), Understanding Machine Learning, 2nd. Cambridge University Press, New York, NY, p.397, [ISBN: 978-1107512825].</p> <p>Sebastian Raschka, Vahid Mirjalili. (2019), Python Machine Learning, Packt Publishing, Birmingham, p.770, [ISBN: 978-1789955750].</p>	
<i>Supplementary Book Resources</i>	
<p>Kartik Hosanagar. (2019), A Human's Guide to Machine Intelligence, Penguin, London, p.272, [ISBN: 9780525560890].</p> <p>Trevor Hastie, Robert Tibshirani, Jerome Friedman. (2017), The Elements of Statistical Learning, 2nd ed.. Springer, New York, NY, p.767, [ISBN: 978-0387848570].</p> <p>John D. Kelleher, Brian Mac Namee, Aoife D'Arcy. (2020), Fundamentals of Machine Learning for Predictive Data Analytics, 2nd ed.. MIT Press, Cambridge, MA, p.856, [ISBN: 9780262044691].</p> <p>Wes McKinney. (2017), Python for Data Analysis, O'Reilly Media, Sebastopol, CA, p.550, [ISBN: 978-1491957660].</p> <p>Aurélien Géron. (2019), Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow, O'Reilly, Sebastopol, CA, p.819, [ISBN: 978-1492032649].</p> <p>CHARU C. AGGARWAL. (2019), Neural Networks and Deep Learning, Springer, p.524, [ISBN: 978-3030068561].</p> <p>Ian Goodfellow, Yoshua Bengio, Aaron Courville. (2016), Deep Learning, MIT Press, Cambridge, MA, p.775, [ISBN: 978-0262035613].</p> <p>Dipanjan Sarkar. (2016), Text Analytics with Python, Apress, Bangalore, p.385, [ISBN: 978-1484223871].</p> <p>Benjamin Bengfort, Tony Ojeda, Rebecca Bilbro. (2018), Applied Text Analysis with Python, O'Reilly Media, Sebastopol, CA, p.310, [ISBN: 978-1491963043].</p> <p>Valentina E. Balas, Sanjiban S. Roy, Dharmendra Sharma, Pijush Samui. (2019), Handbook of Deep Learning Applications, Springer, Cham, p.383, [ISBN: 978-3-030-11478-7].</p>	
<i>Recommended Article/Paper Resources</i>	
<p>Lipton, Z. C. &amp; Steinhardt, J.. (2019), Troubling trends in machine learning scholarship, Queue, 17, p.80, <a href="https://doi.org/10.1145/3317287.3328534">https://doi.org/10.1145/3317287.3328534</a></p> <p>Raschka, S, Patterson, J., &amp; Nolet, C.. (2020), Machine learning in Python: Main developments and technology trends in data science, machine learning, and artificial intelligence, Information, 11, p.193, <a href="https://doi.org/10.3390/info11040193">https://doi.org/10.3390/info11040193</a></p>	
<i>Supplementary Article/Paper Resources</i>	
<p>Joshi, A. V.. (2020), Emerging trends in machine learning, Machine Learning and Artificial Intelligence, p.12713, <a href="https://doi.org/10.1007/978-3-030-26622-6_13">https://doi.org/10.1007/978-3-030-26622-6_13</a></p> <p>Jordan, M. I. &amp; Mitchell, T. M.. (2015), Machine learning: Trends, perspectives, and prospects, Science, 349, p.25526, <a href="https://doi.org/10.1126/science.aaa8415">https://doi.org/10.1126/science.aaa8415</a></p>	
<i>Other Resources</i>	
<p>[Website], Machine Learning Repository, Center for Machine Learning and Intelligent Systems, <a href="https://archive.ics.uci.edu/ml/index.php">https://archive.ics.uci.edu/ml/index.php</a></p> <p>[Website], Kaggle platform for predictive modelling competitions, <a href="https://www.kaggle.com">https://www.kaggle.com</a></p> <p>[Website], Central Statistics Office, <a href="http://www.cso.ie">http://www.cso.ie</a></p> <p>[Website], Eurostat, <a href="http://ec.europa.eu/eurostat">http://ec.europa.eu/eurostat</a></p> <p>[Website], Data.gov, <a href="https://www.data.gov">https://www.data.gov</a></p> <p>[Website], Google Dataset Search, <a href="https://datasetsearch.research.google.co m/">https://datasetsearch.research.google.co m/</a></p> <p>[Website], Google Cloud Public Datasets, <a href="https://cloud.google.com/public-datasets/">https://cloud.google.com/public-datasets/</a></p>	
Discussion Note:	