H9DMML2: Data Mining and Machine Learning II

Module Code:	H9DMML2
Long Title	Data Mining and Machine Learning II APPROVED
Title	Data Mining and Machine Learning II
Module Level:	LEVEL 9
EQF Level:	7
EHEA Level:	Second Cycle
Credits:	10
Module Coordinator:	MICHAEL BRADFORD
Module Author:	Jenette Carson
Departments:	School of Computing
Specifications of the qualifications and experience required of staff	PhD/MSc degree in a computing or cognate discipline. May have industry experience also.

Learning Outcomes

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

#	Learning Outcome Description
LO1	Critically analyse advanced data mining and knowledge discovery methodologies in order to assess best practice guidance when applied to complex data mining problems
LO2	Investigate and evaluate key concepts and advanced data mining techniques and assess when to apply such techniques on complex datasets and problem domains.
LO3	Contextualise, research and utilise current data mining approaches, applications and technologies in order to provide strategies to address processing of datasets with a variety of characteristics
LO4	Critically review and apply appropriate data mining research and assess research methods

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

H9DMML2: Data Mining and Machine Learning II

Module Content & Assessment

Indicative Content
General Strategies Revisited Increasing data complexity and size with fundamental methods. . Considerations of Complexity on Computing Requirements
General Strategies Revisited Dimensionality Reduction (PCA, MCA, etc.). Feature Engineering. Measuring Predictor Importance
General Strategies Revisited Understanding, Detecting and Handling (massive) class imbalance.. Understanding Factors that can Affect Model Performance; e.g. Type III errors, selection bias, measurement errors, improper variable encoding. Ethically assessing biases.. .
Advanced Regression Models Regression revision, and penalised models
Advanced Regression Models Generalised Linear Modelling
Advanced Regression Models Automated Linear Modelling via Bagging and Boosting
Ensembles Ensembles:. Random Forest. Voting. Stacking.
Ensembles Bagging and Boosting Methods (e.g. XGBoost, AdaBoost, CART aggregation etc.)
Black Box Methods Support Vector Machines and Support Vector Regression
Black Box Methods Neural Networks:. Classic Topologies and Activation Functions. Back Propagation. Gradient Descent and Stochastic Gradient Descent. Hyperparameter Optimisation techniques
Black Box Methods Algorithmic Accountability, Ethical issues with black-box methods
Deep Regression Models A brief introduction to deep learning applied to regression problems (e.g. GLMNet). Special emphasis to be played on when these methods are(n't) appropriate (e.g. data volumes required).

Assessment Breakdown	%
Coursework	50.00%
End of Module Assessment	50.00%

Assessments

Full Time

Coursework

Assessment Type:	Formative Assessment	**% of total:**	Non-Marked
Assessment Date:	n/a	**Outcome addressed:**	1,2,3,4
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:**	50
Assessment Date:	n/a	**Outcome addressed:**	3,4
Non-Marked:	No		

Assessment Description:
Propose and execute a research project using data mining techniques as a team of 3-4 participants.

End of Module Assessment

Assessment Type:	Terminal Exam	**% of total:**	50
Assessment Date:	End-of-Semester	**Outcome addressed:**	1,2
Non-Marked:	No		

Assessment Description:
The examination will be a minimum of three hours in duration and may include a mix of: short answer questions, vignettes, essay based questions and case study based questions requiring the application of core module competencies. Marks will be awarded based on clarity, appropriate structure, relevant examples, depth of topic knowledge, and evidence of outside core text reading.

No Workplace Assessment

Reassessment Requirement

Repeat examination
Reassessment of this module will consist of a repeat examination. It is possible that there will also be a requirement to be reassessed in a coursework element.

Reassessment Description
The repeat strategy for this module is by repeat assessment/project that covers all learning outcomes.

H9DMML2: Data Mining and Machine Learning II

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)	202	Every Week	202.00
			Total Weekly Contact Hours	48.00

Module Resources

Recommended Book Resources

Hastie, T., Tibshirani, R. & Friedman, J.. (2016), The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed), Springer Series in Statistics.

James, G., Witten, D., Hastie, T. & Tibshirani, R.. (2017), An Introduction to Statistical Learning: with Applications in R, Springer Texts in Statistics.

Kuhn, M. & Johnson, K.. (2013), Applied Predictive Modeling, Springer.

Shalev-Shwartz, S. & Ben-David, S.. (2014), Understanding Machine Learning: From Theory to Algorithms, Cambridge University Press.

Supplementary Book Resources

Downey, B.. (2014), Think Stats: Exploratory Data Analysis, (2nd ed).

Goodfellow, I., Bengio, Y., & Courville, A.. (2016), Deep Learning, The MIT Press.

Hearty, J.. (2016), Advanced Machine Learning with Python, Packt Publishing Ltd.

Leskovec, J. Rajaraman, A., & Ullman, J.. (2014), Mining of Massive Datasets, Cambridge University Press.

Wickham, H. & Grolemund, G.. (2017), R for Data Science: Import, Tidy, Transform, Visualize, and Model Data, O'Reilly.

This module does not have any article/paper resources

Other Resources

[Website], Datacamp,
http://www.datacamp.com

[Website], KD Nuggest,
http://www.kdnuggets.com

[Website], R Bloggers,
http://www.r-bloggers.com

Discussion Note:	