H6DMML: Data Mining and Machine Learning

Module Code:	H6DMML
Long Title	Data Mining and Machine Learning APPROVED
Title	Data Mining and Machine Learning
Module Level:	LEVEL 6
EQF Level:	5
EHEA Level:	Short Cycle
Credits:	10
Module Coordinator:	Arghir Moldovan
Module Author:	Arghir Moldovan
Departments:	School of Computing
Specifications of the qualifications and experience required of staff	MSc and/or PhD degree in computer science 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	Contrast fundamental data mining and machine learning concepts and techniques, and discuss their applicability to different problems.
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	Summarise, critique and present the results from data mining and machine learning.

Dependencies
Module Recommendations
No recommendations listed
Co-requisite Modules
No Co-requisite modules listed

Entry requirements	Learners should have attained the knowledge, skills and competence gained from stage 1 of the BSc (Hons) in Data Science

H6DMML: Data Mining and Machine Learning

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 and ethical implications of machine learning 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:	Continuous 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:	Continuous Assessment	**% of total:**	40
Assessment Date:	n/a	**Outcome addressed:**	1,2
Non-Marked:	No		

Assessment Description:
This assessment will evaluate learner's comprehension of fundamental data mining and machine learning theory and concepts, their applicability and limitations to different problems. In addition, 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.

Assessment Type:	Project	**% of total:**	60
Assessment Date:	n/a	**Outcome addressed:**	1,2,3,4,5
Non-Marked:	No		

Assessment Description:
Learners will be assessed through a practical project that will evaluate all learning outcomes. Learners will have to identify or and extract one or more datasets; apply data pre-processing, transformation and exploration techniques; apply suitable machine learning techniques to extract knowledge from the datasets; and report and interpret the findings.

No End of Module Assessment

No Workplace Assessment

Reassessment Requirement

Coursework Only
This module is reassessed solely on the basis of re-submitted coursework. There is no repeat written examination.

Reassessment Description
The reassessment strategy for the Data Mining and Machine Learning module will consist of a project that will assess all learning outcomes. Students who fail the module will be afforded an opportunity to do the repeat project over the Summer months.

H6DMML: Data Mining and Machine Learning

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	Per Semester	2.00
Tutorial	Other hours (Practical/Tutorial)	24	Per Semester	2.00
Independent Learning	Independent learning (hours)	202	Per Semester	16.83
			Total Weekly Contact Hours	4.00

Module Resources

Recommended Book Resources

Witten, I. H., Frank, E., Hall, M. A. & Pal, C. J.. (2016), Data Mining: Practical machine learning tools and techniques (4th ed), Morgan Kaufmann.

Lantz, B.. (2015), Machine learning with R (2nd ed), Packt Publishing Ltd.

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.

Supplementary Book Resources

Mueller, A. C.. (2016), Introduction to machine learning with Python, O'Reilly.

Hofmann, M., & Klinkenberg, R.. (2013), RapidMiner: Data Mining Use Cases and Business Analytics Applications, CRC Press.

Han, J., Pei, J., & Kamber, M.. (2011), Data mining: concepts and techniques (3rd ed), Elsevier.

Berthold, M., & Hand, D. J.. (2003), Intelligent data analysis: an introduction, Springer Science & Business Media.

This module does not have any article/paper resources

Other Resources

[Website], UC Irvine Machine Learning Repository http://archive.ics.uci.edu/ml/.

[Website], Kaggle platform for predictive modelling competitions https://www.kaggle.com/.

[Website], Website: Datasets for Data Mining and Data Science http://www.kdnuggets.com/datasets/index. html.

Discussion Note:	