A8DMML: Data Mining & Machine Learning

Module Code:	A8DMML
Long Title	Data Mining & Machine Learning APPROVED
Title	Data Mining & Machine Learning
Module Level:	LEVEL 8
EQF Level:	6
EHEA Level:	First Cycle
Credits:	5
Module Coordinator:	Simon Caton
Module Author:	Madita Feldberger
Departments:	School of Computing
Specifications of the qualifications and experience required of staff	

Learning Outcomes	
On successful completion of this module the learner will be able to:	
#	**Learning Outcome Description**
LO1	Extract, transform, explore, and clean data in preparation for data mining and machine learning
LO2	Evaluate and apply statistical methods for prediction and forecasting in various problem domains
LO3	Build and evaluate data mining and machine learning models in various problem domains
LO4	Extract, interpret and evaluate information and knowledge from data for industry contexts
LO5	Articulate and evaluate Industry-focused questions using various data artefacts and methods from statistical learning, data mining and machine learning
LO6	Summarise, critique and present the results from data mining in various problem domains

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

Entry requirements	

A8DMML: Data Mining & Machine Learning

Module Content & Assessment

Indicative Content
Intro to data mining, key tools and core methodologies n/a
Data Understanding I Types of data (Categorical, Functional, Numerical, Hierarchical, Time Series etc.), Structured vs. Unstructured Data, Descriptive and Inferential Statistics Revisited for Data, Understanding, Data Mining, and Machine Learning Exploratory Data Analysis
Data Understanding II Identifying and Handling Missing Values and Outliers Feature Engineering, and Dimensionality Reduction (Principal Component Analysis and Linear Discriminant Analysis) Normalisation methods Sampling and under sampling
Univariate and multivariate regression Using Linear and Logistic Regression for Univariate and Multivariate Predictive Analytics, Applying Regression, Auto Regression and Vector Auto Regression for Time Series now- and forecasting
Time series I: Univariate data Using Linear and Logistic Regression for Univariate and Multivariate Predictive Analytics, Applying Regression, Auto Regression and Vector Auto Regression for Time Series now- and forecasting
Time Series II: Multivariate data Using Linear and Logistic Regression for Univariate and Multivariate Predictive Analytics, Applying Regression, Auto Regression and Vector Auto Regression for Time Series now- and forecasting
Clustering Evaluation measures for unsupervised methods Exclusive (e.g. k-means / k-medoids) and Fuzzy Clustering (e.g. c-means / c-mediods) using various distance measures.
Association Rule Mining Association Rule Mining
Introduction to Classification Models Evaluation measures for supervised methods Hold-out, k-fold cross validation, and model bootstrapping, K-nearest neighbours
Decision Trees Decision Trees: C5.0, CART, and Random Forests
Naïve Bayes and Intro to Bayesian Classification Naïve Bayes and principals of Bayesian Classification
Introduction to Text Mining Text (Pre)processing and Cleaning, Sentiment Analysis, Entity Extraction

Assessment Breakdown	%
Coursework	100.00%

Assessments

Full Time

Coursework

Assessment Type:	Formative Assessment	**% of total:**	Non-Marked
Assessment Date:	n/a	**Outcome addressed:**	
Non-Marked:	Yes		

Assessment Description:
Formative assessment will be included by the provision of class based problem solving exercises and short answer questions. Feedback will be provided individually or as a group in written and 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:	CA 1 (0380)	**% of total:**	20
Assessment Date:	n/a	**Outcome addressed:**	1,2
Non-Marked:	No		

Assessment Description:
The first test will assess apprentices' competence in data understanding and the application of regression methods to an unseen data set.

Assessment Type:	CA 2 (0390)	**% of total:**	20
Assessment Date:	n/a	**Outcome addressed:**	2,3
Non-Marked:	No		

Assessment Description:
The second test will assess apprentices' knowledge, understanding and practical competence in time series analysis and forecasting as well as unsupervised machine learning.

Assessment Type:	CA 3 (0420)	**% of total:**	20
Assessment Date:	n/a	**Outcome addressed:**	3,4
Non-Marked:	No		

Assessment Description:
The third test will assess apprentices' knowledge, understanding and practical competence in supervised machine learning.

Assessment Type:	Project (0050)	**% of total:**	40
Assessment Date:	n/a	**Outcome addressed:**	5,6
Non-Marked:	No		

Assessment Description:
Learners will be assessed through a team project with both practical and research elements.

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.

A8DMML: Data Mining & Machine Learning

Module Workload
Module Target Workload Hours 0 Hours

Module Resources

Recommended Book Resources

James, G., Witten, D., Hastie, T., & Tibshirani, R.. (2013), An introduction to statistical learning, Vol. 6. New York: Springer.

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.

Lantz, B.. (2013), Machine learning with R., Packt Publishing Ltd.

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

Supplementary Book Resources

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

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

Leskovec, J., Rajaraman, A., & Ullman, J. D.. (2014), Mining of massive datasets, Cambridge University Press.

Raschka, S.. Python machine learning, Packt Publishing Ltd.

This module does not have any article/paper resources

Other Resources

Stanford University,
http://infolab.stanford.edu/~ullman/mini ng/2008/index.html

UC IRVINE MACHINE LEARNING REPOSITORY,
http://archive.ics.uci.edu/ml/

KAGGLE: PLATFORM FOR PREDICTIVE MODELING COMPETITIONS,
https://www.kaggle.com/

Discussion Note:	