H8DWM: Data and Web Mining

Module Code:	H8DWM
Long Title	Data and Web Mining APPROVED
Title	Data and Web Mining
Module Level:	LEVEL 8
EQF Level:	6
EHEA Level:	First Cycle
Credits:	10
Module Coordinator:	Simon Caton
Module Author:	Margarete Silva
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	Analyse the characteristics of a data sets and their attributes, investigate what transformations and statistical operations can be carried out on each and identify factors that impact on data quality
LO2	Investigate a variety of data mining techniques, and identifying their practical applicability to various problem domains
LO3	Independently research current trends and developments in knowledge discovery related technologies and use this skill to critically analyse publications to assess the relative merits of various technologies
LO4	Investigate how web search engines crawl, index, rank web content, how the web is structured
LO5	Develop an in-depth knowledge of the fundamental web data mining concepts and techniques, and how previously acquired knowledge of data mining applies to the web

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

H8DWM: Data and Web Mining

Module Content & Assessment

Indicative Content

1. Data Analysis and Mining Overview (15%)
Data vs. information Data mining and machine learning Structural descriptions and rules for classification and association Exploration of sample datasets Fielded applications (e.g., ranking web pages, loan applications, screening images, load forecasting, machine fault diagnosis, market basket analysis) Generalization as search Data mining and ethics

2. Data Transformations (15%)
Attribute selection and discretization Projections (e.g., Principal component analysis, random projections, partial least-squares, text, time series) Sampling Handling dirty data

3. Knowledge Representation and Machine Learning Schemes (50%)
Tables Linear models Trees Rules based systems for knowledge representation Instance-based representation Inferring rudimentary rules Statistical modelling Historical evolution and foundations of AI Approaches to machine learning (e.g., decision tree learning, association rule learning, clustering) Utilising machine learning application software environments (e.g., Weka, R, RapidMiner etc.) for data mining and data visualisation

4. Extracting Data from the Web (20%)
Web crawler operations Search engines implementation Identification of search trends Search Engine Optimisation (SEO) Web usage, web content, and web structure mining Social media data mining

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

Assessments

Full Time

Coursework

Assessment Type:	Continuous Assessment (0200)	**% of total:**	20
Assessment Date:	n/a	**Outcome addressed:**	1,2,3
Non-Marked:	No		

Assessment Description:
Literature Review

Assessment Type:	Project (0050)	**% of total:**	30
Assessment Date:	n/a	**Outcome addressed:**	1,2,3
Non-Marked:	No		

Assessment Description:
Group Project

End of Module Assessment

Assessment Type:	Terminal Exam	**% of total:**	50
Assessment Date:	End-of-Semester	**Outcome addressed:**	1,2,3,4,5
Non-Marked:	No		

Assessment Description:
End-of-Semester Final Examination

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.

H8DWM: Data and Web Mining

Module Workload

Module Target Workload Hours 0 Hours

Workload: Full Time

Workload Type	Workload Description	Hours	Frequency	Average Weekly Learner Workload
Lecture	No Description	2	Every Week	2.00
Tutorial	No Description	2	Every Week	2.00
Independent Learning Time	No Description	6.5	Every Week	6.50
			Total Weekly Contact Hours	4.00

Module Resources

Recommended Book Resources

Bing Liu. Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data, Springer, [ISBN: 3642194591.].

Ian H. Witten, Eibe Frank, Mark A. Hall. Data Mining: Practical Machine Learning Tools and Techniques, Third Edition, Morgan Kaufmann, [ISBN: 0123748569.].

Matthew A. Russell. Mining the Social Web, O'Reilly Media, p.360, [ISBN: 1449388345].

Brett Lantz.. (2015), Machine learning with R, 2. Packt Pub, Birmingham, UK, p.454, [ISBN: 9781784393908].

Supplementary Book Resources

Michael R. Berthold (Editor), David J. Hand (Editor). Intelligent Data Analysis, Springer, [ISBN: 3642077072.].

Jiawei Han, Micheline Kamber, Jian Pei. Data Mining: Concepts and Techniques, Third Edition, Morgan Kaufmann, [ISBN: 0123814790.].

Rajaraman A., Ullman J.,. (2011), Mining of Massive Datasets, Free on-line edition available at: http://infolab.stanford.edu/~ullman/mmds.html. Cambridge Press.

Kevin Warwick. Artificial Intelligence: The Basics, Routledge, p.192, [ISBN: 0415564832].

Stuart J. Russell and Peter Norvig; contributing writers, Ernest Davis... [et al.]. (2010), Artificial intelligence, Prentice Hall, Upper Saddle River, N.J., [ISBN: 0136042597].

Pang-Ning Tan, Michael Steinbach, Vipin Kumar. (2006), Introduction to data mining, Pearson Addison Wesley, Boston, [ISBN: 0321321367.].

This module does not have any article/paper resources

Other Resources

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

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

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

Discussion Note: