

Short Title:	Algorithms for Data Science APPROVED
Full Title:	Algorithms for Data Science
Module Code:	ADSA H6013
ECTS credits:	10
NFQ Level:	9
Module Delivered in	1 programme(s)
Module Contributor:	Geraldine Gray
Module Description:	This module is aimed at learners who want to study advanced concepts relating to data science. Using both lectures and independent research, the module will address a number of issues relating to understanding and optimising the performance of data mining algorithms.
Learning Outcomes:	
<i>On successful completion of this module the learner will be able to</i>	
<ol style="list-style-type: none"> 1. Discuss in depth a variety of data mining techniques, and their applicability to various problem domains 2. Evaluate a business objective and related dataset to assess the appropriateness of a number data mining algorithms in achieving that objective 3. Work through the mining and evaluation stages of a data mining methodology, selecting the most appropriate mining technique, and optimising algorithm parameters to maximise performance 4. Independently research current trends and developments in knowledge discovery related technologies 5. Critically analyse relevant publications to assess the relative merits of methodologies used and conclusions made 6. Self-evaluate work done. 	

Module Content & Assessment

Indicative Content

Data Mining model functions

Classification and prediction, clustering, dependency modelling, sequence modelling and data summarisation. Matching the model function to the type of analysis required. Case studies of KDD applications.

Data Mining model representations

The algorithms underlining the mining model functions such as: Classification & Prediction: decision trees and rules, neural networks, SVM's, k-nearest neighbour, Bayesian networks, regression analysis. Clustering: self-organised networks, partitional clustering methods, hierarchical clustering, grid based methods, density based clustering, distance measures. Association analysis: Apriori, FP growth, generalised sequential pattern discovery. Model adaptations for Big Data.

Evaluation

Evaluation of classification results: precision, recall, confusion matrix, accuracy, Kappa, cost matrix, ROC, AUC, Lift and Gain charts, confidence, error functions. Evaluation of clustering results, difficulty in defining goodness of fit. Evaluation of dependency analysis: objective and subjective measures of rules interestingness.

Indicative Assessment Breakdown

	%
Course Work Assessment %	100.00%

Course Work Assessment %

Assessment Type	Assessment Description	Outcome addressed	% of total	Assessment Date
Reflective Journal	Students must prepare a portfolio of literary reviews and analysis covering a range of topics across all areas of the syllabus, and give an oral presentation of at least one of their research areas. Relevant topics include: • Research current trends in classification, and discuss in detail developments in at least one classification algorithm. • Research current trends in clustering, and discuss in detail developments in at least one clustering algorithm. • Research into one other data mining technique such as sequence analysis, association mining or visual data mining. • Research into evaluation techniques.	1,4,5	60.00	n/a
Practical/Skills Evaluation	Students will be presented with a number of datasets and mining objectives, designed to cover the range mining functions on the syllabus. From this information, students will be required to select candidate mining algorithms appropriate to the data type and mining objective; compare the results of these mining algorithms, and analyse the effectiveness of adjusting parameter settings on each algorithm. The deliverable would be a report justifying, evaluating and analysing the effectiveness of the algorithms used.	2,3	30.00	n/a
Reflective Journal	Students will be required to complete a self evaluation of work submitted, based on detailed criteria defined at the start of the semester.	6	5.00	n/a
Presentation	Present a selected topic from submitted literary reviews	1,4,5	5.00	n/a

No Final Exam Assessment %

Indicative Reassessment Requirement

Coursework Only

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

ITB reserves the right to alter the nature and timings of assessment

Indicative Module Workload & Resources

Indicative Workload: Full Time	
Frequency	Indicative Average Weekly Learner Workload
Every Week	2.00
Every Week	2.00
Every Week	6.00

Resources

Supplementary Book Resources
<p>Tan, Steinbach, Kumar 2014, <i>Introduction to Data Mining</i>, Pearson [ISBN: 1292026154]</p> <p>Markus Hofmann, Ralf Klinkenberg 2013, <i>Rapidminer: Data Mining Use Cases and Business Analytics Applications</i>, Chapman and Hall/CRC</p> <p>Jiawei Han, Micheline Kamber, Jian Pei 2013, <i>Data Mining: Concepts and Techniques, Third Edition</i>, Morgan Kaufmann [ISBN: 0123814790]</p> <p>Ian H. Witten, Eibe Frank, Mark A. Hall 2011, <i>Data Mining: Practical Machine Learning Tools and Techniques, Third Edition</i>, Morgan Kaufmann [ISBN: 0123748569]</p>
<i>This module does not have any article/paper resources</i>
Other Resources
<p>Journal: Science Direct <i>Computational Statistics and Data Analysis</i></p> <p>Journal: InfoTrack <i>IEEE Transactions on Knowledge and Data Engineering</i></p>

Module Delivered in

Programme Code	Programme	Semester	Delivery
BN_KADSA_R	Master of Science in Computing in Applied Data Science & Analytics	1	Mandatory