

**This Version is No Longer Current**  
The latest version of this module is available [here](#)

## MODULE DESCRIPTOR

### Module Title

Data Mining

Reference	CMM510	Version	8
Created	February 2018	SCQF Level	SCQF 11
Approved	April 2005	SCQF Points	15
Amended	March 2018	ECTS Points	7.5

### Aims of Module

To provide students with an understanding of the main principles underlying Data Mining and Machine Learning techniques and the ability to apply current Data Mining and Machine Learning tools to real datasets.

### Learning Outcomes for Module

On completion of this module, students are expected to be able to:

- 1 Discuss, compare and contrast the advantages and disadvantages of applying a specific data mining technique to a given learning task.
- 2 Use a toolkit to develop a data mining application tailored to a given learning task and evaluate the results obtained.
- 3 Effectively interpret the results of learning through an understanding of the strengths and limitations of data mining technology and the selection of an appropriate evaluation technique.
- 4 Demonstrate knowledge of the state-of-the-art in data mining and an awareness of current areas of research.
- 5 Apply and, where necessary, adapt an appropriate data mining technique to a given problem.

### Indicative Module Content

Basic data mining concepts. Implementation of fundamental learning approaches. Rules involving relations; incorporating domain knowledge in learning. Advanced techniques for evaluating learned concepts. Calculation of confidence intervals for predictive performance. Comparison of data mining schemes. Paired t-test. Minimum Description Length principle and its application to clustering. Attribute selection. Combining learned results. Applications.

### Module Delivery

This is a lecture based course, supplemented with laboratory sessions, where a data mining toolkit is applied to varied learning tasks and tutorials where additional understanding is gained through practical exercises which supplement the lectures.

**Indicative Student Workload**

	Full Time	Part Time
Contact Hours	48	48
Non-Contact Hours	102	102
Placement/Work-Based Learning Experience [Notional] Hours	N/A	N/A
TOTAL	150	150
Actual Placement hours for professional, statutory or regulatory body		

**ASSESSMENT PLAN**

If a major/minor model is used and box is ticked, % weightings below are indicative only.

**Component 1**

Type:	Examination	Weighting:	50%	Outcomes Assessed:	1, 3, 4, 5
Description:	Closed book examination.				

**Component 2**

Type:	Practical Exam	Weighting:	50%	Outcomes Assessed:	2
Description:	Assessed lab.				

**MODULE PERFORMANCE DESCRIPTOR****Explanatory Text**

The calculation of the overall grade for this module is based on 50% weighting of C1 and 50% weighting of C2. An overall minimum grade D is required to pass the module

		Examination:						NS
		A	B	C	D	E	F	
Practical Exam:	A	A	A	B	B	C	E	
	B	A	B	B	C	C	E	
	C	B	B	C	C	D	E	
	D	B	C	C	D	D	E	
	E	C	C	D	D	E	E	
	F	E	E	E	E	E	F	
NS		Non-submission of work by published deadline or non-attendance for examination						

**Module Requirements**

Prerequisites for Module	None except for course entry requirements.
Corequisites for module	None.
Precluded Modules	None.

**INDICATIVE BIBLIOGRAPHY**

- 1 OLSON, DL. (2019), Descriptive Data Mining, 2nd Edition. Springer Nature.
- 2 WITTEN, IH. FRANK, E. HALL, MA, PAL, C. (2017) Data Mining - Practical Machine Learning Tools and Techniques, 4th ed. Morgan Kaufman.
- 3 NISBET, R. MINER, G. YALE, KP. (2018), Handbook of Statistical Analysis and Data Mining Applications, 2nd Edition, Academic Press.
- 4 AGGARWAL, CC. (2015). Data Mining: the Textbook. Springer.
- 5 CICHOSZ, P. (2015) Data Mining Algorithms: Explained Using R, Wiley.
- 6 JAMES, G. WITTEN, D. HASTIE, T. TIBSHIRANI, R. (2021) An introduction to statistical learning : with applications in R. Springer