

## Module 7: Data Mining Algorithms and Techniques

<b>Stage</b>	1						
<b>Semester</b>	2						
<b>Module Title</b>	Data Mining Algorithms and Techniques						
<b>Module Number</b>	7						
<b>Module Status</b>	Mandatory						
<b>Module ECTS Credits</b>	10						
<b>Module NFQ level</b>	9						
<b>Pre-Requisite Module Titles</b>	None						
<b>Co-Requisite Module Titles</b>	None						
<b>Capstone Module</b>	No						
<b>List of Module Teaching Personnel</b>	Mr Alex Cronin Ms Aine McManus Dr Waseem M. Akhtar						
<b>Contact Hours</b>	<b>Non-contact Hours</b>						<b>Total Effort (hours)</b>
72	128						200
<b>Lecture</b>	<b>Practical</b>	<b>Tutorial</b>	<b>Seminar</b>	<b>Assignment</b>	<b>Placement</b>	<b>Independent Work</b>	
36	36			60		68	
<b>Allocation of Marks (Within the Module)</b>							
	<b>Continuous Assessment</b>	<b>Project</b>	<b>Practical</b>	<b>Final Examination</b>	<b>Total</b>		
<b>Percentage Contribution</b>	50			50	100		

### Intended Module Learning Outcomes

On successful completion of this module the learner will be able to:

1. Demonstrate advanced knowledge of the data mining process and its applications from a problem solving perspective
2. Demonstrate advanced knowledge of different data mining tasks and algorithms and where they should be applied
3. Critically evaluate the different models of data mining in respect of their use, accuracy and performance
4. Implement various data mining algorithms and techniques
5. Employ appropriate data mining tools on large datasets to generate information
6. Analyse data mining results and determine if improvements can be made.

## Module Objectives

This module aims to give learners a thorough understanding of different data mining techniques, algorithms and tools necessary to infer information from large datasets. The learners will understand the underpinning concepts and principles that make these algorithms work. The learners will encounter and will implement various data mining techniques.

The learners will be expected to take different sets of data and apply the appropriate data mining techniques to them to infer information. From this they should be able to hypothesise information about their datasets and test those hypotheses in a controlled scientific manner.

## Module Curriculum

- **Overview of Data Mining**  
Learning from data: why, what and how? From data to information to knowledge: Issues and Paradigms, explaining the past and predicting the future, finding interesting structures in data, machine learning
- **Data Mining Tasks**  
Exploratory Data Analysis (EDA), Descriptive Modelling, Predictive Modelling (Classification and Regression), Segmentation (clustering), Discovering Patterns and Rules, Retrieval by Content, Visualisation.
- **Components of Data Mining Algorithms**  
The task the algorithm to address (e.g. classification, clustering etc.), Model or Pattern Structure fitting to the data (e.g. linear regression model), Score functions used to judge the quality of the fitted model or patterns (e.g. accuracy, BIC, etc.), Optimizations and Search Methods, Data Management Strategy used for storing, indexing and retrieving data.
- **Measurements and Models**  
Nature and type of datasets and measurements, dealing with uncertainty, models for structured data, statistical models, statistical data mining, predictions, Bayesian Probability model
- **Data Mining Algorithms**  
Overview of data mining algorithms (e.g. CART, Back propagation, A Priori) in terms of tasks, structure, score function, search methods,
- **Data Mining in Regression Framework**  
Regression Modelling, Linear Regression, ANOVA tables, Multiple Regression, Logistic Regression
- **Data Mining Tools:**  
The Weka Data Mining Workbench; the Explorer, the Knowledge Flow

Interface, the Experimenter

- **Data Mining Real World Applications and Case Studies**

An overview of real world applications and case studies of data mining relevant to the content of this module

## **Reading Lists and other learning materials**

### **Recommended Reading**

Witten I H, Frank E, Hall M A., 2011, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann Publishes Inc.

Larose D T., 2006, *Data Mining Methods and Models*, Wiley-Blackwell

Hand D, Mannila H, Smyth P., 2001, *Principles of Data Mining*, MIT Press

Kantardzic M., 2011, *Data Mining: Concepts, Models, Methods and Algorithms*, Wiley Blackwell

### **Secondary Reading**

Yanchang Z, Yonhua C., 2013, *Data Mining Applications with R*, Academic Press

Additional reading material will include peer reviewed research papers and most recent and relevant case studies.

## **Module Learning Environment**

Lectures are carried out in class rooms / lecture halls in the College. Lab tutorials are carried out in computer labs throughout the Campus. All have the software required to deliver the programme.

### **Library**

All learners have access to an extensive range of physical and electronic (remotely accessible) library resources. The library monitors and updates its resources on an on-going basis, in line with the College's Library Acquisition Policy. Lecturers update reading lists for this course on an annual basis as is the norm with all courses run by Griffith College.

## **Module Teaching and Learning Strategy**

Learners are taught using a combination of classes and practicals. Classes are used to explain the concepts and exemplify the techniques.

Practicals give the learners the opportunity to implement the ideas that are discussed in class and to practice mobile development.

In addition to classes and practicals, learners will need to put in at least two hours homework each week.

### Module Assessment Strategy

This module is 50% continuous assessment consisting of three assessment elements. There will also be a 50% examination.

Element No.	Weighting	Type	Description	Learning Outcomes Assessed
1.	15%	Workbooks Submission	This will involve a series of tutorials given on weekly basis. They aim at enhancing the understanding of concepts and ideas using practical problems and case studies.	2, 4
2.	15%	Assignment	This will involve a programming work and/or producing a written technical paper format report.	1,3
3.	30%	Programming and report writing	Implementation of a data mining project from start to finish; this will involve selecting a topic and a relevant dataset; defining the aims and objectives of mining; designing and implementing right mining techniques and reporting the results.	4, 5, 6
4.	50%	Examination	End of module examination	1, 2, 3, 5, 6