

620-639 Data Mining

Credit Points:	12.50
Level:	9 (Graduate/Postgraduate)
Dates & Locations:	2009, This subject commences in the following study period/s: Semester 2, - Taught on campus.
Time Commitment:	Contact Hours: 36 hours comprising one two-hour lecture per week and one one-hour practical class per week. Total Time Commitment: Not available
Prerequisites:	None
Corequisites:	None
Recommended Background Knowledge:	It is recommended students complete a second year statistics subject (such as 620-202 [2008] Statistics or its equivalent) and have had some exposure to computer packages.
Non Allowed Subjects:	None
Core Participation Requirements:	It is University policy to take all reasonable steps to minimise the impact of disability upon academic study and reasonable steps will be made to enhance a student's participation in the University's programs. Students who feel their disability may impact upon their participation are encouraged to discuss this with the subject coordinator and the Disability Liaison Unit.
Coordinator:	Dr Owen Dafydd Jones
Subject Overview:	<p>Data Mining refers to the management and analysis of large data sets.</p> <p>Data Mining became possible with the advent of large-scale data collection and the computing power necessary to process it. It involves all of the following steps</p> <ol style="list-style-type: none"> 1. Data Warehousing 2. Data Cleaning 3. Data Description and Visualisation 4. Data Analysis and Interpretation <p>This course deals only with step 4 of the Data Mining process: data analysis and interpretation. It considers techniques for Rule Finding, Classification, Regression and Clustering. The themes that run through the course are:</p> <ol style="list-style-type: none"> 1. Model fitting and selection and how to avoid overfitting 2. Scalable algorithms that can be used with very large data sets 3. How to accommodate high-dimensional data 4. Actionability and interpretability of models
Objectives:	<p>After completing this subject, students should:</p> <ul style="list-style-type: none"> # understand many of the techniques used to analyse large data sets; # have acquired skills and techniques widely used in modern data mining; and # have gained the ability to pursue further studies in this and related areas.
Assessment:	Up to 40 pages of written assignments (20%: two assignments worth 10% each, due mid and late in semester), a three-hour written examination (80%, in the examination period).
Prescribed Texts:	None.
Recommended Texts:	TBA.

Breadth Options:	This subject is not available as a breadth subject.
Fees Information:	Subject EFTSL, Level, Discipline & Census Date, http://enrolment.unimelb.edu.au/fees
Generic Skills:	<p>Upon completion of this subject, students should develop:</p> <ul style="list-style-type: none"># problem-solving skills (especially through tutorial exercises and assignments) including engaging with unfamiliar problems and identifying relevant strategies;# analytical skills including the ability to construct and express logical arguments and to work in abstract or general terms to increase the clarity and efficiency of the analysis; and# ability to work in a team, through interactions with other students.
Related Majors/Minors/ Specialisations:	R05 RM Master of Science - Mathematics and Statistics