MONASH University
Information Technology

FIT3073
Data mining

Unit Guide

Semester 1, 2010

Last updated: 19 Feb 2010

Table of Contents

FIT3073 Data mining - Semester 1, 2010

Chief Examiner:

Dr Grace Rumantir
Lecturer
Phone: +61 3 990 31965
Fax: +61 3 990 31077

Contact hours: Thursday and Friday 11am-12pm and 4-5pm

Lecturer(s) / Leader(s):

Caulfield

Ms Oshadi Alahakoon
Fax: +61 3 990 31077

Introduction

Welcome to FIT3073 Data Mining for semester 1, 2010.

Unit synopsis

This unit provides an overview of the techniques used to search for knowledge within a data set using both supervised and unsupervised learning. The techniques include Classification, Prediction, Clustering, Association discovery, Time sequence discovery, Sequential pattern discovery, Visualization, Statistical Methods, Decision Trees, Rule based methods, Neural networks, Machine learning, Genetic Algorithms and Fuzzy Systems. Students are able to choose an appropriate technique to suit a particular situation.

Learning outcomes

At the completion of this unit students will:

- develop student knowledge of the techniques and methods for data exploration in large databases, both those currently being used and those which are presently being researched;
- be familiar with the currently available techniques for the extraction of information from large databases;
- have an understanding of the purpose of data mining;
- have an understanding of the major techniques for data mining;
- have developed the knowledge to allow them to apply a process to the acquisition of knowledge from a data store;
- appreciate the potential for data mining techniques to permit access to private information and understand this must be done only in the proper context;
- practice ethical behaviour when when conducting data mining exercises;
- have developed the skill to choose an appropriate technique for a particular situation;
- have the skills to use a number of implementations of data mining software.

Contact hours

2 hrs lectures/wk, 2 hrs laboratories/wk

Workload

- two-hour lecture
- two-hour tutorial
- minimum of 3-4 hours of personal study per week.

Unit relationships

Prerequisites

FIT1004 or CSE2132 or equivalent

Prohibitions

CSE3212

Teaching and learning method

Teaching approach

- Explanation and discussion of the theoretical aspect of the unit will be conducted in the lecture.
- Practical aspect that support the theoretical aspects of the unit will be conducted in the tutorial.
- Reading the supplied articles or book chapters will be undertaken by students during their personal study hours.

Timetable information

For information on timetabling for on-campus classes please refer to MUTTS, http://mutts.monash.edu.au/MUTTS/

Tutorial allocation

On-campus students should register for tutorials/laboratories using the Allocate+ system: http://allocate.its.monash.edu.au/

Unit Schedule

Week	Date*	Topic	Key dates
1	01/03/10	Introduction to Data Mining	
2	08/03/10	Models in Data Mining	
3	15/03/10	Model Representation and Evaluation	
4	22/03/10	Data Preparation	
5	29/03/10	Data Mining Process	
Mid semester break			
6	12/04/10	Classification Algorithm	
7	19/04/10	Clustering	
8	26/04/10	Unit Test	Unit Test (20%)
9	03/05/10	Association Rules I	
10	10/05/10	Association Rules II	
11	17/05/10	Web Mining	Assignment Due (20%)
12	24/05/10	Neural Networks and Data Mining	
13	31/05/10	Revision	

*Please note that these dates may only apply to Australian campuses of Monash University. Off-shore students need to check the dates with their unit leader.

Unit Resources

Prescribed text(s) and readings

There is no prescribed text for this unit

Recommended text(s) and readings

- Roiger R.J. & Geatz M.W. (2003) Data Mining: A Tutorial-Based Primer, Addison-Wesley/Pearson Education Inc.
- Berry J.A. & Linoff G. (2000) Mastering Data Mining: The Art and Science of Customer Relationship Management, John Wiley & Sons, Inc.
- Berry J.A. & Linoff G. (1997) Data Mining Techniques: For Marketing, Sales, and Customer Support, John Wiley & Sons, Inc.
- Cabena P., Hadjinian P., Stadler R., Verhees J. & Zanasi A.(1998) Discovering Data Mining: From Concept to Implementation, Prentice Hall
- Dunham M.H. (2003) Data Mining:Introductory and Advanced Topics, Prentice-Hall/Pearson Education Inc.
- Han J. & Kamber M. (2000) Data Mining: Concepts and Techniques, Morgan Kaufmann - available as e-book from Monash Library
- Kennedy R.L., Lee Y., Van Roy B., Reed C.D. & Lippman R.P. (1997) Solving Data Mining Problems Through Pattern Recognition, Prentice Hall
- Witten I.H. & Frank E. (2000) Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufmann - available as e-book from Monash Library

Required software and/or hardware

WEKA Data Mining Software, can be downloaded from http://www.cs.waikato.ac.nz/ml/weka/

Equipment and consumables required or provided

Students studying off-campus are required to have the minimum system configuration specified by the Faculty as a condition of accepting admission, and regular Internet access. On-campus students, and those studying at supported study locations may use the facilities available in the computing labs. Information about computer use for students is available from the ITS Student Resource Guide in the Monash University Handbook. You will need to allocate up to **n** hours per week for use of a computer, including time for newsgroups/discussion groups.

Study resources

Study resources we will provide for your study are:

- Lecture and tutorial notes will be available on MUSO.

Assessment

Overview

Examination (3 hours): 60%; In-semester assessment: 40%

Faculty assessment policy

To pass a unit which includes an examination as part of the assessment a student must obtain:

- 40% or more in the unit's examination, and
- 40% or more in the unit's total non-examination assessment, and
- an overall unit mark of 50% or more.

If a student does not achieve 40% or more in the unit examination or the unit non-examination total assessment, and the total mark for the unit is greater than 50% then a mark of no greater than 49-N will be recorded for the unit.

Assignment tasks

Assignment coversheets

Assignment coversheets are available via "Student Forms" on the Faculty website:
http://www.infotech.monash.edu.au/resources/student/forms/
You MUST submit a completed coversheet with all assignments, ensuring that the plagiarism declaration section is signed.

Assignment submission and return procedures, and assessment criteria will be specified with each assignment.

- **Assignment task 1**

  Title:
  Unit Test
  Description:
  Mid-semester test.
  Weighting:
  20%
  Due date:
  Week 8 lecture (27 April 2010)

- **Assignment task 2**

  Title:
  Application of Data mining
  Description:
  Based on a given data set, students will be expected to perform data mining and write a report to explain the process undertaken.
  Weighting:
  20%
  Due date:
  Tuesday, 18 May 2010 (Week 11), at 11.55pm

Examination

- **Weighting:** 60%
  Length: 3 hours
  Type (open/closed book): Closed book

See Appendix for End of semester special consideration / deferred exams process.

Due dates and extensions

Please make every effort to submit work by the due dates. It is your responsibility to structure your study program around assignment deadlines, family, work and other commitments. Factors such as normal work pressures, vacations, etc. are not regarded as appropriate reasons for granting extensions. Students are advised to NOT assume that granting of an extension is a matter of course.

Students requesting an extension for any assessment during semester (eg. Assignments, tests or presentations) are required to submit a Special Consideration application form (in-semester exam/assessment task), along with original copies of supporting documentation, directly to their lecturer within two working days before the assessment submission deadline. Lecturers will provide specific outcomes directly to students via email within 2 working days. The lecturer reserves the right to refuse late applications.

A copy of the email or other written communication of an extension must be attached to the assignment submission.

Refer to the Faculty Special consideration webpage or further details and to access application forms: http://www.infotech.monash.edu.au/resources/student/equity/special-consideration.html

Late assignment

Assignments received after the due date will be subject to a penalty of **10% day of the total available marks in the submission.**

Return dates

Students can expect assignments to be returned within two weeks of the submission date or after receipt, whichever is later.

Appendix

Please visit the following URL: http://www.infotech.monash.edu.au/units/appendix.html for further information about:

- Continuous improvement
- Unit evaluations
- Communication, participation and feedback
- Library access
- Monash University Studies Online (MUSO)
- Plagiarism, cheating and collusion
- Register of counselling about plagiarism
- Non-discriminatory language
- Students with disability
- End of semester special consideration / deferred exams