



**MONASH** University  
Information Technology

**FIT3002**  
**Applications of data mining**

**Unit Guide**

**Semester 1, 2010**

The information contained in this unit guide is correct at time of publication. The University has the right to change any of the elements contained in this document at any time.

*Last updated: 12 Feb 2010*

# Table of Contents

<u><a href="#">FIT3002 Applications of data mining - Semester 1, 2010</a></u> .....	1
<u>Chief Examiner:</u> .....	1
<u>Lecturer(s) / Leader(s):</u> .....	1
<u>Gippsland</u> .....	1
<u>South Africa</u> .....	1
<u>Malaysia</u> .....	1
<u>Introduction</u> .....	2
<u>Unit synopsis</u> .....	2
<u>Learning outcomes</u> .....	2
<u>Contact hours</u> .....	2
<u>Workload</u> .....	2
<u>Unit relationships</u> .....	3
<u>Prerequisites</u> .....	3
<u>Prohibitions</u> .....	3
<u>Teaching and learning method</u> .....	4
<u>Teaching approach</u> .....	4
<u>Timetable information</u> .....	4
<u>Tutorial allocation</u> .....	4
<u>Off-Campus Learning or flexible delivery</u> .....	4
<u>Unit Schedule</u> .....	4
<u>Improvements to this unit</u> .....	5
<u>Unit Resources</u> .....	6
<u>Prescribed text(s) and readings</u> .....	6
<u>Recommended text(s) and readings</u> .....	6
<u>Required software and/or hardware</u> .....	6
<u>Equipment and consumables required or provided</u> .....	7
<u>Study resources</u> .....	7
<u>Assessment</u> .....	8
<u>Overview</u> .....	8
<u>Faculty assessment policy</u> .....	8
<u>Assignment tasks</u> .....	8
<u>Examination</u> .....	9
<u>Due dates and extensions</u> .....	9
<u>Late assignment</u> .....	9
<u>Return dates</u> .....	10
<u>Appendix</u> .....	11

# **FIT3002 Applications of data mining - Semester 1, 2010**

## **Chief Examiner:**

### **Associate Professor Kai Ming Ting**

Director of Undergraduate Studies

Phone: +61 3 990 26241

## **Lecturer(s) / Leader(s):**

### **Gippsland**

### **Associate Professor Kai Ming Ting**

Director of Undergraduate Studies

Phone: +61 3 990 26241

### **South Africa**

### **Mr Neil Manson**

Lecturer

Phone: +27 11 950 4035

Fax: +27 11 950 4033

### **Malaysia**

**Elsa Phung**

## Introduction

Welcome to FIT3002 Applications of Data Mining. This 6 point unit is core to BITS Business Systems major program and elective to all other undergraduate degree programs in the Faculty of IT. The unit has been designed to provide you with an overview of data mining: its needs and motivation, process, basic principles, operations, case studies, key and emerging application areas, hands-on experience using data mining tools, and an understanding of current research issues.

## Unit synopsis

In the modern corporate world, data is viewed not only as a necessity for day-to-day operation, it is seen as a critical asset for decision making. However, raw data is of low value. Succinct generalisations are required before data gains high value. Data mining produces knowledge from data, making feasible sophisticated data-driven decision making. This unit will provide students with an understanding of the major components of the data mining process, the various methods and operations for data mining, knowledge of the applications and technical aspects of data mining, and an understanding of the major research issues in this area.

## Learning outcomes

At the completion of this unit students will have -  
A knowledge and understanding of:

- the motivation and the need for data mining;
- characteristics of major components of the data mining process;
- the basic principles of methods and operations for data mining;
- case studies to bridge the connection between hands-on experience and real-world applications;
- key and emerging application areas;
- current major research issues.

Developed the skills to

- use data mining tools to solve data mining problems.

## Contact hours

2 hrs lectures/wk, 2 hrs laboratories/wk

## Workload

For on campus students, workload commitments are:

- two-hour lecture and
- two-hour tutorial (or laboratory) (requiring advance preparation)
- a minimum of 2-3 hours of personal study per one hour of contact time in order to satisfy the reading and assignment expectations.
- You will need to allocate up to 5 hours per week in some weeks, for use of a computer, including time for newsgroups/discussion groups.

Off-campus students generally do not attend lecture and tutorial sessions, however, you should plan to spend equivalent time working through the relevant resources and participating in discussion groups

FIT3002 Applications of data mining - Semester 1, 2010

each week.

## **Unit relationships**

### **Prerequisites**

FIT1004 or equivalent

### **Prohibitions**

CSE3212, GCO3828

## Teaching and learning method

### Teaching approach

This unit will be delivered via a weekly two-hour lecture. Lecturers may go through specific examples, give demonstrations and present slides that contain theoretical concepts.

In tutorials/practicals students will discuss in-depth fundamental and interesting aspects about data mining and have hands-on experience using data mining tools. The tutorials/practicals are particularly useful in helping students consolidate concepts and practise their problem solving skills.

### 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/>

### Off-Campus Learning or flexible delivery

Off-Campus students should treat the Unit Book (consisting of 12 modules) as their primary source for self-directed study. The modules contain text which is directed to leading you through the learning for each week. Also refer to the Unit Study Plan on the unit web page for further detail.

Online Discussion Forums are provided for the primary purpose of enabling off-campus students as well as on-campus students to engage with each other and the lecturer in Australia. The lecturer will expect all students to read these forums at least twice per week. In the forums, you may ask questions about the topics or exercises of each module, or to clarify interpretation of assignment tasks and marking criteria.

### Unit Schedule

Week	Date*	Topic	Key dates
1	01/03/10	The Need for Data Mining	
2	08/03/10	Model Building	
3	15/03/10	Model Representation	
4	22/03/10	Data Mining Process	
5	29/03/10	Performance Evaluation	
Mid semester break			
6	12/04/10	Engineering the input and output	
7	19/04/10	Algorithms	
8	26/04/10	Implementation Issues	
9	03/05/10	Market basket analysis	
10	10/05/10	Cluster Analysis & Anomaly Detection	

11	17/05/10	Case Studies	
12	24/05/10	Data Mining Applications & Research Issues (additional reading)	
13	31/05/10	N.A.	

\*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.

### **Improvements to this unit**

A topic on "cluster analysis and anomaly detection" and additional reading on application have been added in 2008 to broaden students' knowledge in this area.

## Unit Resources

### Prescribed text(s) and readings

Witten, I.H. & Frank, E. Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann Publishers, second edition, 2005. ISBN: 0-12-088407-0

Text books are available from the [Monash University Book Shops](#). Availability from other suppliers cannot be assured. The Bookshop orders texts in specifically for this unit. You are advised to purchase your text book early.

### Recommended text(s) and readings

1. Kennedy, R.L., Lee, Y. Roy, B.V., Reed, C.D. & Lippman, R.P., Solving Data Mining Problems through Pattern Recognition, Prentice Hall, 1998.
2. Cabena, P., Hadjinian, P., Stadler, R., Verhees, J. & Zanasi, A., Discovering Data Mining: from concept to implementation, Prentice Hall, 1997.
3. Berry, J.A.M. & Linoff, G. Data Mining Techniques for Marketing, Sales, and Customer Support, John Wiley & Sons, 1997.
4. Tan, P-N, Steinbach, M. & Kumar, V. Introduction to Data Mining, Addison Wesley, 2006.
5. Han, J. & Kamber, M. Data Mining: Concepts and Techniques, Morgan Kaufmann, Second Edition, 2006.
6. Dunham, M.H., Data Mining: Introductory and Advance Topics, Pearson Education, 2003.
7. Groth, R., Data Mining: Building competitive advantage, Prentice Hall, 2000.
8. Berson, A., Smith, S. & Thearling, K., Building Data Mining Applications for CRM, McGraw Hill. 2000.
9. Berry, J.A.M. & Linoff, G. Mastering Data Mining: The Art and Science of Customer Relationship Management, John Wiley & Sons, 2000.
10. Mena, J. Data Mining Your Website. Digital Press, 1999.
11. Westphal, C. & Blaxton, T. Data Mining Solutions, John Wiley & Sons, 1998.
12. Quinlan, J.R. C4.5: Program for Machine Learning, Morgan Kaufmann, 1993.
13. Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P. & Uthurusamy, R. Advances in Knowledge Discovery and Data Mining, AAAI Press/MIT Press, 1996.

### Required software and/or hardware

1. Software Title: WEKA, version 3.6
2. Magnum OPUS version 4



Both are freeware and they are made available in the GSIT CD-ROM or retrievable from the websites stated in the relevant unit home page.

## **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 6 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:

A Unit Book containing the unit information and 12 Study Guides.

A CD-ROM sent at the start of the semester, with software required for all units.

## 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.

The unit is assessed with two assignments and a three-hour closed book examination. To pass the unit you must pass each individual hurdle:

- 40% or more in the unit's examination and
- 40% or more in the unit's 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 assessment then a mark of no greater than 44-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:**

Assignment 1

**Description:**

This assignment requires students to use the data mining tool, WEKA, to build a good model from a given set of data, and write a report describing the data mining process.

**Weighting:**

20%

**Due date:**

14 April 2010

• **Assignment task 2**

**Title:**

Assignment 2

**Description:**

This assignment requires students to use the data mining tool, WEKA, to explore several models and then choose one that will likely to produce the largest profit within the budgetary constraint for a mass mailing campaign.

**Weighting:**

20%

**Due date:**

5 May 2010

## 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 5% a day. Assignments received later than one week after the due date will not be accepted.

## **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