



**MONASH** University  
Information Technology

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

**Unit Guide**

**Semester 1, 2011**

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: 02 Mar 2011*

# Table of Contents

<a href="#"><u>FIT3002 Applications of data mining - Semester 1, 2011</u></a> .....	1
<a href="#"><u>Mode of Delivery</u></a> .....	1
<a href="#"><u>Contact Hours</u></a> .....	1
<a href="#"><u>Workload</u></a> .....	1
<a href="#"><u>Unit Relationships</u></a> .....	1
<a href="#"><u>Prohibitions</u></a> .....	1
<a href="#"><u>Prerequisites</u></a> .....	1
<a href="#"><u>Chief Examiner</u></a> .....	2
<a href="#"><u>Campus Lecturer</u></a> .....	2
<a href="#"><u>Gippsland</u></a> .....	2
<a href="#"><u>South Africa</u></a> .....	2
<a href="#"><u>Sunway</u></a> .....	2
<a href="#"><u>Learning Objectives</u></a> .....	2
<a href="#"><u>Graduate Attributes</u></a> .....	2
<a href="#"><u>Assessment Summary</u></a> .....	3
<a href="#"><u>Teaching Approach</u></a> .....	3
<a href="#"><u>Feedback</u></a> .....	3
<a href="#"><u>Our feedback to You</u></a> .....	3
<a href="#"><u>Your feedback to Us</u></a> .....	3
<a href="#"><u>Previous Student Evaluations of this unit</u></a> .....	3
<a href="#"><u>Required Resources</u></a> .....	3
<a href="#"><u>Unit Schedule</u></a> .....	4
<a href="#"><u>Assessment Policy</u></a> .....	4
<a href="#"><u>Assessment Tasks</u></a> .....	5
<a href="#"><u>Participation</u></a> .....	5
<a href="#"><u>Examinations</u></a> .....	6
<a href="#"><u>Examination 1</u></a> .....	6
<a href="#"><u>Assignment submission</u></a> .....	6
<a href="#"><u>Extensions and penalties</u></a> .....	6
<a href="#"><u>Returning assignments</u></a> .....	6
<a href="#"><u>Policies</u></a> .....	6
<a href="#"><u>Student services</u></a> .....	7

# **FIT3002 Applications of data mining - Semester 1, 2011**

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.

## **Mode of Delivery**

- Gippsland (Day)
- Gippsland (Off-campus)
- Sunway (Day)
- South Africa (Day)

## **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 each week.

## **Unit Relationships**

### **Prohibitions**

CSE3212, GCO3828

### **Prerequisites**

FIT1004 or FIT2010 or equivalent

## **Chief Examiner**

**Kai Ming Ting**

## **Campus Lecturer**

### **Gippsland**

**Kai Ming Ting**

### **South Africa**

**Sakkie van Zyl**

### **Sunway**

**Elsa Phung**

## **Learning Objectives**

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.

## **Graduate Attributes**

Monash prepares its graduates to be:

1. responsible and effective global citizens who:
  - a. engage in an internationalised world
  - b. exhibit cross-cultural competence
  - c. demonstrate ethical values

critical and creative scholars who:

- a. produce innovative solutions to problems
- b. apply research skills to a range of challenges
- c. communicate perceptively and effectively

## Assessment Summary

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

Assessment Task	Value	Due Date
Assignment 1	20%	6 April 2011
Assignment 2	20%	4 May 2011
Examination 1	60%	To be advised

## Teaching Approach

### Lecture and tutorials or problem classes

This teaching and learning approach provides facilitated learning, practical exploration and peer learning.

## Feedback

### Our feedback to You

Types of feedback you can expect to receive in this unit are:

- Informal feedback on progress in labs/tutes
- Graded assignments with comments
- Other: Solutions to review questions and assignments

### Your feedback to Us

Monash is committed to excellence in education and regularly seeks feedback from students, employers and staff. One of the key formal ways students have to provide feedback is through SETU, Student Evaluation of Teacher and Unit. The University's student evaluation policy requires that every unit is evaluated each year. Students are strongly encouraged to complete the surveys. The feedback is anonymous and provides the Faculty with evidence of aspects that students are satisfied and areas for improvement.

For more information on Monash's educational strategy, and on student evaluations, see:

<http://www.monash.edu.au/about/monash-directions/directions.html>

<http://www.policy.monash.edu/policy-bank/academic/education/quality/student-evaluation-policy.html>

## Previous Student Evaluations of this unit

If you wish to view how previous students rated this unit, please go to

<https://emuapps.monash.edu.au/unitevaluations/index.jsp>

## Required Resources

1. Software Title: WEKA, version 3.6

2. Magnum OPUS version 4

Both are freeware from the websites stated in the relevant practical web pages.

## Unit Schedule

Week	Date*	Activities	Assessment
0	21/02/11		No formal assessment or activities are undertaken in week 0
1	28/02/11	The Need for Data Mining	Practical work and Review Questions
2	07/03/11	Model Building	Practical work and Review Questions
3	14/03/11	Model Representation	Practical work and Review Questions
4	21/03/11	Data Mining Process	Review Questions
5	28/03/11	Performance Evaluation	Review Questions
6	04/04/11	Engineering the input and output	Practical work and Review Questions; Assignment 1 due 6 April 2011
7	11/04/11	Algorithms	Practical work and Review Questions
8	18/04/11	Implementation Issues	Review Questions
Mid semester break			
9	02/05/11	Market basket analysis	Practical work and Review Questions; Assignment 2 due 4 May 2011
10	09/05/11	Cluster Analysis	Review Questions
11	16/05/11	Anomaly Detection	Review Questions
12	23/05/11	Case Studies and Data Mining Applications	Review Questions
	30/05/11	SWOT VAC	No formal assessment is undertaken SWOT VAC

\*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.

## 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

## Assessment Tasks

### Participation

Assignment tasks are required to be completed by students in pairs.

#### • Assessment 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%

**Criteria for assessment:**

To get a Pass grade, students must perform data preparation/preprocessing, produce several different models and choose the best model, and submit a clearly written report describing the process.

To get a better grade, students must show that they have performed extra data analysis and preprocessing, explored a wide range of different models and describe how the final model is produced and how it can be applied for future predictions.

More detailed criteria will be provided in the sample marksheet on the assignment web page.

**Due date:**

6 April 2011

#### • Assessment 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 be likely to produce the largest profit within the budgetary constraint for a mass mailing campaign. Students are required to write a report to describe the process and analysis involved.

**Weighting:**

20%

**Criteria for assessment:**

- ◆ Must have a clear problem definition section that defines the inputs (and their types: nominal or numeric) and output; evaluation method and performance measure used (train and test using the given data sets and choose model based on profit).
- ◆ Produce several different models.
- ◆ Choose the best model which maximises profit in all parts of the process.
- ◆ A clearly written report which shows the high level process taken.

More detailed criteria will be provided in the sample marksheet on the assignment web page.

**Due date:**

4 May 2011

## Examinations

### • Examination 1

**Weighting:**

60%

**Length:**

3 hours

**Type (open/closed book):**

Closed book

**Electronic devices allowed in the exam:**

None

## Assignment submission

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.

## Extensions and penalties

Submission must be made by the due date otherwise penalties will be enforced.

You must negotiate any extensions formally with your campus unit leader via the in-semester special consideration process:

<http://www.infotech.monash.edu.au/resources/student/equity/special-consideration.html>.

## Returning assignments

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

## Policies

Monash has educational policies, procedures and guidelines, which are designed to ensure that staff and students are aware of the University's academic standards, and to provide advice on how they might uphold them. You can find Monash's Education Policies at:

<http://policy.monash.edu.au/policy-bank/academic/education/index.html>

Key educational policies include:

- Plagiarism  
(<http://www.policy.monash.edu/policy-bank/academic/education/conduct/plagiarism-policy.html>)
- Assessment  
(<http://www.policy.monash.edu/policy-bank/academic/education/assessment/assessment-in-coursework-p>)
- Special Consideration  
(<http://www.policy.monash.edu/policy-bank/academic/education/assessment/special-consideration-policy.h>)
- Grading Scale  
(<http://www.policy.monash.edu/policy-bank/academic/education/assessment/grading-scale-policy.html>)



- Discipline: Student Policy (<http://www.policy.monash.edu/policy-bank/academic/education/conduct/student-discipline-policy.html>)
- Academic Calendar and Semesters (<http://www.monash.edu.au/students/key-dates/>);
- Orientation and Transition (<http://www.infotech.monash.edu.au/resources/student/orientation/>); and
- Academic and Administrative Complaints and Grievances Policy (<http://www.policy.monash.edu/policy-bank/academic/education/management/complaints-grievance-policy>)

## Student services

The University provides many different kinds of support services for you. Contact your tutor if you need advice and see the range of services available at [www.monash.edu.au/students](http://www.monash.edu.au/students). The Monash University Library provides a range of services and resources that enable you to save time and be more effective in your learning and research. Go to <http://www.lib.monash.edu.au> or the library tab in my.monash portal for more information. Students who have a disability or medical condition are welcome to contact the Disability Liaison Unit to discuss academic support services. Disability Liaison Officers (DLOs) visit all Victorian campuses on a regular basis

- Website: <http://adm.monash.edu/sss/equity-diversity/disability-liaison/index.html>;
- Telephone: 03 9905 5704 to book an appointment with a DLO;
- Email: [dlu@monash.edu](mailto:dlu@monash.edu)
- Drop In: Equity and Diversity Centre, Level 1 Gallery Building (Building 55), Monash University, Clayton Campus.

## Reading List

### **Textbook:**

Witten, I.H. & Frank, E. Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann Publishers, Second Edition, 2005.

### **References:**

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.