



MONASH University

FIT5045
Knowledge discovery and data mining

Unit Guide

Semester 2, 2009

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 : 16 Jul 2009

Table of Contents

<u>FIT5045 Knowledge discovery and data mining - Semester 2, 2009</u>	1
<u>Chief Examiner</u>	1
<u>Lecturer(s) / Leader(s)</u>	1
<u>Caulfield</u>	1
<u>Additional communication information</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>	2
<u>Prerequisites</u>	2
<u>Prohibitions</u>	3
<u>Relationships</u>	3
<u>Teaching and learning method</u>	4
<u>Timetable information</u>	4
<u>Tutorial allocation</u>	4
<u>Unit Schedule</u>	4
<u>Unit Resources</u>	5
<u>Prescribed text(s) and readings</u>	5
<u>Recommended text(s) and readings</u>	5
<u>Required software and/or hardware</u>	5
<u>Equipment and consumables required or provided</u>	5
<u>Study resources</u>	5
<u>Assessment</u>	6
<u>Overview</u>	6
<u>Faculty assessment policy</u>	6
<u>Assignment tasks</u>	6
<u>Examinations</u>	6
<u>Due dates and extensions</u>	7
<u>Late assignment</u>	7
<u>Return dates</u>	7
<u>Appendix</u>	8

FIT5045 Knowledge discovery and data mining - Semester 2, 2009

Chief Examiner:

Dr Grace Rumantir

Fax: +61 3 8622 8999

Contact hours: Wednesday 2-4pm

Lecturer(s) / Leader(s):

Caulfield

Dr Grace Rumantir

Fax: +61 3 8622 8999

Contact hours: Wednesday 2-4pm

Additional communication information:

Tutor: Minh Le Viet

Introduction

Welcome to FIT5045 Knowledge Discovery and Data Mining for Semester 2, 2009. This 6 point unit is an elective unit to all of the masters by coursework programs in the Faculty of IT. The unit has been designed to provide you with the fundamental principles of data mining and how it can be used to extract hidden patterns from data. It explores various data mining methods and its practical applications using a data mining tool.

Unit synopsis

Modern methods of discovering patterns in large-scale databases are introduced, including classification, clustering and association rules analysis. These are contrasted with more traditional methods of finding information from data, such as data queries. Data pre-processing methods for dealing with noisy and missing data and with dimensionality reduction are reviewed. Hands-on case studies in building data mining models are performed using a popular software package.

Learning outcomes

At the completion of this unit students will:

1. Be able to differentiate between supervised and unsupervised learning;
2. Know how to apply the main techniques for supervised and unsupervised learning;
3. Know how to use statistical methods for evaluating data mining models;
4. Be able to perform data pre-processing for data with outliers, incomplete and noisy data;
5. Be able to extract and analyse patterns from data using a data mining tool.

Contact hours

4 hrs/week

Workload

Students are expected to commit to:

- 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.

Unit relationships

Prerequisites

For MAIT students, [FIT9017](#), [FIT9018](#), [FIT9019](#), [FIT9030](#), [FIT9020](#) and [FIT4037](#).

- + Sound fundamental knowledge in high school maths and statistics
- + Basic database and computer programming knowledge

Prohibitions

CSE5230, FIT5024

Relationships

FIT5045 is a elective unit in the masters by coursework degree programs. Before attempting this unit you must have satisfactorily completed

For MAIT students, FIT9017, FIT9018, FIT9019, FIT9030, FIT9020 and FIT4037, or equivalent.

You should have knowledge of

Sound fundamental knowledge in maths and statistics Basic database and computer programming knowledge . You may not study this unit and CSE5230, FIT5024 in your degree.

Teaching and learning method

Students are expected to attempt a basic maths and statistics quiz on Moodle at the beginning of the semester to determine if the student has sufficient fundamental knowledge in maths and statistics to take this unit.

Apart from the weekly lecture and tutorials, students are expected to attempt weekly quizzes on Moodle covering the topic presented in the previous week. There will be 10 weekly quizzes that will carry a total 3% bonus mark for the unit.

Each quiz will only be available in one week and feedback will be automatically sent to students who have attempted the quiz once the due date has passed. Moodle has no facility to enable the lecturer to extract closed quizzes. Therefore it is not possible for students who do not attempt a quiz before the closing date to get access to the questions.

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.cc.monash.edu.au/>

Unit Schedule

Week	Topic	Key dates
1	Unit Administration and Introduction to Data Mining	
2	Model Building	
3	Model Evaluation	
4	Data Preprocessing (1)	
5	Data Preprocessing (2)	
6	Classification	
7	Unit Test	In the lecture
8	Clustering	
9	Association Rules Mining (1)	
10	Association Rules Mining (2)	Hurdle: Stage 1 of Assignment due at the start of the lecture
Mid semester break		
11	Mining Stream, Time Series and Sequence Data	
12	Data Mining Visualization	Assignment due at the start of the lecture
13	Revision	

Unit Resources

Prescribed text(s) and readings

There is no one prescribed textbook for this unit. Students are expected to access the relevant chapters of the books on the recommended reading lists. Two of the books are available as online e-books in the library.

Text books are available from the Monash library and 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

Online e-books in the library:

J. Han & M. Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann, 2006
I.H. Witten & E. Frank, Data Mining: Practical Machine Learning Tools and Techniques, 2nd edition, Morgan Kaufmann, 2005

Other books:

R. Roiger and M. Geatz, Data Mining A Tutorial-based Primer, Pearson Education, Inc., 2003
P. Tan, M. Steinbach, V. Kumar, Introduction to Data Mining, Pearson Education, Inc., 2006
G. Gupta, Introduction to Data Mining and Case Studies, Prentice-Hall, New Delhi, 2006
A.B.M. Shawakat Ali and S. A. Wasimi, Data Mining: Methods and Techniques, Thomson Learning, 2007

Required software and/or hardware

You will need to download the data mining tool WEKA version 3.4 (book version) from <http://www.cs.waikato.ac.nz/ml/weka/>

You will need to have Java <http://www.java.com/> installed to run WEKA on your computer.

Equipment and consumables required or provided

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 **5** 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:

Study resources we will provide for your study are available on Moodle.

Assessment

Overview

Assignments: 100%

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 44% 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:

Application of Data Mining

Description:

Students are to use WEKA to perform various data mining methods on a particular data set.

Weighting:

20%

Due date:

Week 12 at the beginning of the lecture

Examinations

• Examination 1

Weighting: 20%

Length: 1.5 hours

Type (open/closed book): Closed book

Remarks:

To be done in the form of a unit test in Week 7 lecture time slot.

• **Examination 2**

Weighting: 60%

Length: 3 hours

Type (open/closed book): Closed book

Remarks:

to be conducted in the formal examination period

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% per day, including weekends.

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