MONASH University

**FIT3073
Data mining**

Unit guide

Semester 1, 2009

Table of Contents

FIT3073 Data mining - Semester 1, 2009

Unit leader :

Grace Rumantir

Lecturer(s) :

Caulfield

- Grace Rumantir

Tutors(s) :

Caulfield

- Minh Le Viet

Introduction

Welcome to FIT3073 Data Mining for semester 1, 2009.

Unit synopsis

ASCED Discipline Group classification: 020303 Database management

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

Knowledge and Understanding

To 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. For students to become familiar with the currently available techniques for the extraction of information from large databases.

At the completion of study the students will:

- 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.

Attitudes, Values and Beliefs

- 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.

Practical Skills

- 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.

Workload

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

Unit relationships

Prerequisites

Before attempting this unit you must have satisfactorily completed

FIT1004 or CSE2132 or equivalent

Relationships

FIT3073 is an elective unit in the system development major in BITS degree.

Before attempting this unit you must have satisfactorily completed FIT1004 or CSE2132 or equivalent

You may not study this unit and CSE3212 in your degree.

Continuous improvement

Monash is committed to 'Excellence in education' (Monash Directions 2025 - http://www.monash.edu.au/about/monash-directions/directions.html) and strives for the highest possible quality in teaching and learning.

To monitor how successful we are in providing quality teaching and learning Monash regularly seeks feedback from students, employers and staff. One of the key formal ways students have to provide feedback is through Unit Evaluation Surveys. The University's Unit Evaluation policy (http://www.policy.monash.edu/policy-bank/academic/education/quality/unit-evaluation-policy.html) requires that every unit offered is evaluated each year. Students are strongly encouraged to complete the surveys as they are an important avenue for students to "have their say". The feedback is anonymous and provides the Faculty with evidence of aspects that students are satisfied and areas for improvement.

Faculties have the option of administering the Unit Evaluation survey online through the my.monash portal or in class. Lecturers will inform students of the method being used for this unit towards the end of the semester.

Student Evaluations

If you wish to view how previous students rated this unit, please go to
http://www.adm.monash.edu.au/cheq/evaluations/unit-evaluations/

Improvements to this unit

A MonQuest evaluation will be conducted in Week 11.

Unit staff - contact details

Unit leader

Dr Grace Rumantir
Fax +61 3 8622 8999
Contact hours : Thursday and Friday 11am-12pm and 4-5pm

Lecturer(s) :

Dr Grace Rumantir
Fax +61 3 8622 8999
Contact hours : Thursday and Friday 11am-12pm and 4-5pm

Tutor(s) :

Minh Le Viet

Teaching and learning method

- 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.

Tutorial allocation

On-campus students should register for tutorials/laboratories using Allocate+.

Communication, participation and feedback

Monash aims to provide a learning environment in which students receive a range of ongoing feedback throughout their studies. You will receive feedback on your work and progress in this unit. This may take the form of group feedback, individual feedback, peer feedback, self-comparison, verbal and written feedback, discussions (on line and in class) as well as more formal feedback related to assignment marks and grades. You are encouraged to draw on a variety of feedback to enhance your learning.

It is essential that you take action immediately if you realise that you have a problem that is affecting your study. Semesters are short, so we can help you best if you let us know as soon as problems arise. Regardless of whether the problem is related directly to your progress in the unit, if it is likely to interfere with your progress you should discuss it with your lecturer or a Community Service counsellor as soon as possible.

Unit Schedule

Week	Topic	Key dates
1	Introduction to Data Mining	
2	Models in Data Mining	
3	Model Representation and Evaluation	
4	Data Preparation	
5	Data Mining Process	
6	Classification Algorithm	
Mid semester break		
7	Unit Test	Unit Test (20%)
8	Association Rules I	
9	Association Rules II	
10	Clustering I	
11	Clustering II	Assignment Due (20%)
12	Mining Data Stream	
13	Revision	

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 PTR 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 Kennedy R.L., Lee Y., Van Roy B., Reed C.D. & Lippman R.P. (1997) Solving Data Mining Problems Through Pattern Recognition, Prentice Hall PTR Witten I.H. & Frank E. (2000) Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufmann

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.

Library access

The Monash University Library site contains details about borrowing rights and catalogue searching. To learn more about the library and the various resources available, please go to http://www.lib.monash.edu.au.

The Educational Library and Media Resources (LMR) is also a very resourceful place to visit at http://www.education.monash.edu.au/library/

Monash University Studies Online (MUSO)

All unit and lecture materials are available through MUSO (Monash University Studies Online). Blackboard is the primary application used to deliver your unit resources. Some units will be piloted in Moodle. If your unit is piloted in Moodle, you will see a link from your Blackboard unit to Moodle (http://moodle.monash.edu.au) and can bookmark this link to access directly. In Moodle, from the Faculty of Information Technology category, click on the link for your unit.

You can access MUSO and Blackboard via the portal: http://my.monash.edu.au

Click on the Study and enrolment tab, then Blackboard under the MUSO learning systems.

In order for your Blackboard unit(s) to function correctly, your computer needs to be correctly configured.

For example:

- Blackboard supported browser
- Supported Java runtime environment

For more information, please visit: http://www.monash.edu.au/muso/support/students/downloadables-student.html

You can contact the MUSO Support by phone : (+61 3) 9903 1268

For further contact information including operational hours, please visit: http://www.monash.edu.au/muso/support/students/contact.html

Further information can be obtained from the MUSO support site: http://www.monash.edu.au/muso/support/index.html

Assessment

Unit assessment policy

To pass this unit, a student must obtain :

- 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

Ifa student does not achieve 40% or more in the unit examination or theunit non-examination assessment then a mark of no greater than 44-Nwill be recorded for the unit.

Assignment tasks

- ## Assignment Task

  Title : Unit Test

  Description :

  Mid-semester test.
  Weighting : 20%

  Criteria for assessment :

  Due date : Week 7 lecture (23 April 2009)
- ## Assignment Task

  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%

  Criteria for assessment :

  Assessment will be marked based on:

  - ♦ the evidence of understanding of data mining process.
  - ♦ the evidence of understanding of preparing data for the data mining process
  - ♦ the evidence of understanding in creating models for data mining.
  - ♦ the evidence of understanding in evaluating the performance of the models proposed.

  Due date : Sunday, 24 May 2009 (Week 11), at 11.55pm

Examinations

- **Examination 1**

  Weighting : 60%

  Length : 3 hours

  Type (open/closed book) : Closed book

Assignment submission

Electronic assignment submissions are to be done through Moodle on or before the due date.

Assignment coversheets

The assignment coversheet will be available in MUSO.

University and Faculty policy on assessment

Due dates and extensions

The due dates for the submission of assignments are given in the previous section. 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 seldom regarded as appropriate reasons for granting extensions. Students are advised to NOT assume that granting of an extension is a matter of course.

Requests for extensions must be made to the unit lecturer at yourcampus at least two days before the due date. You will be asked to forward original medical certificates in cases of illness, and may beasked to provide other forms of documentation where necessary. A copyof the email or other written communication of an extension must beattached to the assignment submission.

There is no extension for unit test.

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.

Assessment for the unit as a whole is in accordance with the provisions of the Monash University Education Policy at http://www.policy.monash.edu/policy-bank/academic/education/assessment/

We will aim to have assignment results made available to you within two weeks after assignment receipt.

Plagiarism, cheating and collusion

Plagiarism and cheating are regarded as very serious offences. In cases where cheating has been confirmed, students have been severely penalised, from losing all marks for an assignment, to facing disciplinary action at the Faculty level. While we would wish that all our students adhere to sound ethical conduct and honesty, I will ask you to acquaint yourself with Student Rights and Responsibilities (http://www.infotech.monash.edu.au/about/committees-groups/facboard/policies/studrights.html) and the Faculty regulations that apply to students detected cheating as these will be applied in all detected cases.

In this University, cheating means seeking to obtain an unfair advantage in any examination or any other written or practical work to be submitted or completed by a student for assessment. It includes the use, or attempted use, of any means to gain an unfair advantage for any assessable work in the unit, where the means is contrary to the instructions for such work.

When you submit an individual assessment item, such as a program, a report, an essay, assignment or other piece of work, under your name you are understood to be stating that this is your own work. If a submission is identical with, or similar to, someone else's work, an assumption of cheating may arise. If you are planning on working with another student, it is acceptable to undertake research together, and discuss problems, but it is not acceptable to jointly develop or share solutions unless this is specified by your lecturer.

Intentionally providing students with your solutions to assignments is classified as "assisting to cheat" and students who do this may be subject to disciplinary action. You should take reasonable care that your solution is not accidentally or deliberately obtained by other students. For example, do not leave copies of your work in progress on the hard drives of shared computers, and do not show your work to other students. If you believe this may have happened, please be sure to contact your lecturer as soon as possible.

Cheating also includes taking into an examination any material contrary to the regulations, including any bilingual dictionary, whether or not with the intention of using it to obtain an advantage.

Plagiarism involves the false representation of another person's ideas, or findings, as your own by either copying material or paraphrasing without citing sources. It is both professional and ethical to reference clearly the ideas and information that you have used from another writer. If the source is not identified, then you have plagiarised work of the other author. Plagiarism is a form of dishonesty that is insulting to the reader and grossly unfair to your student colleagues.

Register of counselling about plagiarism

The university requires faculties to keep a simple and confidential register to record counselling to students about plagiarism (e.g. warnings). The register is accessible to Associate Deans Teaching (or nominees) and, where requested, students concerned have access to their own details in the register. The register is to serve as a record of counselling about the nature of plagiarism, not as a record of allegations; and no provision of appeals in relation to the register is necessary or applicable.

Non-discriminatory language

The Faculty of Information Technology is committed to the use of non-discriminatory language in all forms of communication. Discriminatory language is that which refers in abusive terms to gender, race, age, sexual orientation, citizenship or nationality, ethnic or language background, physical or mental ability, or political or religious views, or which stereotypes groups in an adverse manner. This is not meant to preclude or inhibit legitimate academic debate on any issue; however, the language used in such debate should be non-discriminatory and sensitive to these matters. It is important to avoid the use of discriminatory language in your communications and written work. The most common form of discriminatory language in academic work tends to be in the area of gender inclusiveness. You are, therefore, requested to check for this and to ensure your work and communications

are non-discriminatory in all respects.

Students with disabilities

Students with disabilities that may disadvantage them in assessment should seek advice from one of the following before completing assessment tasks and examinations:

- Faculty of Information Technology Student Service staff, and / or
- your Unit Coordinator, or
- Disabilities Liaison Unit

Deferred assessment and special consideration

Deferred assessment (not to be confused with an extension for submission of an assignment) may be granted in cases of extenuating personal circumstances such as serious personal illness or bereavement. Information and forms for Special Consideration and deferred assessment applications are available at http://www.monash.edu.au/exams/special-consideration.html. Contact the Faculty's Student Services staff at your campus for further information and advice.