FIT5024
Applications of data mining

Unit guide

Semester 1, 2009

Last updated : 10 Feb 2009

Table of Contents

FIT5024 Applications of data mining - Semester 1, 2009

Unit leader :

Kai Ming Ting

Lecturer(s) :

Gippsland

- Kai Ming Ting

Introduction

Welcome to FIT5024 Applications of Data Mining. This 6 point unit is elective to all postgraduate 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 generalizations 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:

Knowledge 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

Skills in using data mining tools to solve data mining problems

Understanding of current major research issues

Workload

Unit relationships

Prerequisites

Before attempting this unit you must have satisfactorily gained entry to: MIT, MBS and MAIT , or equivalent. Knowledge of Database is helpful but not essential.

Relationships

You may not study this unit and COT 5230 Data Mining or translation set GCO5828 Applications of Data Mining 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.monash.edu.au/unit-evaluation-reports/

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 staff - contact details

Unit leader

Associate Professor Kai Ming Ting
Director of Undergraduate Studies
Phone +61 3 990 26241

Lecturer(s) :

Associate Professor Kai Ming Ting
Director of Undergraduate Studies
Phone +61 3 990 26241

Teaching and learning method

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	The Need for Data Mining	
2	Model Building	
3	Model Representation	
4	Data Mining Process	
5	Performance Evaluation	
6	Engineering the input and output	
Mid semester break		
7	Algorithms	
8	Implementation Issues	
9	Market basket analysis	
10	Cluster Analysis & Anomaly Detection	
11	Case Study	
12	Data Mining Applications & Research Issues	
13	Study Week	

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.

Recommended text(s) and readings

Kennedy, R.L., Lee, Y. Roy, B.V., Reed, C.D. & Lippman, R.P., Solving Data Mining Problems through Pattern Recognition, Prentice Hall, 1998.

Cabena, P., Hadjinian, P., Stadler, R., Verhees, J. & Zanasi, A., Discovering Data Mining: from concept to implementation, Prentice Hall, 1997. Berry, J.A.M. & Linoff, G. Data Mining Techniques for Marketing, Sales, and Customer Support, John Wiley & Sons, 1997. Tan, P-N, Steinbach, M. & Kumar, V. Introduction to Data Mining, Addison Wesley, 2006. Han, J. & Kamber, M. Data Mining: Concepts and Techniques, Morgan Kaufmann, Second Edition, 2006.Dunham, M.H., Data Mining: Introductory and Advance Topics, Pearson Education, 2003. Groth, R., Data Mining: Building competitive advantage, Prentice Hall, 2000. Berson,. A., Smith, S. & Thearling, K., Building Data Mining Applications for CRM, McGraw Hill. 2000. Berry, J.A.M. & Linoff, G. Mastering Data Mining: The Art and Science of Customer Relationship Management, John Wiley & Sons, 2000. Mena, J. Data Mining Your Website. Digital Press, 1999. Westphal, C. & Blaxton, T. Data Mining Solutions, John Wiley & Sons, 1998.

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.

Software may be:

- downloaded from On Unit Web site
- purchased at academic price at good software retailers

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:

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

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

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

The unit is assessed with three 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 Task

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 : 10%

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.
Due date : 8 April 2009

- ## Assignment Task

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 : 10%

Criteria for assessment :

Students are required to write a report that describes the process which should include data preparation and preprocessing, the models employed, the model selection process and the suggestion as to how the campaign should be done based on the chosen model.
Due date : 6 May 2009

- ## Assignment Task

Title : Assignment 3

Description :

Each student can choose either

(**A**) To conduct an in-depth study of a specific method for data mining, or

(**B**) To conduct a review of a current research topic.

Weighting : 20%

Criteria for assessment :

Students are required to write a report on the chosen topic which should contain:

- An introduction
- A detail description of the chosen topic
- A comparison of different methods (if applicable)
- A reference section

Due date : 20 May 2009

Examinations

- ## Examination 1

Weighting : 60%

Length : 3 hours

Type (open/closed book) : Closed book

Assignment submission

Assignments will be submitted electronically via WebFace (http://wfsubmit.its.monash.edu.au/)

Assignment coversheets

Students must obtain a coversheet from WebFace (http://wfsubmit.its.monash.edu.au/) before submitting their assignment electronically.

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.

Late assignment

Assignments received after the due date will be subject to a penalty of 10% 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.

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.