MONASH University
Information Technology

FIT3002
Applications of data mining

Unit Guide

Semester 1, 2014

Last updated: 21 Feb 2014

Table of Contents

FIT3002 Applications of data mining - Semester 1, 2014

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)
- Malaysia (Day)
- South Africa (Day)

Workload Requirements

Minimum total expected workload equals 12 hours per week comprising:

(a.) Contact hours for on-campus students:

- One 2-hour workshop
- One 2-hour laboratory (for 6 weeks)

(b.) Study schedule for off-campus students:

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

(c.) Additional requirements (all students):

- A minimum of 8 hours independent study per week for completing lab and project work, private study and revision.

Unit Relationships

Prohibitions

CSE3212, GCO3828

Prerequisites

FIT1004 or FIT2010 or equivalent

Chief Examiner

Dr Grace Rumantir

Campus Lecturer

Gippsland

Kai Ming Ting

South Africa

Neil Manson

Malaysia

Jojo Wang

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 the Student Evaluation of Teaching and Units (SETU) survey. 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, see:

www.monash.edu.au/about/monash-directions/ and on student evaluations, see:
www.policy.monash.edu/policy-bank/academic/education/quality/student-evaluation-policy.html

Previous Student Evaluations of this Unit

Based on previous student feedback topics on cluster analysis and anomaly detection, and additional reading on their application, were added in to broaden students' knowledge in these areas.

If you wish to view how previous students rated this unit, please go to
https://emuapps.monash.edu.au/unitevaluations/index.jsp

Academic Overview

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.

Unit Schedule

Week	Activities	Assessment
0		No formal assessment or activities are undertaken in week 0
1	The Need for Data Mining	Practical work and Review Questions
2	Model Building	Practical work and Review Questions
3	Model Representation	Practical work and Review Questions
4	Data Mining Process	Review Questions
5	Performance Evaluation	Review Questions
6	Engineering the input and output	Practical work and Review Questions; Assignment 1 due 9 April 2014
7	Algorithms	Practical work and Review Questions
8	Implementation Issues	Practical work and Review Questions
9	Market basket analysis	Practical work and Review Questions; Assignment 2 due 7 May 2014
10	Cluster Analysis	Review Questions
11	Anomaly Detection	Review Questions
12	Case Studies and Data Mining Applications	Review Questions
	SWOT VAC	No formal assessment is undertaken in SWOT VAC
	Examination period	LINK to Assessment Policy: http://policy.monash.edu.au/policy-bank/academic/education/assessment/assessment-in-coursework-policy.html

*Unit Schedule details will be maintained and communicated to you via your learning system.

Teaching Approach

- **Workshops**
  This teaching and learning approach provides facilitated learning, exploration and peer learning.
- **Laboratory-based classes**
  This approach provides practical exploration and peer learning.

Assessment Summary

Examination (3 hours): 50%; In-semester assessment: 50%

Assessment Task	Value	Due Date
Assignment 1	15%	9 April 2014
Assignment 2	15%	7 May 2014
Weekly participation assessment	20%	Weekly

Unit Schedule

Examination 1	50%	To be advised

Assessment Requirements

Assessment Policy

Faculty Policy - Unit Assessment Hurdles
(http://intranet.monash.edu.au/infotech/resources/staff/edgov/policies/assessment-examinations/assessment-hurd

Academic Integrity - Please see resources and tutorials at
http://www.monash.edu/library/skills/resources/tutorials/academic-integrity/

Assessment Tasks

Participation

Students are assessed on their participations in workshops and online activities in Moodle throughout the semester. For on-campus students, this includes presentations and weekly participation in the discussion in the workshops. Off-campus students are required to do similar activities online in Moodle.

This assessment will be made weekly and it is worth 20% of the total assessment.

Two assignment tasks are required to be completed by students in pairs. These assignments are worth 30% of the total assessment.

- **Assessment task 1**

  Title:
  Assignment 1
  Description:
  This assignment requires students to use the data mining tool WEKA to build a good performing model from a given set of data, and write a report describing the data mining process.
  Weighting:
  15%
  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.

  Members of a group are expected to contribute equally to the group work. Each member is to submit a peer review form independently to provide their assessment of every member's contribution. Normally the marks awarded to group members will be the same, however if the peer review indicates an uneven contribution, the marker will investigate and if necessary apportion marks accordingly.

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

9 April 2014

- **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:

15%

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.

Members of a group are expected to contributed equally to the group work. Each member is to submit a peer review form independently to provide their assessment of every member's contribution. Normally the marks awarded to group members will be the same, however if the peer review indicates an uneven contribution, the marker will investigate and if necessary apportion marks accordingly.

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

Due date:

7 May 2014

- **Assessment task 3**

Title:

Weekly participation assessment

Description:

Students are required to participate in workshops and online activities in Moodle throughout the semester. For on-campus students this includes presentations and weekly participation in the discussions in the workshops. Off-campus students are required to do similar activities online.

Weighting:

20%

Criteria for assessment:

This is an individual assessment and assessed on participation rather than correctness.

Due date:

Weekly

Examinations

- **Examination 1**

Weighting:
50%
Length:
3 hours
Type (open/closed book):
Closed book
Electronic devices allowed in the exam:
None

Learning resources

Reading list

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 & Pei, J. M. Data Mining: Concepts and Techniques, Morgan Kaufmann, Third Edition, 2011.

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.

14. Vasant Dhar. Data Science and Prediction, Communications of the ACM, Vol. 56, No. 12, December 2013, pages 64-73.

Monash Library Unit Reading List (if applicable to the unit)
http://readinglists.lib.monash.edu/index.html

Faculty of Information Technology Style Guide

Feedback to you

Examination/other end-of-semester assessment feedback may take the form of feedback classes, provision of sample answers or other group feedback after official results have been published. Please check with your lecturer on the feedback provided and take advantage of this prior to requesting individual consultations with staff. If your unit has an examination, you may request to view your examination script booklet, see
http://intranet.monash.edu.au/infotech/resources/students/procedures/request-to-view-exam-scripts.html

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

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.monash.edu.au/exams/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.

Assignment submission

It is a University requirement
(http://www.policy.monash.edu/policy-bank/academic/education/conduct/student-academic-integrity-managing-pla
for students to submit an assignment coversheet for each assessment item. Faculty Assignment coversheets can be found at http://www.infotech.monash.edu.au/resources/student/forms/. Please check with your Lecturer on the submission method for your assignment coversheet (e.g. attach a file to the online assignment submission, hand-in a hard copy, or use an online quiz). Please note that it is your responsibility to retain copies of your assessments.

Online submission

If Electronic Submission has been approved for your unit, please submit your work via the learning site for this unit, which you can access via links in the my.monash portal.

Required Resources

Please check with your lecturer before purchasing any Required Resources. Limited copies of prescribed texts are available for you to borrow in the library, and prescribed software is available in student labs.

1. Software Title: WEKA, version 3.6
2. Magnum OPUS version 4.6

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

Prescribed text(s)

Limited copies of prescribed texts are available for you to borrow in the library.

Witten, I.H., Frank, E. & Hall, M.A.. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*. (3rd Edition) Morgan Kaufmann Publishers.

Other Information

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: www.policy.monash.edu.au/policy-bank/academic/education/index.html

Key educational policies include:

- Student Academic Integrity Policy and Student Academic Integrity: Managing Plagiarism and Collusion Procedures ; http://www.policy.monash.edu/policy-bank/academic/education/conduct/student-academic-integrity-policy.h
- Assessment in Coursework Programs; http://www.policy.monash.edu/policy-bank/academic/education/assessment/assessment-in-coursework-po
- Special Consideration; http://www.policy.monash.edu/policy-bank/academic/education/assessment/special-consideration-policy.ht
- 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/dates/
- Orientation and Transition; http://intranet.monash.edu.au/infotech/resources/students/orientation/
- Academic and Administrative Complaints and Grievances Policy; http://www.policy.monash.edu/policy-bank/academic/education/management/complaints-grievance-policy.h

Faculty resources and policies

Important student resources including Faculty policies are located at http://intranet.monash.edu.au/infotech/resources/students/

Graduate Attributes Policy

http://www.policy.monash.edu/policy-bank/academic/education/management/monash-graduate-attributes-policy.ht

Student Charter

www.opq.monash.edu.au/ep/student-charter/monash-university-student-charter.html

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 http://www.monash.edu.au/students. For Malaysia see http://www.monash.edu.my/Student-services, and for South Africa see http://www.monash.ac.za/current/.

Monash University Library

The Monash University Library provides a range of services, resources and programs that enable you to save time and be more effective in your learning and research. Go to www.lib.monash.edu.au or the library tab in my.monash portal for more information. At Malaysia, visit the Library and Learning Commons at http://www.lib.monash.edu.my/. At South Africa visit http://www.lib.monash.ac.za/.

Disability Liaison Unit

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://www.monash.edu/equity-diversity/disability/index.html
- Telephone: 03 9905 5704 to book an appointment with a DLO; or contact the Student Advisor, Student Commuity Services at 03 55146018 at Malaysia
- Email: dlu@monash.edu
- Drop In: Equity and Diversity Centre, Level 1, Building 55, Clayton Campus, or Student Community Services Department, Level 2, Building 2, Monash University, Malaysia Campus