FIT5045
Knowledge discovery and data mining

Unit Guide

Semester 2, 2012

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: 06 Jul 2012
FIT5045 Knowledge discovery and data mining - Semester 2, 2012

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.

Mode of Delivery

- Caulfield (Day)
- Gippsland (Off-campus)

Contact Hours

2 hrs lectures/wk, 2 hrs laboratories/wk

Workload

Students will be expected to spend a total of 12 hours per week during semester on this unit as follows:

For on-campus students:

- 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

CSE5230, FIT5024

Prerequisites

Sound fundamental knowledge in maths and statistics. Basic database and computer programming knowledge.
Chief Examiner

Dr Grace Rumantir

Campus Lecturer

Caulfield

Grace Rumantir

Consultation hours: Thursday 2-4pm (in H7.08)

Gippsland

Kai Ming Ting

Tutors

Caulfield

Lauchlin Wilkinson

Consultation hours: TBA
Academic Overview

Outcomes

At the completion of this unit students will:

- be able to differentiate between supervised and unsupervised learning;
- know how to apply the main techniques for supervised and unsupervised learning;
- know how to use statistical methods for evaluating data mining models;
- be able to perform data pre-processing for data with outliers, incomplete and noisy data;
- be able to extract and analyse patterns from data using a data mining tool;
- have an understanding of the difference between discovery of hidden patterns and simple query extractions in a dataset;
- have an understanding of the different methods available to facilitate discovery of hidden patterns in a dataset;
- have developed the ability to preprocess data in preparation for data mining experiments;
- have developed the ability to evaluate the quality of data mining models;
- be able to appreciate the need to have representative sample input data to enable learning of patterns embedded in population data;
- be able to appreciate the need to provide quality input data to produce useful data mining models;
- have acquired the skill to use the common features in data mining tools;
- have acquired the skill to use the visualisation features in a data mining tools to facilitate knowledge discovery from a data set;
- have acquired the skill to compare data mining models based on the results on a set of performance criteria;
- be able to work in a team to extract knowledge from a common data set using different data mining methods and techniques.

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%

<table>
<thead>
<tr>
<th>Assessment Task</th>
<th>Value</th>
<th>Due Date</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20%</td>
<td></td>
</tr>
</tbody>
</table>
Academic Overview

For Caulfield on-campus students: Unit Test; For Gippsland off-campus students: Analysis of Case Studies

For all students: Group Assignment 20% For all students: 14 October 2012
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
- Test results and feedback
- Quiz results
- Solutions to tutes, labs 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.policy.monash.edu/policy-bank/academic/education/quality/student-evaluation-policy.html

Previous Student Evaluations of this unit

This unit was offered for the first time in Semester 2, 2009. The student reviews were good, but the unit will continually undergo improvements to ensure continual provision and delivery of up-to-date quality material.

Students will be requested to provide periodic informal anonymous feedback on the unit in Week 4 and Week 8. In Week 11 the Monquest and in Week 13 the Unit Evaluation evaluations will be conducted.

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

Students are to download the latest version of the free Data Mining Software WEKA from http://www.cs.waikato.ac.nz/ml/weka/ to work on their assignment and the tutorial exercises on their personal computers. WEKA is installed in the student labs used for the tutorials for this unit.
## Unit Schedule

<table>
<thead>
<tr>
<th>Week</th>
<th>Activities</th>
<th>Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td>No formal assessment or activities are undertaken in week 0</td>
</tr>
<tr>
<td>1</td>
<td>Unit Administration and Introduction to Data Mining</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Model Building</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Model Evaluation</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Data Preprocessing</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Data Preprocessing</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Classification</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Clustering</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>For Caulfield on-campus students: Unit Test in lecture; For all students: Anomaly Detection (non-examinable topic)</td>
<td>For Caulfield on-campus students: Unit Test in lecture 13 September 2012; For Gippsland off-campus students: Analysis of Case Studies due 16 September 2012</td>
</tr>
<tr>
<td>9</td>
<td>Association Rules Mining (1)</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Association Rules Mining (2)</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Web Mining</td>
<td>For all students: Group Assignment due 14 October 2012</td>
</tr>
<tr>
<td>12</td>
<td>Data Mining and Information Visualization</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SWOT VAC</td>
<td>No formal assessment is undertaken in SWOT VAC</td>
</tr>
</tbody>
</table>

*Unit Schedule details will be maintained and communicated to you via your MUSO (Blackboard or Moodle) learning system.*
Assessment Requirements

Assessment Policy

Faculty Policy - Unit Assessment Hurdles

Academic Integrity - Please see the Demystifying Citing and Referencing tutorial at http://lib.monash.edu/tutorials/citing/

Assessment Tasks

Participation

• Assessment task 1

  Title: For Caulfield on-campus students: Unit Test; For Gippsland off-campus students: Analysis of Case Studies

  Description: For Caulfield on-campus students: Closed-book unit test to be conducted in the lecture time slot in Week 8.

  For Gippsland off-campus students: Students are to answer questions in relation to case studies provided.

  Weighting: 20%

  Criteria for assessment: Correctness in answering the questions

  Due date: For Caulfield on-campus students: 13 September 2012 (in lecture); For Gippsland off-campus students: 16 September 2012

• Assessment task 2

  Title: For all students: Group Assignment

  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 best models for a given data set.

  Weighting: 20%

  Criteria for assessment: The assignment will be completed in groups of two students.

  Students will be assessed on:

  ◆ The degree to which the submission meet the assignment specification
  ◆ The quality of the data preprocessing and the design of experiments
  ◆ How well the experiments are conducted and summarised
  ◆ How well the results of the experiments are analysed and documented
Assessment Requirements

Further assessment criteria and marking sheet will be made available on the unit Moodle site.

Members in each group will receive the same marks. If there are issues/concerns about individual contributions within a group, a peer evaluation form will be used.

Due date:
For all students: 14 October 2012

Examinations

- Examination 1

  Weighting:
  60%

  Length:
  3 hours

  Type (open/closed book):
  Closed book

  Electronic devices allowed in the exam:
  None

Assignment submission

It is a University requirement (http://www.policy.monash.edu/policy-bank/academic/education/conduct/plagiarism-procedures.html) 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).

Online submission

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

Extensions and penalties

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


Returning assignments

Students can expect assignments to be returned within two weeks of the submission date or after receipt, whichever is later.
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: 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)
- Special Consideration (http://www.policy.monash.edu/policy-bank/academic/education/assessment/special-consideration-policy.html)
- 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/);
- and
- Codes of Practice for Teaching and Learning (http://www.policy.monash.edu/policy-bank/academic/education/conduct/suppdocs/code-of-practice-teaching-and-learning.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 www.monash.edu.au/students. For Sunway see http://www.monash.edu.my/Student-services, and for South Africa see http://www.monash.ac.za/current/.

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. At Sunway, visit the Library and Learning Commons at http://www.lib.monash.edu.my/. At South Africa visit http://www.lib.monash.ac.za/.

Academic support services may be available for students who have a disability or medical condition. Registration with the Disability Liaison Unit is required. Further information is available as follows:

- Website: http://monash.edu/equity-diversity/disability/index.html
- Email: dlu@monash.edu
- Drop In: Equity and Diversity Centre, Level 1 Gallery Building (Building 55), Monash University, Clayton Campus, or Student Community Services Department, Level 2, Building 2, Monash University, Sunway Campus
- Telephone: 03 9905 5704, or contact the Student Advisor, Student Community Services at 03 55146018 at Sunway