MONASH University
Information Technology

FIT3152
Data science

Unit Guide

Semester 2, 2015

Last updated: 23 Jul 2015

Table of Contents

FIT3152 Data science - Semester 2, 2015

In recent years the world has seen an explosion in the quantity and variety of data routinely recorded and analysed by research and industry, prompting some social commentators to refer to this phenomenon as the rise of "big data," and the analysts and practitioners who investigate the data as "data scientists."

The data may come from a variety of sources, including scientific experiments and measurements, or may be recorded from human interactions such as browsing data or social networks on the Internet, mobile phone usage or financial transactions. Many companies too, are realising the value of their data for analysing customer behaviour and preferences, recognising patterns of behaviour such as credit card usage or insurance claims to detect fraud, as well as more accurately evaluating risk and increasing profit.

In order to obtain insights from big data new analytical techniques are required by practitioners. These include computationally intensive and interactive approaches such as visualisation, clustering and data mining. The management and processing of large data sets requires the development of enhanced computational resources and new algorithms to work across distributed computers.

This unit will introduce students to the analysis and management of big data using current techniques and open source and proprietary software tools. Data and case studies will be drawn from diverse sources including health and informatics, life sciences, web traffic and social networking, business data including transactions, customer traffic, scientific research and experimental data. The general principles of analysis, investigation and reporting will be covered. Students will be encouraged to critically reflect on the data analysis process within their own domain of interest.

Mode of Delivery

- Clayton (Day)
- Clayton (Online)
- Malaysia (Day)

Workload Requirements

Minimum total expected workload equals 12 hours per week comprising:

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

- Two hours of lectures
- One 2-hour laboratory

(b.) Additional requirements (all students):

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

See also Unit timetable information

Unit Relationships

Prerequisites

FIT1006 or ETC1000 or equivalent. (For example BUS1100, ETC1010, ETC2010, ETF2211, ETW1000, ETW1010, ETW1102, ETW2111, ETX1100, ETX2111, ETX2121, MAT1097, STA1010)

Chief Examiner

Dr John Betts

Campus Lecturer

Clayton

Dr John Betts

Dr Sue Bedingfield

Mr Parthan Kasarapu

Malaysia

Dr Jojo Wong

Tutors

Clayton

Mr Rui Jie Chow (RJ)

Mr Dilpreet Singh

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

Past students have commented that learning R, RStudio and RapidMiner were highlights of the course, as was the guest lecture. These remain in this year's offering, with an increased emphasis on real world problems and applications of data science.

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

On successful completion of this unit, students should be able to:

- demonstrate the ability to transform real world problems into ones that can then be solved using data analytics techniques;
- cleanse and prepare data for analysis;
- analyse large data sets using a range of statistical, graphical and machine-learning techniques;
- validate and critically assess the results of analysis;
- interpret the results of analysis and communicate these to a broad audience;
- employ open source and proprietary software for data analytics;
- critically assess the appropriateness of analytical methods for a given task;
- identify opportunities for organisations to employ data analytics to understand current practice and identify potential opportunities;
- critically evaluate the limitations and benefits of data analytics.

Unit Schedule

Week	Activities	Assessment
0		No formal assessment or activities are undertaken in week 0
1	Introduction to Data Science. Introduction to R and RStudio. Review of basic statistics using R	Tutorial Participation assessed Weekly
2	Exploring data using graphics in R	
3	Data manipulation in R	
4	Linear regression in R	
5	Network analysis	Group Assignment (Initial report) due 28 August 2015
6	Guest Lecture	
7	Classification using decision trees	
8	Comparing classification models, evaluating algorithms	
9	K-Means and hierarchical clustering	
10	Text analysis	Individual Assignment due 9 October 2015
11	Student Presentations	Group Assignment (Presentation) due Week 11 lecture and (Final report) due 16 October 2015
12	Review of the course and exam preparation	
	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

Lecture and tutorials or problem classes

This teaching and learning approach helps students to initially encounter information at lectures, discuss and explore the information during tutorials, and practice in a hands-on lab environment.

Assessment Summary

Examination (2 hours): 60%; In-semester assessment: 40%

Assessment Task	Value	Due Date
Group Assignment	20%	Initial report due 28 August 2015. Presentation due Week 11 lecture. Final report due 16 October 2015

Unit Schedule

Individual Assignment	10%	9 October 2014
Tutorial Participation	10%	Weekly
Examination 1	60%	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

- **Assessment task 1**

  Title:
  Group Assignment
  Description:
  Students will work in groups to analyse a large data set and report their findings, and give a brief presentation of their project results during the Week 11 lecture.
  Weighting:
  20%
  Criteria for assessment:

  - Understanding of the real-world problem, and how the data might be used to solve the problem.
  - Cleansing and pre-processing the data.
  - Visual representation of the data, and initial insights into the data. (Initial report at this milestone)
  - Accuracy and reliability of the model.
  - Reporting and communication of results.

  As this is a group project, students in each group will allocate a weighting of the final results to each member of the group based on a consensus estimate of each member's contribution.
  Due date:
  Initial report due 28 August 2015. Presentation due Week 11 lecture. Final report due 16 October 2015
  Remarks:
  This assessment relates to Learning Outcomes 1, 2, 3, 4 and 5.

- **Assessment task 2**

  Title:
  Individual Assignment
  Description:
  Students will individually analyse a data set and report their findings.
  Weighting:
  10%
  Criteria for assessment:

- Understanding of the problem, and how the data might be used to solve the problem.
- Cleansing and pre-processing the data.
- Visual representation of the data, and initial insights into the data.
- Accuracy and reliability of the model.
- Reporting and communication of results.

Due date:
9 October 2014
Remarks:
This assessment relates to Learning Outcomes 1, 2, 3, 4 and 5.

- **Assessment task 3**

Title:
Tutorial Participation
Description:
Students will be assessed on their participation during tutorials.
Weighting:
10%
Criteria for assessment:
On Campus Students

- Participation in tutorials
- Completion of class exercises
- Contribution to class discussions

Off Campus Students

- Particpation in online forums
- Contributing to a Wiki

Due date:
Weekly
Remarks:
This assessment relates to Learning Outcomes 1, 2, 3, 4 and 5.

Examinations

- **Examination 1**

Weighting:
60%
Length:
2 hours
Type (open/closed book):
Closed book
Electronic devices allowed in the exam:
Electronic calculators permitted in the exam.
Remarks:
This assessment relates to Learning Outcomes 3, 4 and 5.

Learning resources

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

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
- Solutions to tutes, labs 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 electronic submission). 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 system for this unit, which you can access via links in the my.monash portal.

Recommended text(s)

W. N. Venables, D. M. Smith. (2013). *An Introduction to R*. () Available from: http://www.cran.r-project.org/doc/manuals/R-intro.pdf.

M. Allerhand. (2011). *A tiny handbook of R*. () SpringerLink (Online service), Online access via Library.

Pang-Ning Tan, Michael Steinbach, Vipin Kumar. (2006). *Introduction to data mining*. () Addison-Wesley.

Luis Torgo. (2011). *Data mining with R: learning with case studies*. () Chapman & Hall CRC.

Foster Provost and Tom Fawcett. (2013). *Data Science for Business*. () O'Reilly Media, Inc..

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

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.h

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