COMP257

Data Science

S2 Day 2018

Dept of Computing

Contents

General Information

Unit convenor and teaching staff
Convener, Lecturer
Steve Cassidy
steve.cassidy@mq.edu.au
Contact via Email
E6A377
By Appointment

Lecturer
Jia Wu
jia.wu@mq.edu.au
Contact via Email
E6A71
By Appointment

Tutor
Sonit Singh
sonit.singh@mq.edu.au

Credit points
3

Prerequisites
(COMP115 or COMP125) AND (STAT150 or STAT170 or STAT171)

Corequisites

Co-badged status
ITEC657

Unit description
This unit introduces students to the fundamental techniques and tools of data science, such as the graphical display of data, predictive models, evaluation methodologies, regression, classification and clustering. The unit provides practical experience applying these methods using industry-standard software tools to real-world data sets. Students who have completed this unit will be able to identify which data science methods are most appropriate for a real-world data set, apply these methods to the data set, and interpret the results of the analysis they have performed.

Important Academic Dates

Information about important academic dates including deadlines for withdrawing from units are available at https://www.mq.edu.au/study/calendar-of-dates

Learning Outcomes

On successful completion of this unit, you will be able to:

Identify the appropriate Data Science analysis for a problem and apply that method to the problem.

Interpret Data Science analyses and summarise and identify the most important aspects of a Data Science analysis.

Present the results of their Data Science analyses both verbally and in written form.

Discuss the broader implications of Data Science analyses.

General Assessment Information

If you receive special consideration for the final exam, a supplementary exam will be scheduled in the week of **December 17-21 2018**. By making a special consideration application for the final exam you are declaring yourself available for a resit during the supplementary examination period and will not be eligible for a second special consideration approval based on pre-existing commitments. Please ensure you are familiar with the policy prior to submitting an application. Approved applicants will receive an individual notification one week prior to the exam with the exact date and time of their supplementary examination.

Assessment Tasks

Name	Weighting	Hurdle	Due
Workshop Checkpoints	10%	Yes	Every Week
Portfolio	20%	No	Weeks 4, 7, 10, 12
Data Science Project	30%	No	Weeks 8, 13
Final Exam	40%	No	TBC

Workshop Checkpoints

Due: **Every Week**
Weighting: **10%**
This is a hurdle assessment task (see assessment policy for more information on hurdle assessment tasks)

When you attend each weekly workshop you will be asked to complete a task, either practical or involving discussion with the class. **You must complete the checkpoint at least 8 out of 13**

weekly classes to pass the course.

On successful completion you will be able to:

- Interpret Data Science analyses and summarise and identify the most important aspects of a Data Science analysis.

Portfolio

Due: **Weeks 4, 7, 10, 12**
Weighting: **20%**

The portfolio assessment will consist of **three** small data analysis problems that you will be given through the semester. These will involve writing code to analyse one or more data sets. You will submit versions of these for feedback and an initial mark (5%) in weeks 4, 7 and 10, and then submit a final version in week 12 as a portfolio including some reflection on the work you've done through the semester for the final 5%.

On successful completion you will be able to:

- Identify the appropriate Data Science analysis for a problem and apply that method to the problem.
- Interpret Data Science analyses and summarise and identify the most important aspects of a Data Science analysis.
- Present the results of their Data Science analyses both verbally and in written form.
- Discuss the broader implications of Data Science analyses.

Data Science Project

Due: **Weeks 8, 13**
Weighting: **30%**

In groups of 2-3, students will be given one or more datasets and are asked to develop an analysis of this data and present a report. This project should include using more than one dataset, finding data out in the world, cleaning and analysing the data, training a predictive model and using the model to make some conclusions. The report should be reproducible, all methods not only documented but available as an executable archive along with the data.

- Proposal and scoping document (week 8): 5%
- Final report (week 13): 20%
- Project presentation (week 13): 5%

On successful completion you will be able to:

- Identify the appropriate Data Science analysis for a problem and apply that method to the problem.

- Interpret Data Science analyses and summarise and identify the most important aspects of a Data Science analysis.
- Present the results of their Data Science analyses both verbally and in written form.
- Discuss the broader implications of Data Science analyses.

Final Exam

Due: **TBC**
Weighting: **40%**

The exam will assess your knowledge and understanding of the data analysis and machine learning methods covered in the semester.

On successful completion you will be able to:

- Interpret Data Science analyses and summarise and identify the most important aspects of a Data Science analysis.
- Discuss the broader implications of Data Science analyses.

Delivery and Resources

Classes

There will be one two hour lecture each week and one two hour workshop in the computing laboratory. You are expected to attend both classes as they provide complimentary learning activities each week. In practical classes you will write code and experiment with various data sets; in lectures we will discuss the methods you are learning and how the results of your analysis can be interpreted.

Textbooks

We will refer to the following texts during the semester:

Introduction to Data Science A Python Approach to Concepts, Techniques and Applications **Igual**, Laura, **Seguí**, Santi (electronic edition available via MQ Library)

Computational and Inferential Thinking: The Foundations of Data Science By Ani Adhikari and John DeNero (available on GitBooks)

You will be given readings from these and other sources each week.

Technology Used and Required

We will make use of Python 3.6 for data analysis, including a range of modules such as *scikit-learn, pandas, numpy* that provide additional features. These can all be installed via the Anaconda Python distribution. We will discuss this environment and the installation process in the first week of classes.

We will use Jupyter Notebook as a way of developing and presenting the analysis results. This

is included in the full Anaconda distribution.

Project Work

A major part of the assessment in this unit is based on a project that you will complete in groups. This will allow you to explore the techniques you are learning in class in a real-world data analysis exercise.

Unit Schedule

The indicative list of topics is shown here, this is subject to change based on feedback from the class.

1	Overview of DS, Learning Python, Notebooks	SC
2	Data formats, Python input and output	SC
3	Descriptive Statistics, simple visualisation	SC
4	Causality and correlation; Visualisation	SC
5	Predictive Modelling: Linear and Logistic Regression	SC
6	Time Series Modelling	SC
7	Feature sets and spaces; Unsupervised learning	SC
8	Supervised Learning: K-Nearest Neighbours	JW
9	Naive Bayes Classifiers	JW
10	Applying Machine Learning	JW
11	Learning Decision Trees	JW
12	Dealing with Big Data; Data Science in Business	JW
13	Project Presentations	SC/JW

Policies and Procedures

Macquarie University policies and procedures are accessible from Policy Central (https://staff.m

q.edu.au/work/strategy-planning-and-governance/university-policies-and-procedures/policy-central). Students should be aware of the following policies in particular with regard to Learning and Teaching:

- Academic Appeals Policy
- Academic Integrity Policy
- Academic Progression Policy
- Assessment Policy
- Fitness to Practice Procedure
- Grade Appeal Policy
- Complaint Management Procedure for Students and Members of the Public
- Special Consideration Policy *(**Note:** The Special Consideration Policy is effective from 4 December 2017 and replaces the Disruption to Studies Policy.)*

Undergraduate students seeking more policy resources can visit the Student Policy Gateway (https://students.mq.edu.au/support/study/student-policy-gateway). It is your one-stop-shop for the key policies you need to know about throughout your undergraduate student journey.

If you would like to see all the policies relevant to Learning and Teaching visit Policy Central (https://staff.mq.edu.au/work/strategy-planning-and-governance/university-policies-and-procedures/policy-central).

Student Code of Conduct

Macquarie University students have a responsibility to be familiar with the Student Code of Conduct: https://students.mq.edu.au/study/getting-started/student-conduct

Results

Results shown in *iLearn*, or released directly by your Unit Convenor, are not confirmed as they are subject to final approval by the University. Once approved, final results will be sent to your student email address and will be made available in eStudent. For more information visit ask.mq.edu.au.

Student Support

Macquarie University provides a range of support services for students. For details, visit http://students.mq.edu.au/support/

Learning Skills

Learning Skills (mq.edu.au/learningskills) provides academic writing resources and study strategies to improve your marks and take control of your study.

- Workshops
- StudyWise
- Academic Integrity Module for Students
- Ask a Learning Adviser

Student Services and Support

Students with a disability are encouraged to contact the Disability Service who can provide appropriate help with any issues that arise during their studies.

Student Enquiries

For all student enquiries, visit Student Connect at ask.mq.edu.au

IT Help

For help with University computer systems and technology, visit http://www.mq.edu.au/about_us/offices_and_units/information_technology/help/.

When using the University's IT, you must adhere to the Acceptable Use of IT Resources Policy. The policy applies to all who connect to the MQ network including students.

Graduate Capabilities

Capable of Professional and Personal Judgement and Initiative

We want our graduates to have emotional intelligence and sound interpersonal skills and to demonstrate discernment and common sense in their professional and personal judgement. They will exercise initiative as needed. They will be capable of risk assessment, and be able to handle ambiguity and complexity, enabling them to be adaptable in diverse and changing environments.

This graduate capability is supported by:

Learning outcomes

- Interpret Data Science analyses and summarise and identify the most important aspects of a Data Science analysis.
- Present the results of their Data Science analyses both verbally and in written form.
- Discuss the broader implications of Data Science analyses.

Assessment tasks

- Workshop Checkpoints
- Portfolio
- Data Science Project
- Final Exam

Discipline Specific Knowledge and Skills

Our graduates will take with them the intellectual development, depth and breadth of knowledge, scholarly understanding, and specific subject content in their chosen fields to make them competent and confident in their subject or profession. They will be able to demonstrate, where relevant, professional technical competence and meet professional standards. They will be able

to articulate the structure of knowledge of their discipline, be able to adapt discipline-specific knowledge to novel situations, and be able to contribute from their discipline to inter-disciplinary solutions to problems.

This graduate capability is supported by:

Learning outcomes

- Identify the appropriate Data Science analysis for a problem and apply that method to the problem.
- Interpret Data Science analyses and summarise and identify the most important aspects of a Data Science analysis.
- Present the results of their Data Science analyses both verbally and in written form.

Assessment tasks

- Workshop Checkpoints
- Portfolio
- Data Science Project
- Final Exam

Critical, Analytical and Integrative Thinking

We want our graduates to be capable of reasoning, questioning and analysing, and to integrate and synthesise learning and knowledge from a range of sources and environments; to be able to critique constraints, assumptions and limitations; to be able to think independently and systemically in relation to scholarly activity, in the workplace, and in the world. We want them to have a level of scientific and information technology literacy.

This graduate capability is supported by:

Learning outcomes

- Identify the appropriate Data Science analysis for a problem and apply that method to the problem.
- Interpret Data Science analyses and summarise and identify the most important aspects of a Data Science analysis.
- Discuss the broader implications of Data Science analyses.

Assessment tasks

- Workshop Checkpoints
- Portfolio
- Data Science Project
- Final Exam

Problem Solving and Research Capability

Our graduates should be capable of researching; of analysing, and interpreting and assessing data and information in various forms; of drawing connections across fields of knowledge; and they should be able to relate their knowledge to complex situations at work or in the world, in order to diagnose and solve problems. We want them to have the confidence to take the initiative in doing so, within an awareness of their own limitations.

This graduate capability is supported by:

Learning outcomes

- Identify the appropriate Data Science analysis for a problem and apply that method to the problem.
- Present the results of their Data Science analyses both verbally and in written form.
- Discuss the broader implications of Data Science analyses.

Assessment tasks

- Portfolio
- Data Science Project
- Final Exam

Effective Communication

We want to develop in our students the ability to communicate and convey their views in forms effective with different audiences. We want our graduates to take with them the capability to read, listen, question, gather and evaluate information resources in a variety of formats, assess, write clearly, speak effectively, and to use visual communication and communication technologies as appropriate.

This graduate capability is supported by:

Learning outcomes

- Present the results of their Data Science analyses both verbally and in written form.
- Discuss the broader implications of Data Science analyses.

Assessment tasks

- Portfolio
- Data Science Project
- Final Exam

Changes from Previous Offering

Following a review of the first offering last year, there are a number of changes in this offering.

- More of an emphasis on machine learning including supervised and unsupervised

learning methods

- In-term tests replaced with an end of semester exam; feedback from students and staff indicated that having two in-term tests was harder to manage and distracted from the project and practical work in semester.
- The final exam is worth 40% of the final mark which is more than the Quizzes were worth last year, to balance this the marks for the Portfolio and Project have been reduced. This hopefully gives a bit more balance to the assessment with contributions from three main components.

Assessment Standards

COMP257 will be graded according to the following general descriptions of the letter grades as specified by Macquarie University; following the general description is additional description of the standards specific to this unit.

High Distinction (HD, 85-100): provides consistent evidence of deep and critical understanding in relation to the learning outcomes. There is substantial originality and insight in identifying, generating and communicating competing arguments, perspectives or problem solving approaches; critical evaluation of problems, their solutions and their implications; creativity in application as appropriate to the discipline.

In the context of this unit, the Portfolio and Project tasks provide an opportunity to show your deep and critical understanding of the methods of Data Science in this unit. An HD student will show some originality and insight in these reports; they will show that they have mastered the techniques we have covered and have gone beyond these to discover new methods.

Distinction (D, 75-84): provides evidence of integration and evaluation of critical ideas, principles and theories, distinctive insight and ability in applying relevant skills and concepts in relation to learning outcomes. There is demonstration of frequent originality in defining and analysing issues or problems and providing solutions; and the use of means of communication appropriate to the discipline and the audience.

In the context of this unit, a D student will display some integration and evaluation in the Portfolio and Project reports - these will go beyond being a simple account of a data analysis, they will make it clear that the student has mastered the techniques and methods covered in the unit and understands how and why they are used together.

Credit (Cr, 65-74): provides evidence of learning that goes beyond replication of content knowledge or skills relevant to the learning outcomes. There is demonstration of substantial understanding of fundamental concepts in the field of study and the ability to apply these concepts in a variety of contexts; convincing argumentation with appropriate coherent justification; communication of ideas fluently and clearly in terms of the conventions of the discipline.

In the context of this unit, a Cr. student shows good performance in all assessment tasks, in particular the portfolio and project reports will show their complete understanding of the tools that they have learned to use.

Pass (P, 50-64): provides sufficient evidence of the achievement of learning outcomes. There is demonstration of understanding and application of fundamental concepts of the field of study; routine argumentation with acceptable justification; communication of information and ideas adequately in terms of the conventions of the discipline. The learning attainment is considered satisfactory or adequate or competent or capable in relation to the specified outcomes.

In the context of this unit, the P student shows good performance in most assessment tasks and is able to complete the major parts of all portfolio notebook tasks satisfactorily. The student is able to complete an analysis using the tools we have learned with some guidance.

Fail (F, 0-49): does not provide evidence of attainment of learning outcomes. There is missing or partial or superficial or faulty understanding and application of the fundamental concepts in the field of study; missing, undeveloped, inappropriate or confusing argumentation; incomplete, confusing or lacking communication of ideas in ways that give little attention to the conventions of the discipline.

Changes since First Published

Date	Description
23/07/2018	Add special consideration text.