COMP8410 and COMP3425 Data Mining, Semester 1 2018

19 February to 25 May, 2018

This course is an introduction to data mining and the general skills for selecting and applying data mining algorithms. We cover a breadth of common and emerging techniques with the aim of understanding where they might be useful and how to use them properly, while understanding limitations. There is an emphasis in the course on data-driven practical work over the mathematical and statistical foundations.

Detailed course content will be made available during the course at the Wattle site, https://wattlecourses.anu.edu.au/course/view.php?id=22838

Quick Reference

Mode of Delivery	On-campus, 12 weeks plus exam. Most material is presented on-line, supplemented with a 1-hour lecture and 2-hour lab in most weeks.
Prerequisites	(COMP7240 or COMP6240 or COMP2400) and (COMP6730 or COMP7230). No prereq for students enrolled in the MScQBB or MSc Adv version (sic)
Incompatible courses	COMP3420, COMP3425, COMP8400
Co-Taught courses	COMP3425, COMP8410
Course Convenor	A/Prof Kerry Taylor
Phone	6125 8560
Email	**comp8410@anu.edu.au**
Office hours for consultation	Tuesdays 3-5pm
Research Interests	Semantic Web, Machine Learning, Spatial and IoT data management
Administrator	Bethany Flanders
Email	data.analytics@anu.edu.au
Lecturer	A/Prof Kerry Taylor
Tutor	TBA
Email	
Phone	
Tutor	TBA
Email	
Phone	
Textbook (not required)	Han, Kamber & Pei, **Data Mining: Concepts and Techniques** 3rd Edition, 2011. www.elsevier.com/books/data-mining-concepts-and-techniques/han/978-0-12-381479-1 . It is available in the university library in soft and hard copy and the on-campus bookshop. The second edition would also be adequate.
Other recommended references	Graham Williams, Data Mining with Rattle and R, The Art of Excavating Data for Knowledge Discovery, Springer 2011. http://www.springer.com/gp/book/9781441998897 .

	Witten, Frank, Hall and Pal, Data Mining, Practical Machine Learning Tools and Techniques, 4th Edition, Elsevier 2017. https://www.elsevier.com/books/data-mining/witten/978-0-12-804291-5

p/g only Learning Outcomes COMP8410

Upon successful completion of this course, students will:

1. Critically analyse and justify the steps involved in the data mining process,
2. Anticipate and identify data issues related to data mining,
3. Research, test and apply the principal algorithms and techniques used in data mining,
4. Justify suitable techniques to use for a given data mining problem,
5. Appraise and reflect upon the results of a data mining project using suitable measurements,
6. Investigate application areas and current research directions of data mining,
7. Reflect upon ethical and social impacts of data mining.

u/g only Learning Outcomes COMP3425

Upon successful completion of this course, students will:

1. Critically analyse and justify the steps involved in the data mining process,
2. Anticipate and identify data issues related to data mining,
3. Test and apply the principal algorithms and techniques used in data mining,
4. Justify suitable techniques to use for a given data mining problem,
5. Appraise and reflect upon the results of a data mining project using suitable measurements,
6. Reflect upon ethical and social impacts of data mining.

Assessment Scheme

Assessment components, weighting and due dates

Assessment Task	**Value %**	**Due Date**	**Learning outcomes**
Weekly online quiz	5 redeemable	Normally in class lab only	All
Written Assignment 1	20	5pm Friday 16 March	1, 6, 7
Lab exam	20	Evening Thursday 22 March	1, 2, 3, 4, 5
Practical Assignment 2	20	5pm Friday 18 May	2, 3, 4, 5

Final exam	35	University exam period	All
OVERALL MARK	100		

- For assignment topics please see the Wattle site.
- Weekly online quizzes will be normally be undertaken during the lab class in which you are enrolled. There is no lab class in Week 1 and the Week 1 quiz can be taken in your own time from 8am Monday Week 1 to 11:59pm Friday Week 2. Marks for all quizzes will be totalled and scaled to contribute 5% to the overall course mark. Quizzes are primarily intended for self-learning. Feedback on correct answers is given and multiple attempts are permitted.
- **Note that the lab exam must be attended in person.**

Final course mark

- All assessment components, with the exception of the online quizzes, must be submitted to pass the course. Each such component is a "hurdle" under the ANU Rules. A student **must attempt both assignments, the lab exam, and the final exam** to pass the course.
- For students with a quiz mark that is proportionally lower than the final exam mark, the 5% of the overall mark given for quizzes will be redeemed by 5% for the final exam mark, giving the quiz mark a value of 0% and final exam mark a value of 40% in the overall mark.
- **At least 50% overall** is required to pass the course.
- The lab exam is a hurdle assessment under the ANU Rules, whereby a minimum/pass mark in a particular assessment(s) item (assignment or exam) is required to pass the course, regardless of performance in other items. A mark of **at least 45% for the lab exam** is required to pass the course.
- A **supplementary exam** will be offered to those students who have an overall mark of at least 45% and either
    - An overall mark of less than 50%; or
    - have failed a hurdle assessment.
- Note that **marks may be moderated** so that raw marks for assessment components as well as final marks may be scaled by the convenor or as a result of school or college academic review.

Policy on late assessment and re-marking

- **Assessment submitted after the due date and time will not be accepted and will not be marked.**
- Extensions to the due date for submission will only be granted if requests are made to the convenor at comp8410@anu.edu.au well in advance, stating the reasons for requiring the extension, evidence to support the reason (usually a medical certificate), and the extension period requested.
- Students may consider applying for special consideration. An application form must be completed and lodged online within three business days of the original due date of the assessment task.
- Any appeals or request for re-consideration regarding an assessment piece must be submitted in writing to comp8410@anu.edu.au within two weeks of the assessment results being released.

Academic Misconduct

Students are expected to have read the ANU Academic Misconduct Rule before commencement of the course. **No group work is permitted in any part of the assessment in this course**. Plagiarism will not be tolerated and University procedures will be applied ruthlessly. Therefore your contributions are expected to be yours alone, except for work that is clearly attributed appropriately. You may find this a helpful guide to plagiarism: http://thevisualcommunicationguy.com/2014/09/16/did-i-plagiarize-the-types-and-severity-of-plagiarism-violations/

Support for Students

The University offers a number of support services for students. Information on these is available online from http://students.anu.edu.au/studentlife/

Course Organisation

Please spend a while familiarising yourself with the Wattle course page.

You will see on Wattle that there is a section for each week of the course. Sections may not be visible until the respective time period commences, to help you pace your way through the course. You are expected to work through the course notes by self-study or in study groups if you prefer.

Each section includes a **description** of the topic to be covered and **extensive course notes**. In some sections there is a **recorded video lecture**, sometimes supported by **lecture slides** that are also available in the section. Usually there are **paper-based exercises** to assist you to apply the course topics in practice. The exercises are considered compulsory and you will have trouble if you do not keep up with them. Some exercises build on the results of previous work.

There are also software-based **practical exercises** (titled "practical exercises") embedded in the course notes. You may do these exercises at your own pace if you prefer, or they may be undertaken in labs with the support of your tutor. If you do not have time to complete them in your scheduled lab, please do complete them outside classes. Either way, the practical exercises are mandatory components of the course. **In Week 1 you will be asked to enrol in a lab at a time to suit you and no labs will be held in week 1.**

Most, but not all, of the course material is sourced from the course text, Han, Kamber and Pei. References to **relevant sections of the text** are given so that you may refer to the text for alternative explanations and extension material. For some topics, **additional reading** is prescribed.

Most weeks include an **open-book but individual self-assessment quiz** to be undertaken in labs and that you are advised to attempt as many times as you need to gain confidence in more theoretical aspects of the course topic. These quizzes form a mandatory part of the course assessment and **can only be attempted during the lab in which you are enrolled.** For this reason, **you are strongly advised to complete all other reading and paper-based exercises for the week prior to your scheduled lab** when the quiz will be undertaken.

Communication and getting help

Please see information on contacting the course staff at the top of this document. Wattle forums are available to support learning but are not compulsory. Please pay attention to **course announcements on the News forum** as these are sometimes critical for course completion and

assessment. Feedback will be provided for submitted assignments, generally within two weeks of due date.

ANU is committed to the demonstration of educational excellence and regularly seeks feedback from students. One of the key formal ways students have to provide feedback is through Student Experience of Learning Support (SELS) surveys. The feedback given in these surveys is anonymous and provides the Colleges, University Education Committee and Academic Board with opportunities to recognise excellent teaching, and opportunities for improvement. For more information on student surveys at ANU and reports on the feedback provided on ANU courses, see http://unistats.anu.edu.au/surveys/selt/students/ and http://unistats.anu.edu.au/surveys/selt/results/learning/ .

Workload

An ANU 6 unit Course is designed for around **130 hours** of student effort over the 12 weeks. For this course, this includes **4 hours per week of semester for self-study** when you are expected to work through the extensive course materials posted on Wattle. Typically, around 3 hours of face-to-face lecture and laboratory work is required, although the pattern varies in some weeks. The time budget also includes assignment work: each assignment is designed for 15 hours. Any remaining time should be used for additional reading such as the text book, recommended papers, self-study, and review and reflection.

Required Resources

A laptop or desktop is required for accessing the course material on Wattle and for completing the practicals and assignments. University student labs may be used. Rattle and R will be used extensively in this course and are available in the CSIT labs or, as they are freely available, you may install them on your own equipment.

Additional Course Costs

The only additional costs would be the purchase of one or more of the recommended books for this course. Successful completion of the course does not require such purchase.