

# Unit of Study Outline

**Unit Code** QBUS3820

**Unit Title** Machine Learning and Data Mining in Business

**Semester 1, 2018**

**Pre-requisite Units:** QBUS2810 or ECMT2110 or DATA2002

**Co-requisite Units:**

**Prohibited Units:**

**Assumed Knowledge and/or Skills:**

**Unit Coordinator:** Assoc. Prof. Peter Radchenko

**Address:** Room 4158, Abercrombie Precinct(H70), The University of Sydney NSW 2006

**Email:** peter.radchenko@sydney.edu.au **Phone:**

**Consultation Hours:** Please go to Blackboard for details of all staff consultation times.

**Class Day(s):** Please go to the University of Sydney timetables for class times and locations.

**Required Text / Resources:**

**There are no required textbooks. Recommended reading:**

1. James, G., Witten, D., Hastie, T., Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R*, Springer, New York. Free download available at <http://www-bcf.usc.edu/~gareth/ISL>

2. T Hastie, R Tibshirani and J Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Second Edition. Springer, Springer Series in Statistics, 2009; available for free at <https://web.stanford.edu/~hastie/ElemStatLearn/>

**Tutorials:** Students will practice data analysis and machine learning techniques in labs using Python.

This unit of study outline MUST be read in conjunction with

The Business School Unit of Study Common Policy and implementation information that applies to every unit of study offered by the Business School (<http://sydney.edu.au/business/currentstudents/policy>). All assessment rules, such as standards used, penalties etc, are covered.

The Business School Student Administration Manual - for information about all processes such as illness, appeals etc ( [http://sydney.edu.au/business/currentstudents/student\\_information/student\\_administration\\_manual](http://sydney.edu.au/business/currentstudents/student_information/student_administration_manual)) When deciding applications and appeals relating to these matters it will be assumed that every student has taken the time to familiarise themselves with these key policies and procedures.

The Business School seeks feedback from students and staff in order to continually improve all units offered. For information on previously collected feedback and innovations made in response to this feedback, please see [http://sydney.edu.au/business/learning/planning\\_and\\_quality/feedback/student](http://sydney.edu.au/business/learning/planning_and_quality/feedback/student)

## 1. Unit of Study Information

Advances in information technology have made available rich information data sets, often generated automatically as a by-product of the main institutional activity of a firm or business unit. Data Mining deals with inferring and validating patterns, structures and relationships in data, as a tool to support decisions in the business environment. This unit offers an insight into the main statistical methodologies for the visualisation and the analysis of business and market data, providing the information requirements for specific tasks such as credit scoring, prediction and classification, market segmentation and product positioning. Emphasis is given to empirical applications using modern software tools.

## 2. Program Learning Outcomes and Unit Learning Outcomes

The Program Learning Outcomes for this Program are located at <http://sydney.edu.au/business/about/accreditations/AoL>

### Unit Learning Outcomes

Unit Learning Outcomes	Program Learning Outcomes
<b>On the successful completion of the Unit you should be able to:</b>	
<p>The objectives are that, at the completion of the unit, students</p> <p>(1) know the statistical theory required for business data mining and data analysis</p> <p>(2) can identify which statistical tool is most relevant for specific business analytic tasks</p> <p>(2) can identify advantages and limitations of each method</p> <p>(3) can extract information from large volumes of data readily available from the business environment,</p> <p>(4) can obtain and interpret a meaningful analytical result using a software package such as Python</p> <p>(5) can work productively in a team</p> <p>(6) can present and write about their findings effectively</p>	<p>1. Depth of disciplinary expertise</p> <p>2. Critical thinking and problem solving</p> <p>3. Communication</p> <p>4. Information literacy</p> <p>8. An integrated professional, ethical and personal identity</p>

## 3. Assessment

Assessment Title	Assessment Type	Individual/Group	Assessment Conditions	Program Learning Outcomes Assessed	Length	Weight	Due Time	Due Date	Closing Date
Group Project	Assignment	Group	Compulsory	1, 2, 3, 4	N/A	20%	4:00pm	08-Jun-2018	22-Jun-2018
Online Quizzes	Small continuous assessment	Individual	Compulsory	2	N/A	15%		Weekly	Weekly
Mid semester exam	Tutorial quiz, small test or online task	Individual	Compulsory	1, 2, 3	N/A	20%	11:30am	30-Apr-2018	30-Apr-2018
Final Exam	Final exam	Individual	Compulsory	1, 2, 3	N/A	45%		Final Exam Period	Final Exam Period
Academic Honesty								Week 4	

### Assessment details

#### Group Project

- Task Description
 

Form a team of 3 or 4 students. The purpose of this project is to simulate real work experience as a business analytics professional team. The team will design a statistical model using the data analysis and machine learning methods taught in this unit; a training dataset will be provided. More detailed instructions will be given in class.
- Assessment Criteria
  1. Presentation, communication & style (written)
  2. Methodology used is the most appropriate to the aims and objectives of the task
  3. Analysis
  4. Critical reasoning / critical thinking

- Feedback - What, when and how feedback will be provided for this assessment  
Feedback may be provided if necessary.

### Online Quizzes

---

- Task Description  
There will be two online quizzes, each weighing 7.5%. The purpose of the quizzes is to provide on-going feedback on students' learning, understanding and progressing with the concepts and ideas presented in the unit. This assignment should also encourage regular work by students.
- Assessment Criteria
  1. Use of literature/ Knowledge of theory
  2. Problem solving
  3. Critical reasoning / critical thinking
- Feedback - What, when and how feedback will be provided for this assessment  
Answers of the online quizzes will be released after the due date.

### Mid semester exam

---

- Task Description  
The mid-term exam will be run during the lecture class in week 8. It will cover all the material from weeks 1-7 (inclusive). A more detailed description of the mid-semester exam will be given in weeks 6 and 7.
- Assessment Criteria
  1. Use of literature/ Knowledge of theory
  2. Problem solving
  3. Critical reasoning / critical thinking
- Feedback - What, when and how feedback will be provided for this assessment  
Answers to the exam questions will be released within two weeks of the exam date.

### Final Exam

---

- Task Description  
The final exam will cover all the material and concepts discussed in this unit over the full 13 weeks of the semester. A more-detailed description of the final exam will be provided in class.
- Assessment Criteria
  1. Clarity of expression (incl. accuracy, spelling, grammar, punctuation)
  2. Use of literature/ Knowledge of theory
  3. Analysis
  4. Problem solving
  5. Critical reasoning / critical thinking
- Feedback - What, when and how feedback will be provided for this assessment  
There will be a standard final exam review session.

#### 4. Other Resources for Students

All lectures and seminars are recorded and will be available within the LMS for student use. Please note the Business School does not own the system and cannot guarantee that the system will operate or that every class will be recorded. Students should ensure they attend and participate in all classes.

##### **PASS**

Peer-Assisted Study Sessions (PASS) is a peer-facilitated, cooperative student support scheme aiming to improve students' academic performance by providing extra, free learning opportunities with trained student facilitators. Each session includes problem solving practice in areas directly related to understanding the unit concepts more thoroughly. PASS features small class sizes and a friendly personal approach to help you reinforce key course requirements in a collaborative, active and social learning environment. Registrations open in the first week of semester.

For more information or to register, please visit:

<https://sydney.edu.au/students/peer-assisted-study-sessions.html>

Email all enquiries about the PASS program to: [business.pass@sydney.edu.au](mailto:business.pass@sydney.edu.au)

##### **Maths in Business**

The Business School provides a free series of workshops with student facilitators open to all students interested in mastering both basic and upper intermediate level mathematics. The program includes workshops from basic to intermediate level maths and Excel, both of which are fundamental skills for any aspiring business student. The Maths in Business program helps ensure high achievement throughout your business degree. Most foundational units of study require a firm understanding of maths, and this is essential to senior units in accounting, business analytics and finance. Registrations open in the first week of semester.

For more information or to register, please visit: <https://sydney.edu.au/students/maths-in-business.html>

Email all enquiries about the Maths in Business program to: [business.maths@sydney.edu.au](mailto:business.maths@sydney.edu.au)

## 5. Unit Schedule

Week	List of Topics	Assessments Due
1 5 Mar 2018	Introduction to data science and statistical machine learning (notation; key concepts and problems).	
2 12 Mar 2018	Linear regression (the MLR model; least squares estimation; linear regression and linear algebra; statistical inference; interpretation; review of key concepts).	
3 19 Mar 2018	Regression modelling I (K-Nearest Neighbors; bias-variance trade-off; non-parametric vs. parametric approaches).	
4 26 Mar 2018	Regression modelling II (diagnostics; data transformations; categorical predictors; interactions).	
<b>Common week 2 Apr to 8 Apr</b>		
5 9 Apr 2018	Model selection (Mallow's Cp; AIC; BIC; validation set; cross-validation; bootstrap).	Quiz 1
6 16 Apr 2018	Linear model selection and regularization (variable selection; ridge regression; the LASSO; elastic net; principal components regression).	
7 23 Apr 2018	Classification I (logistic regression; evaluating classification models).	
8 30 Apr 2018	In-class mid-term exam.	Mid-term exam
9 7 May 2018	Classification II (linear discriminant analysis; quadratic discriminant analysis; regularized logistic regression; K-nearest neighbors).	
10 14 May 2018	Nonlinear modelling (polynomial regression; regression splines; smoothing splines; local regression; generalized additive models).	
11 21 May 2018	Tree-based methods (decision trees; bagging, random forests, boosting).	Quiz 2
12 28 May 2018	Support Vector Machines (maximal margin classifiers, support vector classifiers and machines; relationship to logistic regression).	
13 4 Jun 2018	Discussion and Review.	Project assignment