UNSW Course Outline

DATA9001 Fundamentals of Data Science - 2024

Published on the 12 May 2024

General Course Information

Course Code : DATA9001
Year : 2024
Term : Term 2
Teaching Period : T2
Is a multi-term course? : No
Faculty : Faculty of Science
Academic Unit : School of Mathematics & Statistics
Delivery Mode : In Person
Delivery Format : Standard
Delivery Location : Kensington
Campus : Sydney
Study Level : Postgraduate
Units of Credit : 6

Useful Links

Handbook [Class Timetable](Class-Timetable)

Course Details & Outcomes

Course Description

This course provides a broad overview of Data Science as a platform for further studies in Data
Science and an understanding and appreciation of Data Science in the modern world.

Students will study the fundamentals of Data Science as it is applied in Computer Science, Economics, and Mathematics and Statistics. They will be introduced to topics such as databases, data analytics, data mining, Bayesian statistics, statistical software, econometrics, machine learning and business forecasting.

The content of this course will be delivered via weekly live lectures with academics from three different Schools: The School of Mathematics and Statistics, the School of Economics and the School of Computer Science and Engineering. These concepts will be further explored through a series of tutorials/workshops.

Course Aims

The aim of the course is to provide a broad overview of probability theory, different statistical methods, regression analysis, and modern data science techniques. This course will provide a platform for further studies in Data Science and Machine Learning.

Course Learning Outcomes

Course Learning Outcomes
CLO1 : Apply probability rules in a given setting to calculate key quantities.
CLO2 : Use key theoretical tools to explore the properties of random variables.
CLO3 : Apply key methods of statistical inference in applied settings.
CLO4 : Use R/RStudio to perform statistical computations and simulations.
CLO5 : Apply various data visualisation tools, perform regression analysis and draw causal inference from data.
CLO6 : Apply fundamental data science techniques and tools, including machine learning, Naïve Bayes classification, Decision trees, K-NN, unsupervised learning and neural networks.

Course Learning Outcomes	Assessment Item
CLO1 : Apply probability rules in a given setting to calculate key quantities.	• Statistics Assignment • Final Exam
CLO2 : Use key theoretical tools to explore the properties of random variables.	• Statistics Assignment • Final Exam
CLO3 : Apply key methods of statistical inference in applied settings.	• Statistics Assignment • Final Exam
CLO4 : Use R/RStudio to perform statistical computations and simulations.	• Statistics Assignment • Final Exam
CLO5 : Apply various data visualisation tools, perform regression analysis and draw causal inference from data.	• Economics Assignment • Final Exam
CLO6 : Apply fundamental data science techniques and tools, including machine learning, Naïve Bayes classification, Decision trees, K-NN, unsupervised learning and neural networks.	• Computer Science Assignment • Final Exam

Learning and Teaching Technologies

Moodle - Learning Management System

Learning and Teaching in this course

Students in DATA9001 are enrolled in one weekly classroom tutorial. The classroom tutorials involve going through selected tutorial questions in detail. The exercises for each week's classroom tutorial are available on Moodle. The main reason for having classroom tutorials is to give you a chance to tackle and discuss problems that you find difficult or do not fully understand. Therefore, it is essential that you try at least a selection of tutorial problems before attending your classroom tutorial so that you know the questions you would like to ask of your

tutor.

Assessments

Assessment Structure

Assessment Item	Weight	Relevant Dates
Computer Science Assignment Assessment Format: Individual	15%	Start Date: Not Applicable Due Date: Typically due in Week 10
Economics Assignment Assessment Format: Individual	15%	Start Date: Not Applicable Due Date: Typically due in Week 7
Statistics Assignment Assessment Format: Individual	15%	Due Date: Typically due in Week 3
Final Exam Assessment Format: Individual	55%	Due Date: Exam Period

Assessment Details
Computer Science Assignment

Assessment Overview

This assignment builds on and extends skills and knowledge acquired in the Computer Science part of the course. The assignment is inspired by real-life scenarios. Specific details of the assessment will be available during the term. This assignment is typically due in Week 10.

The rationale of this assessment is to give you feedback on your progress and mastery of the material, and to obtain measures of your progress towards the stated learning outcomes. You will receive feedback in the form of a marked rubric and individualised comments.

Course Learning Outcomes

- CLO6 : Apply fundamental data science techniques and tools, including machine learning, Naïve Bayes classification, Decision trees, K-NN, unsupervised learning and neural networks.

Detailed Assessment Description

This assignment builds on and extends skills and knowledge acquired in the Computer Science part of the course. The assignment is inspired by real-life scenarios. Specific details of the assessment will be available during the term. This assignment is typically due in Week 10.

The rationale of this assessment is to give you feedback on your progress and mastery of the material, and to obtain measures of your progress towards the stated learning outcomes. You will receive feedback in the form of a marked rubric and individualised comments.

Economics Assignment

Assessment Overview

This is a written assessment that will assess your ability to apply some of the concepts covered in this course to a real-world business problem by performing regression analysis on a dataset and applying various data visualization tools. You are expected to write a short report of under 1,000 words. This assignment is typically due in Week 7.

Specific details of the assessment will be available during the term.

You will receive feedback in the form of a marked rubric and individualised comments.

Course Learning Outcomes

- CLO5 : Apply various data visualisation tools, perform regression analysis and draw causal inference from data.

Detailed Assessment Description

This is a written assessment that will assess your ability to apply some of the concepts covered in this course to a real-world business problem by performing regression analysis on a dataset and applying various data visualization tools. You are expected to write a short report of under 1,000 words. This assignment is typically due in Week 7.

Specific details of the assessment will be available during the term.

You will receive feedback in the form of a marked rubric and individualised comments.

Assignment submission Turnitin type

This assignment is submitted through Turnitin and students do not see Turnitin similarity reports.

Statistics Assignment

Assessment Overview

This assignment will assess your understanding of the material covered from Week 1 to 3. You will be expected to compute probabilities and use key theoretical tools to explore properties of random variables. This assignment is typically due in Week 3.

You will receive feedback in the form of a marked report, with detailed worked solutions for each question after the due date.

The rationale of this assessment is to give you feedback on your progress and mastery of the material, and to obtain measures of your progress towards the stated learning outcomes.

<u>Course Learning Outcomes</u>

- CLO1 : Apply probability rules in a given setting to calculate key quantities.
- CLO2 : Use key theoretical tools to explore the properties of random variables.
- CLO3 : Apply key methods of statistical inference in applied settings.
- CLO4 : Use R/RStudio to perform statistical computations and simulations.

<u>Detailed Assessment Description</u>

This assignment will assess your understanding of the material covered from Week 1 to 3. You will be expected to compute probabilities and use key theoretical tools to explore properties of random variables. This assignment is typically due in Week 3.

You will receive feedback in the form of a marked report, with detailed worked solutions for each question after the due date.

The rationale of this assessment is to give you feedback on your progress and mastery of the material, and to obtain measures of your progress towards the stated learning outcomes.

<u>Assignment submission Turnitin type</u>

This assignment is submitted through Turnitin and students do not see Turnitin similarity reports.

Final Exam

<u>Assessment Overview</u>

The final exam is designed to evaluate your learning and problem-solving skills on all topics delivered across all weeks of the term, including material from lectures, tutorials and workshops. The exam is typically 2hrs long and consists of multiple-choice questions, short numerical and short answer responses - details will be confirmed during the course. The examination will occur during the official university examination period.

The rationale of this assessment is to give you feedback on your progress and mastery of the material, and to obtain measures of your progress towards the stated learning outcomes.

Additional feedback is available through consultation with the course convenor.

Course Learning Outcomes

- CLO1 : Apply probability rules in a given setting to calculate key quantities.
- CLO2 : Use key theoretical tools to explore the properties of random variables.
- CLO3 : Apply key methods of statistical inference in applied settings.
- CLO4 : Use R/RStudio to perform statistical computations and simulations.
- CLO5 : Apply various data visualisation tools, perform regression analysis and draw causal inference from data.
- CLO6 : Apply fundamental data science techniques and tools, including machine learning, Naïve Bayes classification, Decision trees, K-NN, unsupervised learning and neural networks.

Detailed Assessment Description

The final exam is designed to evaluate your learning and problem-solving skills on all topics delivered across all weeks of the term, including material from lectures, tutorials and workshops. The exam is typically 2hrs long and consists of multiple-choice questions, short numerical and short answer responses - details will be confirmed during the course. The examination will occur during the official university examination period.

The rationale of this assessment is to give you feedback on your progress and mastery of the material, and to obtain measures of your progress towards the stated learning outcomes.

Additional feedback is available through consultation with the course convenor.

Assignment submission Turnitin type

Not Applicable

General Assessment Information

For all assessments for this course, no late submissions will be accepted past the deadline.

Grading Basis

Standard

Course Schedule

Teaching Week/Module	Activity Type	Content
Week 1 : 27 May - 2 June	Module	Introduction to probability
Week 2 : 3 June - 9 June	Module	Random variables and random samples
Week 3 : 10 June - 16 June	Module	Inferential statistics and estimation
Week 4 : 17 June - 23 June	Module	Introduction to econometrics and R
Week 5 : 24 June - 30 June	Module	Regression
Week 6 : 1 July - 7 July	Module	Flexibility week
Week 7 : 8 July - 14 July	Module	Forecasting
Week 8 : 15 July - 21 July	Module	Programming and database management systems
Week 9 : 22 July - 28 July	Module	Supervised machine learning
Week 10 : 29 July - 4 August	Module	Unsupervised learning & neural networks

Attendance Requirements

Students are strongly encouraged to attend all classes and review lecture recordings.

Course Resources

Prescribed Resources

There are no prescribed texts for this course.

Recommended Resources

The following list of textbooks is only recommended and **NOT** compulsory for this course:

- A Modern Introduction to Probability and Statistics, F. M. Dekking, C. Kraaikamp, H. P. Lopuhaa⬚ and L. E. Meester

- Introduction to Statistics and Data Analysis, Christian Heumann and Michael Schomaker Shalabh

- Introduction to Statistical Learning, by Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani

- Business Statistics by Sharpe, De Veaux and Velleman (2010). Boston, MA: Addison Wesley.

- The R software: fundamentals of programming and statistical analysis, Lafaye de Micheaux, P., Drouilhet, R., & Liquet, B. (2013).

- Introductory econometrics: A modern approach by Wooldridge, J. M. (2015). Cengage learning.

- Forecasting: principles and practice, Hyndman, R. J., & Athanasopoulos, G. (2018). OTexts.

- Introduction to Data Science, Laura Igual and Santi Seguí

- Python Programming, an Introduction to Computer Science, John M. Zelle, Wartburg College

- Data Science Concepts and Practice, Vijay Kotu and Bala Deshpande, Morgan Kaufmann Publishing

- Introductory econometrics: A modern approach by Wooldridge, J. M. (2015)

Course Evaluation and Development

Student feedback is very important to continual course improvement. This is demonstrated within the School of Mathematics and Statistics by the implementation of the UNSW online student survey myExperience, which allows students to evaluate their learning experiences in an anonymous way. myExperience survey reports are produced for each survey. They are released to staff after all student assessment results are finalised and released to students. Course convenor will use the feedback to make ongoing improvements to the course.

Staff Details

Position	Name	Email	Location	Phone	Availability	Equitable Learning Services Contact	Primary Contact
	Jeffrey Kwan					Yes	Yes
	Martin Eftimoski					No	No
	Hao Xue					No	No

Other Useful Information

Academic Information

Upon your enrolment at UNSW, you share responsibility with us for maintaining a safe, harmonious and tolerant University environment.

You are required to:

- Comply with the University's conditions of enrolment.
- Act responsibly, ethically, safely and with integrity.
- Observe standards of equity and respect in dealing with every member of the UNSW community.
- Engage in lawful behaviour.
- Use and care for University resources in a responsible and appropriate manner.
- Maintain the University's reputation and good standing.

For more information, visit the UNSW Student Code of Conduct Website.

Academic Honesty and Plagarism

Referencing is a way of acknowledging the sources of information that you use to research your assignments. You need to provide a reference whenever you draw on someone else's words, ideas or research. Not referencing other people's work can constitute plagiarism.
Further information about referencing styles can be located at https://student.unsw.edu.au/referencing

Academic integrity is fundamental to success at university. Academic integrity can be defined as a commitment to six fundamental values in academic pursuits: honesty, trust, fairness, respect, responsibility and courage. At UNSW, this means that your work must be your own, and others' ideas should be appropriately acknowledged. If you don't follow these rules, plagiarism may be detected in your work.
Further information about academic integrity, plagiarism and the use of AI in assessments can be located at:

- The Current Students site,
- The ELISE training site, and
- The Use of AI for assessments site.

The Student Conduct and Integrity Unit provides further resources to assist you to understand your conduct obligations as a student: https://student.unsw.edu.au/conduct

Submission of Assessment Tasks

Penalty for Late Submissions
UNSW has a standard late submission penalty of:

- 5% per day,
- for all assessments where a penalty applies,
- capped at five days (120 hours) from the assessment deadline, after which a student cannot

submit an assessment, and
- no permitted variation.

Any variations to the above will be explicitly stated in the Course Outline for a given course or assessment task.

Students are expected to manage their time to meet deadlines and to request extensions as early as possible before the deadline.

Special Consideration

If circumstances prevent you from attending/completing an assessment task, you must officially apply for special consideration, usually within 3 days of the sitting date/due date. You can apply by logging onto myUNSW and following the link in the My Student Profile Tab. Medical documentation or other documentation explaining your absence must be submitted with your application. Once your application has been assessed, you will be contacted via your student email address to be advised of the official outcome and any actions that need to be taken from there. For more information about special consideration, please visit: https://student.unsw.edu.au/special-consideration

Important note: UNSW has a "fit to sit/submit" rule, which means that if you sit an exam or submit a piece of assessment, you are declaring yourself fit to do so and cannot later apply for Special Consideration. This is to ensure that if you feel unwell or are faced with significant circumstances beyond your control that affect your ability to study, you do not sit an examination or submit an assessment that does not reflect your best performance. Instead, you should apply for Special Consideration as soon as you realise you are not well enough or are otherwise unable to sit or submit an assessment.

Faculty-specific Information

Additional support for students

- The Current Students Gateway
- Student Support
- Academic Skills and Support
- Student Wellbeing, Health and Safety
- Equitable Learning Services
- UNSW IT Service Centre
- Science EDI Student Initiatives, Offerings and Guidelines

School-specific Information

School of Mathematics and Statistics and UNSW Policies

The School of Mathematics and Statistics has adopted a number of policies relating to enrolment, attendance, assessment, plagiarism, cheating, special consideration etc. These are in addition to the Policies of The University of New South Wales. Individual courses may also adopt other policies in addition to or replacing some of the School ones. These will be clearly notified in the Course Initial Handout and on the Course Home Pages on the Maths Stats web site. Students in courses run by the School of Mathematics and Statistics should be aware of the School and Course policies by reading the appropriate pages on the web site starting at: The School of Mathematics and Statistics assessment policies

The School of Mathematics and Statistics will assume that all its students have read and understood the School policies on the above pages and any individual course policies on the Course Initial Handout and Course Home Page. Lack of knowledge about a policy will not be an excuse for failing to follow the procedure in it.

Special Consideration - Short Extension Policy

The School of Mathematics and Statistics has carefully reviewed its range of assignments and projects to determine their suitability for automatic short extensions as set out by the UNSW Short Extension Policy. Upon comprehensive examination of our course offerings that incorporate these types of assessments, we have concluded that our current deadline structures already accommodate the possibility of unexpected circumstances that may lead students to require additional days for submission. Consequently, the School of Mathematics and Statistics has decided to universally opt out of the Short Extension provision for all its courses, having pre-emptively integrated flexibility into our assessment deadlines. The decision is subject to revision in response to the introduction of new course offerings. Students may still apply for Special Consideration via the usual procedures.

Computing Lab

The main computing laboratory is room G012 of the Anita B.Lawrence Centre (formerly Red Centre). You can get to this lab by entering the building through the main entrance to the School of Mathematics (on the Mezzanine Level) and then going down the stairs to the Ground Level. A second smaller lab is Room M020, located on the mezzanine level through the glass door (and

along the corridor) opposite the School's entrance.

For more information, including opening hours, see the [computing facilities webpage](#). Remember that there will always be unscheduled periods when the computers are not working because of equipment problems and that this is not a valid excuse for not completing assessments on time.

School Contact Information

Please visit the [School of Mathematics and Statistics website](#) for a range of information.

For information on Courses, please go to "Student life & resources" and either Undergraduate and/or Postgraduate and respective "Undergraduate courses" and "Postgraduate courses" for information on all course offerings.

All school policies, forms and help for students can be located by going to the "Student Services" within "Student life & resources" page. We also post notices in "Student noticeboard" for your information. Please familiarise yourself with the information found in these locations. If you cannot find the answer to your queries on the web you are welcome to contact the Student Services Office directly.

Undergraduate

E: ug.mathsstats@unsw.edu.au

P: 9385 7011 or 9385 7053

Postgraduate

E: pg.mathsstats@unsw.edu.au

P: 9385 7053

Should we need to contact you, we will use your official UNSW email address of in the first instance. **It is your responsibility to regularly check your university email account. Please use your UNSW student email and state your student number in all emails to us.**