MATH 2009 INTRODUCTION TO DATA SCIENCE

Credit Points 10

Legacy Code 301033

Coordinator Liwan Liyanage (https://directory.westernsydney.edu.au/search/name/Liwan Liyanage/)

Description This subject is replaced by equivalent subject COMP 2025 Introduction to Data Science from 2022. Analysis of data is essential for scientific investigation, modelling processes and predicting future events. Data Science is the investigation of the tools required that allow us to perform this modelling and prediction. The increase in accessible data over the past few decades has promoted the use of Data Science, making it a desired skill in many professions. In this subject we further investigate the methods of regression, clustering and classification that form the basis of a data scientist's toolbox.

School Computer, Data & Math Sciences

Discipline Computer Science, Not Elsewhere Classified.

Student Contribution Band HECS Band 2 10cp

Check your fees via the Fees (https://www.westernsydney.edu.au/currentstudents/current_students/fees/) page.

Level Undergraduate Level 2 subject

Pre-requisite(s) For students NOT enrolled in 3769 Bachelor of Data Science or 3770 Bachelor of Applied Data Science - MATH 1028 Statistical Decision Making or MATH 1003 Biometry or MATH 1030 Statistics for Business

Co-requisite(s) For students enrolled in 3769 Bachelor of Data Science or 3770 Bachelor of Applied Data Science - MATH 1033 Thinking About Data

Assumed Knowledge

Computer Programming.

Learning Outcomes

1. Discuss the issues (computational and social) in data science.
2. Choose and apply appropriate data analysis methods for a given set of data.
3. Define the assumptions required behind each data analysis model.
4. Choose and apply appropriate evaluation methods to computed models to gain insight into the effectiveness of the models.
5. Use computer software (R) to achieve the above.

Subject Content

1. Introduction to Data Science
2. Classification
3. Tree-based methods
4. Clustering and unsupervised learning
5. High dimensional data sets
6. Visualisation of model performance
7. Data science ethics
8. Regression

Assessment

The following table summarises the standard assessment tasks for this subject. Please note this is a guide only. Assessment tasks are regularly updated, where there is a difference your Learning Guide takes precedence.

Type	Length	Percent	Threshold	Individual/ Group Task	Mandatory
Multiple Choice	10 min per quiz	20	N	Individual	Y
Numerical Problem Solving	to consist of 10 or so pages of text, not including code and output	30	N	Individual	Y
Viva Voce	15 min per student	10	Y	Individual	Y
Numerical Problem Solving	to consist of 10 or so pages of text, not including code and output	40	N	Individual	Y