

# CS909-15 Data Mining

**20/21**

**Academic year**

20/21

**Department**

Computer Science

**Level**

Taught Postgraduate Level

**Module leader**

Fayyaz Minhas

**Credit value**

15

**Module duration**

10 weeks

**Assessment**

Multiple

**Study location**

University of Warwick main campus, Coventry

---

## Description

### Introductory description

Data Mining.

### Module aims

Understanding of the value of data mining in solving real-world problems;  
Understanding of foundational concepts underlying data mining;  
Understanding of algorithms commonly used in data mining tools;  
Ability to apply data mining tools to real-world problems.

### Outline syllabus

This is an indicative module outline only to give an indication of the sort of topics that may be covered. Actual sessions held may differ.

Introduction to machine learning, basic concepts and motivation;  
Data pre-processing and basic data transformations;  
Regression models (linear regression, logistical regression);  
Classification: decision trees, probabilistic generative models;

Model evaluation, bias-variance trade-off;  
Ensemble methods: boosting, bagging & random forests;  
Dimensionality reduction: Principal Component Analysis (PCA), T-distributed Stochastic Neighbour Embedding (t-SNE);  
Introduction to deep learning, backpropagation, gradient descent;  
Convolutional neural networks;  
Word embeddings;  
Sequence-to-sequence models;  
Attention mechanisms and memory networks;  
Unsupervised deep learning and generative models;  
Transfer learning.

## **Learning outcomes**

By the end of the module, students should be able to:

- Display a comprehensive understanding of different data mining tasks and the algorithms most appropriate for addressing them.
- Evaluate models/algorithms with respect to their accuracy.
- Demonstrate capacity to perform a self-directed piece of practical work that requires the application of data mining techniques.
- Critique the results of a data mining exercise.
- Develop hypotheses based on the analysis of the results obtained and test them.
- Conceptualise a data mining solution to a practical problem.

## **Indicative reading list**

Please see Talis Aspire link for most up to date list.

[View reading list on Talis Aspire](#)

## **Research element**

The students shall be required to explore the literature about latest methods related to classification and deep learning

## **Interdisciplinary**

Data mining lies at the intersection of statistics, computer science and mathematics.

## **Subject specific skills**

Design of data mining solutions  
Learning to develop novel algorithms related to machine learning  
Conducting proper experiment design in machine learning

## **Transferable skills**

Experiment design  
Critical Thinking  
How to conduct literature reviews

---

## Study

### Study time

Type	Required
Lectures	30 sessions of 1 hour (20%)
Practical classes	10 sessions of 1 hour (7%)
Private study	110 hours (73%)
Total	150 hours

### Private study description

Private study should focus on the following components:

- a. Assigned reading
- b. Coding exercises
- c. Assignment solution
- d. Review of the lab component
- e. Revision of the lecture slides

### Costs

No further costs have been identified for this module.

---

## Assessment

You do not need to pass all assessment components to pass the module.

Students can register for this module without taking any assessment.

### Assessment group D1

	Weighting	Study time
Assignment 2	35%	
Assignment 1	25%	
CS909 Data Mining	40%	
CS909 Examination		

## Weighting

## Study time

~Platforms - AEP

### Assessment group R

#### Weighting

#### Study time

CS909 resit examination

100%

CS909 MSc resit examination

~Platforms - AEP

### Feedback on assessment

Formative feedback will be provided in lab sessions and also during lectures where answers are given in class to short exercises.

Summative feedback:

- Written feedback will be provided on the practical assignment and will be given electronically with explanation on the mark given.

[Past exam papers for CS909](#)

---

## Availability

### Pre-requisites

No Warwick module is required as pre-requisite. However familiarity with basic probability and statistics (for example: discrete and continuous random variables, densities and distributions, common distributions including Bernoulli, binomial, uniform and normal distribution, expectations) will be needed.

## Courses

This module is Core for:

- Year 1 of TPSS-C803 Postgraduate Taught Behavioural and and Data Science
- Year 1 of TCSA-G5PA Postgraduate Taught Data Analytics
- Year 1 of TCSA-G5PB Postgraduate Taught Data Analytics (CUSP)

This module is Optional for:

- Year 4 of UCSA-G402 MEng Computing Systems
- Year 5 of UCSA-G403 MEng Computing Systems (Intercalated Year)
- TCYA-G5P5 MSc in Scientific Computing

- Year 1 of G5P5 Scientific Computing
- Year 2 of G5P5 Scientific Computing
- Year 2 of TIMS-L990 Postgraduate Big Data and Digital Futures
- Year 1 of TESA-H641 Postgraduate Taught Communications and Information Engineering
- TCSA-G5PD Postgraduate Taught Computer Science
  - Year 1 of G5PD Computer Science
  - Year 1 of G5PD Computer Science
- Year 1 of TCSA-G5P8 Postgraduate Taught Computer Science and Applications
- Year 1 of TESA-H121 Postgraduate Taught Information Engineering (CUSP)
- Year 1 of TMAA-G1PF Postgraduate Taught Mathematics of Systems
- Year 1 of TSTA-G4P1 Postgraduate Taught Statistics
- Year 1 of TIMA-L99D Postgraduate Taught Urban Analytics and Visualisation
- Year 2 of TIMA-L99C Postgraduate Urban Informatics and Analytics

This module is Option list A for:

- Year 5 of UCSA-G504 MEng Computer Science (with intercalated year)
- Year 4 of UCSA-G402 MEng Computing Systems
- Year 5 of UCSA-G403 MEng Computing Systems (Intercalated Year)
- Year 1 of TIMS-L990 Postgraduate Big Data and Digital Futures
- Year 1 of TMAA-G1PF Postgraduate Taught Mathematics of Systems
- Year 1 of TIMA-L99C Postgraduate Urban Informatics and Analytics
- Year 4 of UCSA-G503 Undergraduate Computer Science MEng