

CS950-10 Foundations of Computational Data Analytics

26/27

Department

Computer Science

Level

Taught Postgraduate Level

Module leader

Weiren Yu

Credit value

10

Module duration

10 weeks

Assessment

Multiple

Study location

University of Warwick main campus, Coventry

Description

Introductory description

This module introduces basic concepts and techniques for data analysis with diverse types, including tabular, text, time series, and geospatial data. Students will gain practical skills in data cleaning, dimensionality reduction, visualisation, and exploratory analysis, as well as methods and algorithms for managing and analysing real-world datasets efficiently. It serves as a taster for students considering further study in the Computational Data Analytics module.

Module aims

The module will provide students with knowledge about the broad foundations of computational data analytics which will help them inform their decision when specialising in one of the areas the MSc Computer Science offers. Data Analytics is a core discipline within computer science, with increasing importance in the age of digital transformation and emerging technologies, with significant economic impact. Because of the highly interdisciplinary nature of Data Analytics students will benefit from being able to pursue working in a wide range of application domains.

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.

Overview of common data modalities: tabular, JSON, time series, text, networks, and geospatial data

Data cleaning and preprocessing: standardisation, type casting, outlier detection

Exploratory data analysis: descriptive statistics, pattern discovery

Improving data quality: handling missing data, noise filtering, class balancing

Dimensionality reduction: feature projection, embeddings, variance preservation

Data partitioning and sampling: train-test splits, cross-validation, stratified sampling

Data visualisation: matplotlib, seaborn, ggplot2, t-SNE, and geospatial mapping

Correlation and dependency analysis: covariance, mutual information, partial correlation

Scalable data analysis: 5 Vs, stream processing, MongoDB, PySpark, TensorFlow, cloud computing, data lakes and warehousing

Learning outcomes

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

- Apply data cleaning, preprocessing, and quality assessment techniques to prepare datasets for analysis.
- Perform exploratory data analysis and visualisation to uncover patterns, anomalies, and structural insights.
- Implement methods for data integration, dimensionality reduction, and feature projection.
- Utilise scalable tools and frameworks to analyse and process large datasets.

Indicative reading list

[Reading lists can be found in Talis](#)

Research element

Coursework will include a research element.

Subject specific skills

in line with the learning objectives students will acquire skills in:

Applying data cleaning, preprocessing, and quality assessment techniques to prepare datasets for analysis;

Performing exploratory data analysis and visualisation to uncover patterns, anomalies, and structural insights;

Implementing methods for data integration, dimensionality reduction, and feature projection;

Utilising scalable tools and frameworks to analyse and process large datasets.

Transferable skills

Being able to apply Data Analytics knowledge and understanding of specialist theoretical and methodological approaches, suggesting and incorporating interrelationships with other relevant disciplines in abstract and unpredictably complex contexts.

Students will obtain the cognitive skills to critically contribute to existing discourses and methodologies in Data Analytics, suggesting new ideas, and designing systematic studies in Data Analytics based on critical analysis and evaluation.

Students will obtain practical skills in organising and communicating information, improving interpersonal, team and networking skills through engaging in classes and computer laboratories. Formative assessment will allow students to strategically enhance their own learning.

Data Analytics is an area with immediate relevance for increasing ethical awareness and its practical application regarding privacy concerns. The associated values will help understanding the importance of personal responsibility and ethical leadership.

Study

Study time

Type	Required
Lectures	20 sessions of 1 hour (20%)
Supervised practical classes	9 sessions of 1 hour (9%)
Private study	29 hours (29%)
Assessment	42 hours (42%)
Total	100 hours

Private study description

Private study, background reading and revision.

Costs

No further costs have been identified for this module.

Assessment

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

Assessment group D

	Weighting	Study time	Eligible for self-certification
Foundations of Computational Data Analytics Coursework	30%	12 hours	No
The coursework will consist of developing computer programs to solve practical problems in computational data analytics.			
Foundations of Computational Data Analytics Exam	70%	30 hours	No
Written 2h exam covering the entire module content, timetabled in January.			

- Answerbook Pink (12 page)
- Students may use a calculator

Assessment group R

	Weighting	Study time	Eligible for self-certification
Foundations of Computational Data Analytics Resit Exam	100%		No

- Answerbook Pink (12 page)
- Students may use a calculator

Feedback on assessment

Individual written feedback on coursework.
Past exam papers.

[Past exam papers for CS950](#)

Availability

Courses

This module is Optional for:

- TCSA-G5PD Postgraduate Taught Computer Science
 - Year 1 of G5PD Computer Science

- Year 1 of G5PG Computer Science with specialism in Artificial Intelligence and Machine Learning
- Year 1 of G5PH Computer Science with specialism in Cyber Security
- Year 1 of G5PI Computer Science with specialism in Data Analytics