

WM9A6-15 Machine Learning and Data Engineering

21/22

Department

WMG

Level

Taught Postgraduate Level

Module leader

Michael Mortenson

Credit value

15

Module duration

2 weeks

Assessment

Multiple

Study locations

University of Warwick main campus, Coventry Primary

Distance or Online Delivery

Description

Introductory description

The practical application of data science and artificial intelligence systems requires the ability to process, engineer and manage the flow of data and the selection/implementation of learning algorithms. This module, using the industry-standard Python language, aims to provide students the necessary skills and competencies to implement efficient and reliable code, and employ best practices in data management, algorithm development and machine learning.

Module aims

This module aims to introduce students to many of the advanced statistical and data engineering techniques made possible by innovations in computing and modern processing power. This includes:

- clustering
- dimension reduction
- regression
- classification

- feature engineering
- natural language processing
- high performance computing
- analysis of algorithms and computational complexity.

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.

Pandas, Dask , Spark and data management: Data cleaning; Data validation; Joining and merging datasets; Feature engineering; Automation.

Computational complexity and analysis of algorithms: Big O notation; Compilation; Vectorisation; Distributed processing; Best practices for programming.

Natural language processing: Working with text data; NLP; Topic models and decomposition.

Clustering and Dimension Reduction: Clustering; Dimension reduction.

Supervised Learning: Regression; Clustering; Ensembles.

Learning outcomes

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

- Develop original, non-trivial Python applications and algorithms.
- Implement robust and efficient data pipelines to extract and transform data from a variety of sources
- Evaluate and optimise data engineering algorithms for better computational performance.
- Automate advanced machine learning techniques and critically evaluate the results.
- Implement and optimise machine learning algorithms for statistical and computational performance.

Indicative reading list

[View reading list on Talis Aspire](#)

Interdisciplinary

A mixture of technology/computing topics, statistics/machine learning, and business topics

International

Topics are of high international demand

Subject specific skills

Programming, databases, data engineering, clustering, dimension reduction, regression, classification, ensemble modelling, computational complexity, cloud computing, IT architecture

Transferable skills

Programming, data analysis, team work, critical analysis, IT architecture

Study

Study time

Type	Required
Lectures	10 sessions of 1 hour 30 minutes (10%)
Practical classes	14 sessions of 1 hour 30 minutes (14%)
Online learning (independent)	10 sessions of 1 hour (7%)
Assessment	104 hours (69%)
Total	150 hours

Private study description

No private study requirements defined for this module.

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 A

	Weighting	Study time	Eligible for self-certification
Data Engineering and Machine Learning Pipeline	20%	14 hours	No
Creating a data engineering/machine learning pipeline. Comprises of application/pipeline code and a short (300 word) description			
Post Module Assignment	80%	90 hours	Yes (extension)
An essay on applications and best practices in data engineering and a programmed implementation of a data pipeline			

Assessment group R

	Weighting	Study time	Eligible for self-certification
Post Module Assignment	100%		Yes (extension)
An essay on applications and best practices in data engineering and a programmed implementation of a data pipeline			

Feedback on assessment

Verbal feedback for in-module element. Written feedback and annotated scripts for post-module element

Availability

Courses

This module is Optional for:

- Year 1 of TWMS-H1S4 Postgraduate Taught e-Business Management (Full-time)