

WM931-15 Data Science & Machine Learning

26/27

Department

WMG

Level

Taught Postgraduate Level

Module leader

Amir Kayhani

Credit value

15

Module duration

4 weeks

Assessment

Multiple

Study locations

University of Warwick main campus, Coventry Primary

Distance or Online Delivery

Description

Introductory description

Data Science and Machine Learning have become key drivers of business change and value generation in the modern digital economy. The ability to derive insights, recommendations and automate actions from a wide range of datasets (traditional and non-traditional - i.e. Big Data) is integral to the competitive advantage of many of the world's largest businesses. This module provides practical exposure to these methods, as well as the underlying theories and concepts.

Module aims

This module aims to enable participants to select, implement and evaluate machine learning algorithms in data science. In particular, the module highlights several of the most common, and in-demand, modern algorithms including classification, regression, clustering, dimension reduction and ensemble methods. Alongside technical knowledge, participants should develop an understanding of the applicability of different types of machine learning to common problems, and best practice for data science and Big Data analytics projects.

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.

Data Science Foundations: Core concepts of Data Science & Machine Learning; Data pre-processing & feature engineering.

Unsupervised learning (e.g., K-means clustering or DBSCAN; Principal Component Analysis; Association Rule Mining (Apriori algorithm))

Classification: Theoretical background; Classification Algorithms (e.g., Decision Trees; Random Forest; KNN; Support Vector Machines; Neural Networks and Deep Learning); Model selection and evaluation.

Regression: Theoretical background; Regression algorithms(e.g., Linear models, Lasso Regression or Gaussian Process Regression); Model selection and evaluation.

Reinforcement Learning: Theoretical background, Reinforcement Learning algorithms(e.g., Q-learning, Deep Reinforcement Learning)

Ensemble Methods (e.g, Bagging, Boosting or Voting)

Natural Language Processing: Theoretical background, Sentiment Analysis, Text-Processing, NER, TF-IDF, Large Language Models.

Handling Imbalanced Data(e.g., Biasing the algorithm, Under Sampling, Oversampling or SMOTE)

Computational Complexity and High Performance Computing

Machine Learning in Cloud(e.g., Fine tuning pre-trained models; Colab or SageMaker)

Introduction to explainability in AI(XAI)

Learning outcomes

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

- Interpret and evaluate various use-cases and the applicability of data science and machine learning.
- Develop a comprehensive understanding of the different stages of data science projects.
- Implement optimised machine learning models and solutions and interpret, evaluate and critique the results.
- Develop comprehension of the core topics of data science, machine learning and artificial intelligence.
- Collaboratively implement and present a data science project, using optimised machine learning techniques, and interpret the results.

Indicative reading list

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

[Specific reading list for the module](#)

Interdisciplinary

International

Data science topics/skills are of high international demand

Subject specific skills

Data science, machine learning, statistics, ensemble learning, software development, data analysis

Transferable skills

Programming, statistics and modelling, team work, critical analysis

Study

Study time

Type	Required
Lectures	12 sessions of 1 hour (8%)
Seminars	10 sessions of 1 hour (7%)
Practical classes	8 sessions of 1 hour (5%)
Online learning (independent)	30 sessions of 1 hour (20%)
Private study	30 hours (20%)
Assessment	60 hours (40%)
Total	150 hours

Private study description

Combination of the following:

- Independent learning materials and activities for programming, machine learning and machine learning in cloud
- Reading list, book chapters and articles

Costs

No further costs have been identified for this module.

Assessment

You must pass all assessment components to pass the module.

Assessment group A4

	Weighting	Study time	Eligible for self-certification
Group Assessment	30%	18 hours	No
In teams, participants create a data science solution on a real-world dataset and present their approach. Peer marking process will be adopted in this assessment.			
Assignment	70%	42 hours	Yes (extension)
The assignment includes the implementation(in Python) of a data science and machine learning project based on a business/industry scenario and the provided dataset and writing a project report (including reflection and consultation).			

Assessment group R4

	Weighting	Study time	Eligible for self-certification
Individual Presentation	30%		No
The student develops a data science solution using a real-world dataset, presents their approach, and provides a reflection on the group work in a recorded video.			
Assignment	70%	42 hours	No

Feedback on assessment

Group and individual presentation – verbal feedback after presentation
Assignment - Annotated scripts returned to students, generic written feedback to group.

Availability

There is currently no information about the courses for which this module is core or optional.