WM931-15 Data Science & Machine Learning

23/24

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: K-means clustering; DBSCAN; Principal Component Analysis
Classification: Theoretical background; Naïve Bayes; Decision Trees; Support Vector Machines; Model selection and evaluation.
Regression: Theoretical background; Linear models; Ridge Regression; Lasso Regression; Gaussian Process Regression; Model selection and evaluation.
Ensemble Methods: Bagging; Boosting; Voting.
Computational Complexity and High Performance Computing: Analysis of algorithms; Apache Spark and PySpark; Batch/Online algorithms

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 best practices for data processing and feature engineering.
- Implement, interpret and critique current, professional standard learning models.
- Automate deployment-ready data science pipelines and algorithms.
- Evaluate and interpret the results of machine learning models and tune them to optimise performance.
- Develop comprehension of the core topics of data science, machine learning and artificial intelligence.

Interdisciplinary

Statistics and computer science topics

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)	60 sessions of 1 hour (40%)
Assessment	60 hours (40%)
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 A2

	Weighting	Study time
Model Development	20%	12 hours
In teams, participants create a data science solution on a real-world dataset and present their approach		
Post Module Assignment	65%	40 hours
A two part submission - the first an essay-style question on a data science/machine learning topic; the second a working program that can model a given dataset		
Feature engineering programming task	15%	8 hours
Limited time test of programming and data skills via a feature engineering task		

Assessment group R2

	Weighting	Study time
Post Module Assignment	100%	

A two part submission - the first an essay-style question on a data science/machine learning topic; the second a working program that can model a given dataset

Feedback on assessment

For In-module work – test scores, verbal feedback after presentation
For post module work - 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.