

Module Details	
Module Title	Statistical Applications of Industrial Big Data
Module Code	COS7049-B
Academic Year	2021/2
Credits	20
School	Department of Computer Science
FHEQ Level	FHEQ Level 7

Contact Hours	
Type	Hours
Directed Study	164
Laboratories	12
Lectures	12
Tutorials	12

Availability	
Occurrence	Location / Period
BDA	University of Bradford / Semester 1

Module Aims
<p>Nowadays large amounts of data are collected from many different sources; such data can be used for enhanced benefits and impact to society by evaluating its quality and relevance, integrating with existing information and digital resources, extracting patterns and creating new knowledge for decision support in engineering, healthcare and wellbeing, and society sustainable development.</p> <p>However, large amounts of data create continuous challenges for relevant and effective usage in industry. This module enables students to hands-on development of specialist knowledge in statistical data analysis required to apply data science principles and to provide data-driven, innovative engineering solutions.</p> <p>The module is intended to Engineering, Management, Data Analytics, Computer Science and similar subject graduates to gain hands-on development of advanced knowledge and skills in the application of statistical methods in support of robust big data-based decision-making.</p> <p>Students will explore how statistical applications of industrial big data resources can support knowledge discovery for decision making in domains such as industry 4.0, product design and development, product quality management. The module will also address legal, social ethical and professional aspects of such projects.</p>

Outline Syllabus

Data quality and data cleansing; Data preparation (statistical evaluation of data quality, data cleaning and data transformation);
 Exploring concepts of data cardinality, dimensionality, imbalance, similarity, feature selection;
 Engineering problem solving using Python Programming.
 Data Pre-Processing
 Exploratory data visualisation and introduction to statistical concepts relevant to exploration of Big Data;
 Features engineering (importance, selection, dimensionality reduction, PCA).
 Statistical Classification
 Discuss the methods to handle high dimensional data, large-samples, sample splitting;
 Data-driven algorithms for statistical models of engineering Big Data (classifiers, decision trees, Na?ve Bayes);
 Bootstrap and bagging methods.
 Special Topics
 Introduction to basket market analysis along with association rules;
 Grouping (k-means, high dimensional clustering, sub-space clustering);
 Text Mining.
 Big data risk assessment and management, intellectual property, legal and ethical issues of big data resources.
 Mini Project
 Academic skills workshops (metacognition, academic writing, digital literacy).
 Independent practice through application to a relevant Engineering Big Data individual project.

Learning Outcomes

Outcome Number	Description
1	Critically analyze solutions for big data statistical analysis and processing.
2	Critically analyze available data and produce results, or guide toward appropriate applications of Big Data Statistical Analysis.
3	Identify correlations and construct statistical models from Engineering Big Data Resources.
4	Interpret the results and explain them to non-specialists, reflect and carry out a critical review of the issues related to legal, social, ethical and professional issues, including data management and data protection.
5	Demonstrate the use practical software tools, with a focus on workflow design and experimentation.

Learning, Teaching and Assessment Strategy

To prepare the students ready for world of work, assessments are designed to measure industry ready skills such as presentation skills, report writing skills, team-work skills (using group coursework to strengthened students? ability to work effectively in teams) and peer evaluation. Throughout the module, students will be set formative assessment activities that will help develop confidence in tackling data analysis problems and in the use of the software tools that will support them. The timely constructive feedback from this formative assessment will support students develop the skills and knowledge required for the summative assessment.

The module will be summatively assessed through a group presentation of a mini group project that requires students to identify correlation and construct statistical models from data resources and interpret the results from these models that can be explained to non-specialists. This will be followed by an individual research project that requires students to critically analyse big data solutions and applications, whilst demonstrating skills in using practical software tools.

If a student requires supplementary assessment for re-assessment, they will be set a range of tasks based on a supplementary scenario and data set to demonstrate evidence for the required learning outcomes.

Mode of Assessment			
Type	Method	Description	Weighting
Summative	Coursework - Written	Group mini project and presentation (10 mins)	20%
Summative	Coursework - Written	Individual Research Project (1500 words)	80%

Reading List
To access the reading list for this module, please visit https://bradford.rl.talis.com/index.html

Please note:

This module descriptor has been published in advance of the academic year to which it applies. Every effort has been made to ensure that the information is accurate at the time of publication, but minor changes may occur given the interval between publishing and commencement of teaching. Upon commencement of the module, students will receive a handbook with further detail about the module and any changes will be discussed and/or communicated at this point.

© University of Bradford 2021

<https://bradford.ac.uk>