UNIVERSITY of BRADFORD

Module Descriptor

Module Details	
Module Title	Statistical Applications of Industrial Big Data
Module Code	COS7049-B
Academic Year	2024/5
Credits	20
School	School of Computer Science, AI and Electronics
FHEQ Level	FHEQ Level 7

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

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

Module Aims

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 the 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.

However, large amounts of data create continuous challenges for relevant and effective usage in industry. 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.

This research-informed module enables students to develop both specialist knowledge and enhanced problem-solving skills in statistical data analysis required to apply data science principles, and to provide data-driven, innovative engineering solutions for decision support and data-enhanced applications.

Students will explore how statistical applications of industrial big data resources can support knowledge discovery for decision making in domains such as industry 4.0/5.0, product design and development, product quality management and product safety. The module will also address legal, social ethical and professional aspects of such application domains. Students will also benefit from a hybrid approach of learning and assessment, benefitting from critical research, planning and working as team members, and solution development and demonstration as individual professionals.

The example topics are aligned with the programme of studies topics through relevant data resources and references. The module assessment will provide opportunities to develop interest, knowledge and skills in problem solving applying statistics to industrial big data challenges and projects within the new wave of data-enhanced engineering developments.

Outline Syllabus

Module Context, Content and Aims; Definitions: the concepts of Data, Information and Knowledge; The concept of Big Data and its features (the V's); Strategies for Big Data Analysis. Hands-on: Colab Basics. Hands-on Intro (python).

Big Data Statistical Exploration; KDD Processes; Challenges in Engineering Big Data Quality (Imbalance; Binning, Averages & Distributions). Hands-on: Big Data Resources; Cardinality vs Dimensionality; Imbalance. Introduction of the industrial big data team and individual project(s).

Big Data Exploration (Statistical Data Visualisations); Hands-on: Statistical Features of the Benchmark Datasets.

Big Data Similarity and Distance Measures. Hands-on: Practical exercise: Data Upload. Measuring similarity for Benchmark Data samples.

Features Engineering. Dimensionality. Cardinality. The Curse of Dimensionality. Hands-on: Analysing benchmark datasets.

Dimensionality Reduction: PCA, FPCA, NLPCA. (Fuzzy) Data Binning. Multi-Domain Data. Hands-on: exercises.

Data-Driven Model Building. Large Sample Methods. High Dimensional Data (Big p). Sub-sampling (Big n). Leveraging. Sample Splitting. Multi-Sample Splitting. Central Limit Theorem. Hands-on: Large-Sample Analysis.

Introduction to Statistical Classification. Classifier Training and Validation. Classification Performance. Hands-on: Classification applied to benchmark data sets. Classification Performance: binary classification; multi-class classification.

Classification Algorithms: Decision Trees. Na?ve Bayes. Discriminant Analysis. Linear Classifiers. Logistic Classifiers. Hands-on: Classifier building on benchmark data. Measuring Accuracy.

Ensemble Classifiers: Big n revisited. Bootstrapping. Bagging methods. Random Forests. Hands-on: Ensemble Classifiers applications.

High Dimensional Regressions. Sample Splitting for High-Dimensional Data. Multi-Sample Splitting. Hands-on: High Dimensional Regression Model Building.

Introduction to Basket Market Analysis. Frequent Pattern Analysis. Association Rules. Confidence. Support. Model Validation. Hands-on: Developing association rules for industry data - analyse patterns.

Grouping (Clustering): k-means and variants. Hierarchical Methods. Density-Based Methods. Grid-Based Methods. Model-Based Methods. Outlier Analysis Hands-on: How statistics applies to clustering.

Optional: Introduction to Text Mining. Sentiment Analysis.

Learning Outcomes	
Outcome Number	Description
01	Identify and critically analyse existing resources and solutions for big data statistical analysis and processing challenges, and propose data, risk team and individual work planning.
02	Critically analyse available data, design experiments, develop solutions, produce and evaluate results, and guide toward appropriate suitably-designed applications of big data statistical analysis and learning.
03	Identify correlations and construct statistical learning models from industrial big data resources.
04	Interpret statistical learning model results and communicate them to the general public (including non-specialists), reflect and carry out a critical review of the issues related to legal, social, ethical and professional issues, including data and risk management and data protection.
05	Demonstrate problem solving with computers, the use of data, practical software tools and code, with a focus on workflow design, experimentation and validation.

Learning, Teaching and Assessment Strategy

Lectures will introduce the module structure, fundamental concepts, definitions, critical reviews, metrics, algorithms and processes relevant for the content, cutting edge technologies and developments.

Tutorials will describe the problem-solving opportunities, processes and use critical reviews and feedback to introduce, discuss and promote practical approaches of the theoretical foundations delivered in Lectures.

Labs will be used to deliver hands-on examples and promote practical solutions interactively.

Consequently, direct contact classes will provide the academic environment to:

1. Prepare the students ready for world of work, module content delivery and assessments are designed to provide, exemplify and measure industry ready skills such as problem solving and demonstration with computers, expertise, multidisciplinary understanding of big data challenges and addressing them with presentation skills, report writing skills, team-work skills (using mini group project to strengthen student ability to work effectively in teams) and peer evaluation.

2. Provide opportunities for formative assessment throughout the module: students will be set 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 and professional career enrichment/development.

3. Interactively assess the module summatively 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 using own results and research references. 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.

4. Structure individual studies through homework and research.

5. 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 individually demonstrate evidence for the required learning outcomes.

It is a requirement of the Institution of Engineering and Technology (IET) that students MUST achieve a mark of at least 30% in assessment components weighted above 30% IN ADDITION to achieving a mark of at least 40% in the module overall. This requirement applies ONLY to students on IET accredited programmes, which is the BDA occurrence/version of the module.

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.

---

https://bradford.ac.uk