MODULE DESCRIPTOR

Module Title

Advanced Data Science

Reference	CM4706	Version	2
Created	January 2023	SCQF Level	SCQF 10
Approved	May 2019	SCQF Points	30
Amended	June 2023	ECTS Points	15

Aims of Module

To provide students with a comprehensive understanding of the main principles and practices underlying the retrieval, extraction and mining of text data, including web data, using advanced analytical techniques to make business decisions.

Learning Outcomes for Module

On completion of this module, students are expected to be able to:

1	Analyse and critically evaluate the main concepts of current technologies for information retrieval and natural language processing.
2	Critically appraise the effectiveness of state-of-the-art techniques for web mining and natural language processing.
3	Design a solution that combines methods from information retrieval and natural language processing to develop a solution to a given real-world task.
4	Plan a comparative study to evaluate and interpret results from designing and developing information retrieval and extraction systems for big data.

Indicative Module Content

Information collection: crawling and document pre-processing. Information retrieval: document Indexing, similarity metrics and clustering. Web Analytics. Comparative analysis of information retrieval and visualisation methods. Text extraction, tokenisation, stemming, bag-of-words, n-gram, statistical language models, vector representations and topic models. Word sense disambiguation, phrase and named entity recognition, POS tagging, shallow parsing, syntax and dependency parsing. Document similarity, clustering and classification, information extraction, sentiment analysis using lexicon-based techniques. Case studies on text classification, topic modelling applied to news articles, intelligent search and browse, sentiment analysis and social media mining.

Module Ref:	CM4706 v2

Module Delivery

The module is delivered in Blended Learning mode using structured online learning materials/activities and directed study, facilitated by regular online tutor support. Workplace Mentor support and work-based learning activities will allow students to contextualise this learning to their own workplace. Face-to-face engagement occurs through annual induction sessions, employer work-site visits, and modular on-campus workshops.

Indicative Student Workload	Full Time	Part Time
Contact Hours	30	N/A
Non-Contact Hours	30	N/A
Placement/Work-Based Learning Experience [Notional] Hours	240	N/A
TOTAL	300	N/A
Actual Placement hours for professional, statutory or regulatory body	240	

ASSESSMENT PLAN

If a major/minor model is used and box is ticked, % weightings below are indicative only.

Component 1

Type:	Coursework	Weighting:	100%	Outcomes Assessed:	1, 2, 3, 4
Description:	The coursework will consist of practical development and a written evaluation of the NLP pipeline as applied to a problem in a business context.				

MODULE PERFORMANCE DESCRIPTOR

Explanatory Text

The calculation of the overall grade for this module is based on 100% weighting of C1. An overall minimum grade D is required to pass the module.

Module Grade	Minimum Requirements to achieve Module Grade:
A	The student must achieve an A in C1.
B	The student must achieve a B in C1.
C	The student must achieve a C in C1.
D	The student must achieve a D in C1.
E	The student must achieve an E in C1.
F	The student must achieve an F in C1.
NS	Non-submission of work by published deadline or non-attendance for examination

Module Requirements

Prerequisites for Module	None, in addition to course entry requirements.
Corequisites for module	None.
Precluded Modules	None.

INDICATIVE BIBLIOGRAPHY

1. BEASLEY M., 2013. Practical Web Analytics for User Experience: How Analytics Can Help You Understand Your Users. Morgan Kaufman.
2. GABER, M.M., COCEA, M., WIRATUNGA, N. and GOKER, A., 2015. Advances in Social Media Analysis. Springer.
3. CERI, S. et al., 2013. Web Information Retrieval. Berlin, Germany: Springer.
4. DIPANJAN, S., 2016. Text Analytics with Python: a practical real-world approach to gaining actionable insights from your data. United States: Apress.