1. **Title of the module**

MAST5956 (MA5956) Big Data and Machine Learning

1. **Division and School/Department or partner institution which will be responsible for management of the module:**

Division of Computing, Engineering and Mathematical Sciences (CEMS)

School of Mathematics, Statistics and Actuarial Science

1. **The level of the module (Level 4, Level 5, Level 6 or Level 7)**

Level 5

1. **The number of credits and the ECTS value which the module represents**

15 credits (7.5 ECTS)

1. **Which term(s) the module is to be taught in (or other teaching pattern)**

Spring

1. **Prerequisite and co-requisite modules**

Pre-requisite: None

Co-requisite: None

1. **The programmes of study to which the module contributes**

A Year in Data Analytics, BSc Data Science (including course with a Year in Industry), BSc Data Science with a Foundation Year

1. **The intended subject specific learning outcomes.  
   On successfully completing the module students will be able to:**
2. demonstrate knowledge and critical understanding of the underlying concepts and principles related to the exploration and analysis of different types of large datasets;
3. use a range of established techniques with a reasonable level of skill to access, explore and pre-process large datasets and to analyse using multivariate statistical and data mining techniques;
4. make appropriate use of IT tools for accessing and analysing large datasets, and for presentation of the results of these analyses, both in written and other forms.
5. **The intended generic learning outcomes.  
   On successfully completing the module students will be able to:**
6. make effective use of IT facilities for solving problems;
7. demonstrate the skills needed to work and communicate in a group, including an understanding of the roles of different individuals within a team;
8. communicate straightforward arguments and conclusions reasonably accurately and clearly;
9. manage their own learning and development;
10. communicate technical and non-technical material competently;
11. demonstrate critical thinking skills.
12. **A synopsis of the curriculum**

This module is designed to provide students with an introduction to the use of data analytics tools on large data sets including the analysis of text data. The module will begin by discussing the principles of text-mining and big data. The module will then discuss the techniques that can be used to explore large data sets (including pre-processing and cleaning) and the use of multivariate statistical techniques for supervised and unsupervised learning. The module will conclude by considering several data mining techniques.

Syllabus: What is “big data”? What is text mining? Exploratory data analysis for large datasets, and pre-processing and cleaning; Multivariate statistical analysis (both unsupervised, e.g. factor analysis or principle component analysis, and supervised, e.g. linear discriminant analysis); Data security; Data mining including techniques such as classification trees, neural networks, clustering, text analysis or network analysis.

1. **Reading list (Indicative list, current at time of publication. Reading lists will be published annually)**

Aggarwal. Data mining: the textbook (2015). Springer.

Han, Kember and Pei. Data Mining: Concepts and Techniques. 3rd Edition (2013). Morgan Kaufmann.

Friedman et al. The Elements of Statistical Learning (2009). Springer.

Hand, Mannila and Smyth. Principles of Data Mining (2001). MIT Press.

Silge and Robinson. Text Mining with R: A Tidy Approach (2017). O’Reilly.

1. **Learning and teaching methods**

30 contact hours of computer-based workshops

120 hours of private study

Total number of study hours: 150.

1. **Assessment methods**
   1. Main assessment methods

The module will be assessed by 100% coursework.

Group project to produce code for processing and analysing data, requiring on average between 20 and 30 hours to complete (40%)

Report on an individual project, which is suitable for a non-technical audience – up to 10 pages (50%)

Individual presentation – up to 10 minutes (10%)

13.2 Reassessment methods

100% coursework

1. ***Map of module learning outcomes (sections 8 & 9) to learning and teaching methods (section12) and methods of assessment (section 13)***

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Module learning outcome** | 8.1 | 8.2 | 8.3 | 9.1 | 9.2 | 9.3 | 9.4 | 9.5 | 9.6 |
| **Learning/ teaching method** |  |  |  |  |  |  |  |  |  |
| **Private Study** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** |
| Workshops | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** |
| **Assessment method** |  |  |  |  |  |  |  |  |  |
| Report on group project | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** | **X** |
| Code and presentation | **X** | **X** | **X** | **X** |  | **X** | **X** | **X** | **X** |

1. **Inclusive module design**

The School recognises and has embedded the expectations of current equality legislation, by ensuring that the module is as accessible as possible by design. Additional alternative arrangements for students with Inclusive Learning Plans (ILPs)/declared disabilities will be made on an individual basis, in consultation with the relevant policies and support services.

The inclusive practices in the guidance (see Annex B Appendix A) have been considered in order to support all students in the following areas:

a) Accessible resources and curriculum

b) Learning, teaching and assessment methods

1. **Campus(es) or centre(s) where module will be delivered**

Canterbury

1. **Internationalisation**

Data analysis is an international language with internationally recognised techniques developed and refined by statisticians and analysts across the globe. Mastery of the subject-specific learning outcomes, 8.1 to 8.3, will equip students to apply the theories and techniques of this module in a wide range of international contexts. The module team is drawn from the School of Mathematics, Statistics and Actuarial Science/School of Social Policy, Sociology and Social Research, which includes many members of staff with international experience of teaching and research collaboration.

In compiling the reading list, consideration has been given to the range of texts that are available internationally and a selection of texts has been identified to complement the delivery of the material.

Examples with an international dimension are included in the module where appropriate.

The support SMSAS/SSPSSR provides to its students is also internationally attuned given our international student body.

**DIVISIONAL OFFICE USE ONLY**

**Revision record – all revisions must be recorded in the grid and full details of the change retained in the appropriate committee records.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Date approved | Major/minor revision | Start date of the delivery of revised version | Section revised | Impacts PLOs (Q6&7 cover sheet) |
| 03/03/2021 | Minor | 2022/23 | 1 | No |
|  |  |  |  |  |