School of Computing

Module Code	M33147
Module Title	INTELLIGENT DATA AND TEXT ANALYTICS
Module Coordinator Other lecturers	Dr Ella Haig<ella.haig@port.ac.uk>
Assessment Item number	Item 1 - Set exercise (coursework) (CW)
Assessment Title	IDTA Coursework 1
Date Issued	2023-10-02

Schedule and Deliverables

Deliverable	Value	Format	Deadline / Date	Late deadline ECF deadline
Coursework	50%	- One report file (.pdf) - A single .zip file containing Python code and/or Knime workflows (and/or equivalent from other software).	2023-11-20 16:00 [GMT/BST]	2023-12-04 16:00 (10 working days after deadline)

Notes and Advice

- The Extenuating Circumstances procedure is there to support you if you have had any circumstances (problems) that have been serious or significant enough to prevent you from attending, completing or submitting an assessment on time. If you complete an Extenuating Circumstances Form (ECF) for this assessment, it is important that you use the correct module code, item number and deadline (not the late deadline) given above.
- ASDAC are available to any students who disclose a disability or require additional support for their academic studies with a good set of resources on the ASDAC moodle site
- The University takes any form of academic misconduct (such as plagiarism or cheating) seriously, so please make sure your work is your own. Please ensure you adhere to our *Code of Student Behaviour* and watch the video on Plagiarism.
- Any material included in your coursework should be TECFAC 08 Plagiarism fully cited and referenced in **APA 7** format. Detailed advice on referencing is available from the library .
- Any material submitted that does not meet format or submission guidelines, or falls outside of the submission deadline could be subject to a cap on your overall result or disqualification entirely.
- If you need additional assistance, you can ask your personal tutor, student engagement officer ana.baker@port.ac.uk, academic tutor xia.han@port.ac.uk or your lecturers.
- If you are concerned about your mental well-being, please contact our Well-being service.

Intelligent Data and Text Analytics Coursework 1

Download the 2011 UK Census dataset from the IDTA module page on Moodle.

1. [20%] Descriptive analytics:
    a. [10%] Display and interpret basic statistics for **all** attributes
    b. [10%] Display and interpret 5 visualisation graphs; include one box plot and four visualisations that show relationships between 2 or more variables
2. [20%] Perform classification using 3 algorithms; present and discuss the results; compare the results of the 3 algorithms.
3. [20%] Perform regression on the "No of hours" attribute using 2 algorithms; present and discuss the results; compare the results of the 2 algorithms.
4. [20%] Perform association rule mining using 1 algorithm and interpret the meaning of at least 5 rules.
5. [20%] Perform clustering using 2 algorithms; present and discuss the results; compare the results of the 2 algorithms.

The **marking scheme** [in 100% breakup of marks] **for each of the tasks** is as follows:

20% - Display of results

40% - Analysis of results

20% - Use of tables and figures when reporting the results to facilitate understanding

10% - Organisation, language style and clarity

10% - The use of at least one method/algorithm not covered in the teaching sessions in addition to the number of methods/algorithms mentioned in the tasks; for example, for Task 1, 5 visualisations are required; to get this 10% you need to include at least one more visualisation that was not covered in addition to the required 5.

Deliverables – to be submitted through Moodle:

1. A **pdf** report documenting all tasks:
    - the report should include up to **2000 words overall**, excluding tables and figures;

- **be concise** – do not include general information, such as an overview of a technique or algorithm – include just the relevant information about applying techniques/ algorithms on the data, the results and their interpretation;
- ***as a guideline*** use approximately 400 words per task; make use of tables and figures, as these do not count towards the word count;
- ***there is no hard requirement of words per task*** – some may be longer than others; the only hard requirement is the total of 2000 words overall.

2. A zip file with Python code and/or Knime workflows (and/or equivalent from other software).

IMPORTANT NOTE

All components of this coursework are **individual work**.

Any unacknowledged copying of either printed material or software from any other person or source (including electronic media) constitutes plagiarism, which is a serious disciplinary offence; any cases of plagiarism will be handled using the University disciplinary procedures.

Please ensure that your coursework is anonymous. Your **NAME must not appear** anywhere on the coursework or the coversheet. Please **use your ID only**.

Deadline: Monday 20 November 2023, 4 pm