

School of Computing

Module Code	M33147
Module Title	INTELLIGENT DATA AND TEXT ANALYTICS
Module Coordinator Other lecturers	Dr Atefeh Khazaei <atefah.khazaei@port.ac.uk> Dr Ella Haig<ella.haig@port.ac.uk> Dr Alaa Mohasseb <alaa.mohasseb@port.ac.uk> Dr Grace Golcarenarenji <gelayol.golcarenarenji@port.ac.uk>
Assessment Item number	Item 2 - Set exercise (coursework) (CW)
Assessment Title	IDTA Coursework 2
Date Issued	2025-11-24



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 equivalent from other software).	2026-01-19, 13:00 [GMT/BST]	2026-02-02, 13:00 (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 [Student Conduct Policy](#) and watch the video on [Plagiarism](#).

- Any material included in your coursework should be fully cited and referenced in **APA 7** format. Detailed advice on referencing is available from the [library](#), also see [TECFAC 08 Plagiarism](#).
- 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 eleni.noussi@port.ac.uk or your lecturers.
- If you are concerned about your mental well-being, please contact our [Well-being service](#)

AI Statement

AI tools, including Generative AI, **should not be used to**:

- generate content to include in your assessment submission
- generate code to include in your assessment submission

AI tools, including Generative AI, **can be used to**:

- develop your understanding of the concepts covered in the module
- practice use of analytics and programming skills
- check/improve the clarity of your language

Use AI tools critically. They make mistakes and their language is not precise. For example, accuracy and precision may be used as synonyms, while in machine learning these are metrics with very specific meaning and using them interchangeably is not appropriate.

Intelligent Data and Text Analytics Coursework 2

Download the data from:

<https://archive.ics.uci.edu/ml/datasets/Sentiment+Labelled+Sentences>; choose **one** of the datasets, i.e., Amazon, imdb or Yelp.

1. **[30%]** Preprocess the textual data to remove potential noise. Use, as a minimum, the following preprocessing methods: 1) removing punctuation, 2) removing numbers, 3) removing stop words, 4) changing the text to lower/upper case and 5) lemmatising.
Describe in detail, with at least 3 examples, the changes to the text for each preprocessing method applied.
2. **[30%]** Using the bag-of-words/terms representation, perform classification using 3 algorithms; present and discuss the results; compare the results of the 3 algorithms.
3. **[10%]** Perform classification using a BERT-based model with fine-tuning and compare the results with the 3 algorithms from Task 2.
4. **[30%]** Perform topic detection using 1 algorithm using 10 as the parameter for the number of topics; present and discuss each topic, and assess the quality of the detected topics.

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

**** 65% - Analysis of the results of the experiments you have conducted**

- Include important findings.
- Critically evaluate the advantages and disadvantages of the methods/algorithms /parameters selected.
- Discuss the effect and impact of the methods/algorithms/parameters used, justifying your approach and methods.
- Base your conclusions on critical analysis and the insights you have gained throughout the module.

**** 20% - Appropriate use of tables and figures when reporting the results**

- Include clear, informative titles.
- Provide meaningful and concise visualisations.
- Present information using the appropriate visualisations.
- Pay attention to consistency between values or details in a table/figure.
- Use labels and legends to ensure your figures are clear

**** 15% - Structure and presentation**

- Organise your report so it is clear and easy to read, using sections and subsections as appropriate.
- Pay attention to the quality of your prose, including the use and flow of language, grammar, spelling, and format.
- Use graphics and tables as appropriate (clear and readable) and choose a page layout that is easy to follow and understand.

Deliverables – to be submitted through Moodle/Wiseflow:

1. A **report** documenting all tasks:
 - the report should include up to **3000 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 750 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 the Python code (and/or equivalent from other software).

IMPORTANT NOTE

- You may complete the coursework **individually, OR in pairs, OR in groups of three**.
- 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 19 January 2026, 13:00

Important Information About Group Formation:

- Team members **must belong to the same class group** (Group 1 on Mondays, Group 2 on Fridays).
- You **must define your group on Moodle**, even if you plan to complete the coursework *individually* (a group with one member).
- Once you have agreed with your teammates, create your group and **record all members' names** on Moodle.
- If you choose to work in a group, **you cannot later switch to individual work**, and the same applies if you start individually, you cannot switch to a group.
- If you choose to complete the coursework as a group, you **must include an appendix** in your report clearly outlining **each member's individual contributions**. This ensures transparency and fair assessment.
- By default, it is assumed that **all team members contribute equally** and thus receive the same mark. However, if the lecturer determines that contributions were unequal, **individual marks may differ accordingly**.
- The **deadline to form your teams** for the **second coursework** is **15 December 2025, 13:00**.