UNIVERSITY OF LEEDS

School of Computing: Assessment brief

Module title	Data Mining
Module code	COMP2121
Assignment title	Assessment for COMP2121 Data Mining, Semester 2 2023/24
Assignment type and description	Assessment 1: **Test** submitted as Minerva MCQ Test, 30% weight, 1 hour to complete, week17 (aka week 4 semester 2) release 10:00 Wednesday 21.2.24, deadline 10:00 Thursday 22.2.24 Assessment 2: **Report**, PDF submitted via Minerva, 70% weight, approx 10 hours per student to complete, release week14 (aka week 1 semester 2) Wednesday 21.2.24 deadline week22 (aka week 9 semester 2) Thursday 25.4.24
Rationale	Summative assessment of student knowledge and understanding of the module syllabus.
Word limit and guidance	**Test**: 30 questions. **Report**: up to 12 pages
Weighting	**Test**: 30% **Report**: 70%
Submission deadline	**Test:** 10:00 Thursday 22.2.24 **Report:** Thursday 25.4.24
Submission method	**Test** submitted as Minerva Test **Report** submitted as Minerva Assignment
Feedback provision	**Test** answers will be discussed in a subsequent lecture **Report** feedback via Minerva
Learning outcomes assessed	These exercises will enable you to: learn theory, methods and terminology used in data mining and text analytics; investigate how to apply AI methods, resources and techniques for implementing and evaluating data mining and text analytics in a practical applied research project; summarize and present your knowledge and ideas to a peer audience, in a research proposal report.
Module lead	Prof Eric Atwell
Other Staff contact	Dr Noorhan Abbas

1. Assignment guidance

You are advised to attempt both assessments: Test, Report. There is no May/June exam.

2. Assessment tasks

For Test, each student will take an individual online test in Minerva. The tests will include Multiple Answer Questions: each question has several suggested answers, and for each possible answer you must select if it is correct or not. If you require special arrangements and/or extra time due to a disability or other special circumstances, please make sure you notify lecturer Eric Atwell AND student support well in advance so they can prepare accordingly.

For the Report, you will develop a 6-month research project proposal, using data mining and text analytics theory, methods and technologies for a practical application of your choice. Lectures and online learning resources will include examples of data mining and text analytics methods, techniques, resources, and applications. You can also include other tools and techniques in your research proposal, as appropriate.

When writing an applied data mining and text analytics research proposal, you can learn from advice provided by the Engineering and Physical Sciences Research Council. EPSRC is a major UK supporter and funder for research projects in Engineering and Physical Sciences, including AI. The EPSRC website gives guidance on writing research project proposals, see https://www.ukri.org/councils/epsrc/guidance-for-applicants/what-to-include-in-your-proposal/case-for-support/

A standard proposal to EPSRC includes several forms and tables, but the core document is "**Proposed research and its context**" to include:
Research hypothesis & objectives,
Background,
Importance and contribution to knowledge,
Pilot study,
Programme and methodology,
And Workplan diagram, eg Gantt Chart
You should use these headings to structure your report, and write a section for each of these: 2 pages for Pilot study and Programme and methodology, 1 page each for other sections, plus 1 page for References; maximum 9 pages in total.

In addition, as an Appendix of up to 3 pages, describe your use of data mining and text analytics tools in developing your Report. This could include: tools used in the small pilot study to trial the methods proposed; use of tools like Google Scholar or ChatGPT in searching for background information and drafting the report (include examples of query and results); use of tools like Grammarly, Word or ChatGPT to check and correct grammar and style.

3. General guidance and study support

See Minerva Learning Resources for the module for guidance and study support on the sections of the **Proposed Research and Context**:

Research hypothesis and objectives: Set out your research idea or hypothesis. Explain why the proposed project is novel and timely, e.g. emphasising the scientific ambition, or any potential transformative outcomes. Identify the overall aims of the project and the measurable objectives against which the outputs, outcomes and impacts of the work will be assessed.

Background: Introduce the background to the proposal and explain its context. Explain how this work relates to past research and current projects at Leeds University, in the UK and abroad.
Importance and contribution to knowledge: Explain how the project may contribute to current or future economic success; to future development of key emerging industries; or addresses key societal challenges. Describe how your research would benefit national and international research, and engage with research in other disciplines to broaden the reach of the new knowledge.
Pilot study: Describe your implementation of an initial pilot study to demonstrate feasibility of the project, including selection of a simple case study, acquiring sample data, development and evaluation of your prototype solution.
Programme and methodology: Describe the work programme including research and user evaluation. Identify the contribution of each member of the research team including users and/or stakeholders. Provide milestones and deliverables that you will use to monitor progress, and explain how the project will be managed.
The research work programme should make use of an appropriate methodology for AI projects, such as CRISP-DM; and should include use of at least two data mining and/or text analytics methods, tools or techniques introduced in the module (eg SketchEngine, Weka, ChatGPT)
1-page workplan diagram must match the written description of the research work programme, showing start, end and duration of each phase or work-package.

4. Assessment criteria and marking process

The mark scheme for your report will reflect the EPSRC grading scheme with marks in the range 0-6 for each criterion, EXCEPT the sections on Pilot study, Programme and methodology, and Appendix: these are most important, and get higher weight:

Background (0-6 marks)
Importance and contribution to knowledge (0-6 marks)
Research hypothesis and objectives (0-6 marks)
Pilot study (0-12 marks)
Programme and methodology, with corresponding workplan diagram (0-12 marks)
Appendix: use of data mining and text analytics tools in developing your Report (0-18 marks)
TOTAL: up to 60 marks

Return of marks and feedback will be via Minerva grade center, approximately 3 weeks after submission.

5. Presentation and referencing

Page limits are strict: up to 12 pages in total. Text must be single-spaced spaced Arial 11 with 2cm margins; you cannot include more text by using smaller fonts or narrower margins. You must present your ideas for all specified sections, and references must be included within the page limits. You must submit via Minerva one copy of the report as a PDF document.
References and citations must be in a consistent format. I recommend Kilgarriff referencing style for Artificial Intelligence papers, see https://blog.kilgarriff.co.uk/?p=71 but you are free to use another format, as long as you apply it consistently to all your references.

The quality of written English will be assessed in this work. As a minimum, you must ensure:

- Paragraphs are used
- There are links between and within paragraphs although these may be ineffective at times
- There are (at least) attempts at referencing

- Word choice and grammar do not seriously undermine the meaning and comprehensibility of the argument
- Word choice and grammar are generally appropriate to an academic text

These are pass/ fail criteria. So irrespective of marks awarded elsewhere, if you do not meet these criteria you will fail overall.

6. Submission requirements

You must submit one copy of the report as a PDF document by the due date and time using the "Submit My Work" link on Minerva. The filename should be your name e.g. EricAtwell.pdf

7. Academic misconduct and plagiarism

Academic integrity means engaging in good academic practice. This involves essential academic skills, such as keeping track of where you find ideas and information and referencing these accurately in your work.

By submitting this assignment you are confirming that the work is a true expression of your own work and ideas and that you have given credit to others where their work has contributed to yours.

8. Assessment/ marking criteria grid

The mark scheme for your report will reflect the EPSRC grading scheme with marks in the range 0-6 for each criterion, EXCEPT the sections on Pilot study, Programme and methodology, and Appendix: these are most important, and get higher weight:

Background (0-6 marks)
Importance and contribution to knowledge (0-6 marks)
Research hypothesis and objectives (0-6 marks)
Pilot study (0-12 marks)
Programme and methodology, with corresponding workplan diagram (0-12 marks)
Appendix: use of data mining and text analytics tools in developing your Report (0-18 marks)
TOTAL: up to 60 marks