

Group Project (Assessment Task B) Specification (2023/24)

Each group typically consists of three students (based on your lab group), working on a project under the supervision of your lab tutor. Your project will be to design, develop, evaluate and present an end-to-end ML solution in Python to a real-world problem. There are two deliverables in the project, *a project proposal and a final submission*. The project proposal is one page with a specified structure. The final submission consists of a research paper-like report with a specified structure plus the Python code (i.e. notebook with the appropriate markdown and code cells) of the project.

The learning outcomes: On successful completion of this assessment task students will achieve the following learning outcomes.

MO2 Select and apply machine learning algorithms to formulate solutions to different types of machine learning problems, taking into account criteria such data availability and characteristics, and problem-specific requirements for balancing speed, accuracy, and explainability.

MO3 Implement and evaluate contemporary machine learning solutions to application problems using a range of contemporary frameworks.

MO4 Demonstrate an awareness of the ethical and societal implications of machine learning solutions.

Assignment Schedule

The group project specification is **handed out in Week 4**.

The group allocation: starts in Week 4 with an allocation deadline: 5pm, Friday in Week 5, 1 March 2024.

The project proposals (10% of the total module mark) are due: 5pm, Friday in Week 7, 15 March 2024.

Submission: Please submit your proposal as a PDF on Blackboard. *Only one person on your group should submit*. Please list your group ID and group members (student numbers and names) on the submission pop-up page.

The project final submissions (60% of the total module mark) are due: 5pm, Friday in Week 12, 3rd May 2024.

Total Group Project Weighting: 70%

Submission: Please submit your final project report as a PDF and your supplementary material as a separate ZIP file on Blackboard. Only one person on your group should submit. *Please list your group ID and group members (student numbers and names) on the submission pop-up page.*

The submission links for both the proposal and final submission are in the assessment folder on the module site on the Blackboard

Project Proposal

The first step is to work together as a team to write a proposal. This may take several rounds of revision to become something that everyone is happy with, including your lab tutor.

The project proposal should be one page (up to 400 words). Your project proposal should describe:

- *The problem*
What is the problem that you will be investigating? Why is it interesting?
- *Context and background*
What is the context and background of your project? What have you read or examined about these?
- *The dataset*
What data will you use? If you are collecting a new dataset, how will you do it? If it is an existing dataset, what is it? Why do you choose it?
- *Methodology*
What algorithms are you proposing to apply? If there are existing implementations, will you use them and how? How do you plan to improve or modify such implementations? You don't have to have an exact answer at this point, but you should have a general sense of how you will approach the problem you are working on.
- *Evaluation*
How will you evaluate your results? Qualitatively, what kind of results do you expect (e.g. plots or figures)? Quantitatively, what kind of analysis will you use to evaluate and/or compare your results (e.g. what performance metrics)?

Ensure your proposal is double column, 10pt font size, and includes the names of all authors (i.e., everyone in your project team).

Choosing a Real-world Problem

If you are having trouble thinking of ideas, are there any problems in the world that you care about that might benefit from Machine Learning, i.e. for social good, making the world a better place? Is there something you

really enjoy doing that would be fun to involve Machine Learning in, i.e. a fun problem?

These are just suggestions which you can ignore and choose a completely different category of problem, and doing so will not affect your assessment in anyway.

Choosing/Creating A Dataset

There exist many datasets that can found online, e.g. on UCI Machine Learning Repository (<https://archive.ics.uci.edu/>), Kaggle (<https://www.kaggle.com/datasets>), and many sites that contain image datasets, e.g. Google Open Images Dataset (<https://opensource.google/>). These datasets have been selected for Machine Learning.

There also exist many real-world datasets that have been created for various purposes, e.g. Open Data Bristol (<https://opendata.bristol.gov.uk/>) contains a lot of datasets most of which are created by Bristol City Council. Indeed, there are many, many datasets for us to explore, which itself can be a good exploratory and discovery process.

Since using an existing dataset removes a lot of work on data gathering, if you are using an existing dataset you will be expected to do more work on other aspects of the project. This may involve implementing algorithms yourself, and then extending them with some ideas you may have. Similarly, if you are using existing libraries and frameworks (e.g., scikit-learn), rather than implementing and extending the algorithms yourself, then you will be expected to do more work on other aspects. For example, this may mean more work on implementing custom and clever feature extraction, selection or preprocessing, or perhaps it may mean more work on setting up an experimental framework for rigorous evaluation and analysis.

Project Final Report

Your project final report is required to be up to 6 pages using the provided template, structured like a paper from an AI/Machine Learning conference.

The following is a suggested structure for your report, as well as the rubric that is used for marking your report:

Title, Author(s)

Abstract Briefly describe your problem, approach, and key results. Should be no more than 200 words.

Introduction (10%) Describe the problem you are working on, the aims and objectives of the project, why it's important, and an overview of your results

Related Work (10%) Discuss published work that relates to your project. How is your approach similar or different from others?

Data (10%) Describe the data you are working with for your project. What type of data is it? Where did it come from? How much data are you working with? Did you have to do any pre-processing, filtering, or other kinds of treatment to use this data in your project?

Methods (30%) Discuss your approach for solving the problem that you set up in the introduction. Why is your approach the right thing to do? Did you consider alternative approaches? You should demonstrate that you have applied ideas and skills built up during the module to tackling your problem of choice. It may be helpful to include figures, diagrams, or tables to describe your method or compare it with other methods.

Experiments (30%) Discuss the experiments that you performed to demonstrate that your approach solves the problem. The exact experiments will vary depending on the project, but you might compare with previously published methods, perform an ablation study to determine the impact of various components of your system, experiment with different hyperparameters or architectural choices, use visualization techniques to gain insights into how your model works, discuss common failure modes of your model, etc. You should include graphs, tables, or other figures to illustrate your experimental results.

Conclusion (5%) Summarize your key results - what have you learned? Suggest ideas for future extensions or new applications of your ideas.

Writing and Formatting (5%) Is your paper clearly written and nicely formatted?

Formatting Guidelines LaTeX styles, and Word template:

https://www.ijcai.org/authors_kit

Expected amount of time individually: 80 hours.

Feedback

Formative feedback will be provided by your lab tutor in lab sessions and project supervision. Summative feedback will be provided by the marking grid and comments in the completed group project assessment form as specified in the marking criteria. The group assessment form will be released within 4 weeks of the submission deadline.

Marking Criteria

	0-40%	40-50%	50-70%	70-100%
Introduction (10%)	Identification of the problem and scope of project, description of results	Sufficient definition of the problem, sufficient description of the project scope and results	Good definition of the problem, good description of the project scope and results	Clear definition of the problem, excellent description of the project scope and results
Related Work (10%)	Identification and discussion of related work	Sufficient review of related work, satisfactory discussion of similarities and differences	Good review of related work, good discussion of similarities and differences	Critical appraisal of related work, excellent discussion of similarities and differences
Data (10%)	Use of appropriate techniques to collect datasets or select existing datasets, treatment of data	Use and justification of appropriate techniques to collect datasets or select existing datasets, sufficient treatment of data	Good use and justification of appropriate techniques to collect datasets or select existing datasets, good treatment of data	Clear justification of appropriate techniques to collect datasets or select existing datasets; Excellent treatment of data
Methods (30%)	Selection and use of methods, basic level of technical breadth and depth, identification of relevant issues including ethical issues	Selection and use of methods, sufficient level of technical breadth and depth, sufficient discussion of relevant issues including ethical issues	Good selection and use of methods, good level of technical breadth and depth, good discussion of relevant issues including ethical issues	Excellent selection and use of methods demonstrating an understanding of alternative methods, high level of technical breadth and depth, excellent discussion of relevant issues including ethical issues
Experimental Evaluation (30%)	Generation and analysis of experimental results	Generation and analysis of experimental results using justified methodology, appraisal and evaluation of results	Generation and analysis of experimental results using clearly justified methodology, good appraisal and evaluation of results	Excellent generation and clear analysis of experimental results using fully justified methodology, excellent appraisal and

				evaluation of results
Conclusion (5%)	Concise summary	Summary showing wider understanding	Good summary showing wider understanding	Demonstration of clear understanding and implications of results
Writing and Formatting (5%)	Use of template and basic writing skills	Satisfactory structure, some graphical/tabular illustrations.	Coherent structure, good use of graphical/tabular illustrations.	Well structured, high-level of clarity, clear presentation and illustrations