



Group Project

Never Stand Still

COMP9417: Machine Learning & Data Mining

Aims

- Acquire more hands-on experience with ML techniques
- Gain more practical skills in handling ML problems
- Exercise communication skills in motivating, reporting and summarising work done on a ML task

Group Formation

Group formation must be completed and registered
by 5 pm, Sunday, 16 March, on Moodle ([link](#)).

Group size: 4-5 students per group (no need to be in the same tutorial)

Project Scope

Aim:

Developing a machine learning model to automatically classify customer comments

- The project scope will be released by **Sunday 23 March**,
- The dataset will be released on **Monday 31 March**

Project Start and Data Release

The datasets will be released on
Monday 31st March
and you can officially start working on your group project

Submission

The final group project submission has three parts:

1. Presentation

- Presentation videos need to be in .mp4 format

2. Report

- must be a single document in PDF format.
- must include names and zIDs of ALL team members

3. Code + Data

- Code must be combined into a single tar or zip archive
- Data should be included.

[All submissions will be via the Moodle page;](#)

Note: ONLY ONE person on the team submits all three parts

Marking

Total: 20 marks available

1. Report (14 marks)
2. Presentation (4 marks)
3. Code (2 marks)

Group Configuration

Each team comprises 4-5 group members, and this group must be declared on Moodle under Group Project Member Selection by **the deadline**.

- Teams can consist of students from different tutorials, and groups can consist of PG and UG students.
- Individual contributions to the project will be assessed through a peer-review process which will be announced later after the reports are submitted. This will be used to scale marks based on contribution. Anyone who does not complete the peer review **by 5 pm Friday, 2 May** will be deemed to have not contributed to the assignment. Peer review is confidential, and group members are not allowed to disclose their reviews to their peers.

Member Contributions

- Please note that **80% of your mark will be weighted based on your individual contribution**. Individual contributions will be assessed through a peer-review process.
- We expect all group members to contribute equally to any work submitted.
- In the case the group feels that one or more of the students have not contributed sufficiently, we will take steps to re-distribute the marks accordingly.
- Some good advice: Keep a record of your contributions throughout the project. Keep a record of all communications with other group members (emails/chat), etc. In the event of a group dispute, we will request evidence from all group members about their contribution.

Deliverables

Presentation

Each team has to submit a 2-minute video presentation of their project:

1. Tell a story about their problem, why it is important, how they modelled it, how they evaluated it, challenges and what they discovered. How they improved their model (if applicable)
2. PowerPoint, Google Slides or PDF used in their video

Code

All code files (in Python) that is required to recreate the findings should be submitted in .zip format. Codes are expected to be well-organised and well-commented. It is suggested to include a readme file to give an instruction on how to run the code to replicate the findings.

Report Structure

Each team must write a detailed report outlining their exploration of the data and approach to modelling. The report is expected to be 5-6 pages **excluding the title page and references** (with a single column, 1.5 line spacing) and easy to read. The body of the report should contain the main parts of the presentation, and any supplementary material should be deferred to the appendix. For example, only include a plot if it is important to get your message across.

The guidelines for the report are as follows:

1. Title Page: title of the project, name of the group and all group members (names and zIDs).
2. Introduction: a brief summary of the task, the main issues for the task and a short description of how you approached these issues.

Report Structure

3. Exploratory Data Analysis: this could be a crucial aspect for any ML project and should be done carefully. Some (potential) questions for consideration: are all features relevant? How can we represent the data graphically in an informative way? What is the distribution of the classes? What are the relationships between the features? ...
4. Methodology: An explanation and justification of methods developed, method selection, feature selection, hyper-parameter tuning, evaluation metrics, design choices, etc. State which method has been selected for the final test and its hyper-parameters.
5. Results: Include the results achieved by the different models implemented in your work using a sensible evaluation metric. Be sure to explain how each model was trained and how you chose your final model.

Report Structure

6. Discussion: Compare different models, their features and their performance. What insights have you gained?
7. Conclusion: Give a brief summary of the project and your findings, and what could be improved on if you had more time.
8. References: List of all literature that you have used in your project, if any. You are encouraged to go beyond the scope of the course content for this project.

Marking Scheme

Category [Mark]	Section [Mark]	Comments
Report [14]	Introduction [1]	Clear introduction of the problem and dataset
	Methods [5]	Clear presentation of methods applied
		Clear presentation of pre-processing and feature extraction (if applicable)
		Justification of the method choices
		Clear description of hyper-parameter tuning (if applicable)
		Clear description of evaluation metrics
	Results [4]	Presentation of results using figures and tables
		Result evaluation using appropriate metrics
		Evaluation of various hyper-parameters and design choices
		Analysis of feature importance
Presentation [4]	Presentation [4]	Comparison of different methods, their features and performances
		Discussion of metrics and which are more appropriate and why
Code [2]	Code [2]	Discussion of future improvements
		A summary of the problem, method, results and discussion
		<ul style="list-style-type: none"> • Story telling [3] • Presentation [1]
		Well organised and well-commented code

Deadline

The deadline to submit the deliverables (video presentation, report and code) is by

5 pm, Monday, 28 April.

Project Help

Consult Python package online documentation for using methods, metrics and scores. There are many other resources on the Internet and in literature related to classification. When using these resources, please keep in mind the guidance regarding plagiarism in the course introduction. General questions regarding the group project should be posted in the Group project forum on the course Moodle page.

Peer Review

Individual contributions to the project will be assessed through a peer-review process which will be announced later after the reports are submitted. This will be used to scale marks based on contribution. Anyone who does not complete the peer review by **the deadline** will be deemed to have not contributed to the assignment.

Peer review is a confidential process, and group members are not allowed to disclose their reviews to their peers.

Important Dates:

Group Formation:	Sunday, 16 March, 5:00 pm
Project scope release:	Sunday 23 March
Dataset release:	Monday 31 March, 5:00
Report, Code and presentation submission:	Monday, 28 April, 5:00 pm
Peer review completion:	Friday, 2 May, 5:00 pm