

ETC3555 - Project 2018

Groups

The groups and time of presentation are given below.

Oct 16, Tuesday, 9:00am, CL_20 Chancellors Walk, Room E365, East Level 3 (Bldg 11)

1. 09:00 am - Group 1: PAUL GORDON HENDY and JACK GRAHAM DAVIES
2. 09:12 am - Group 2: ELIZAVETA MAKSIMOVNA KOSHENKO and ADYE ROHAN DOUGLAS
3. 09:24 am - Group 3: JIANXIANG ZHENG and KAMALPREET SINGH
4. 09:36 am - Group 4: WILLIAM CHAN and JENNIFER HE

Oct 17, Wednesday, 10:00am, CL_20 Chancellors Walk, Room E365, East Level 3 (Bldg 11)

5. 10:00 am - Group 5: XIN QIAN ENG and HONG XIANG YUE
6. 10:12 am - Group 6: HAN YANG LIM and MITCHELL RYAN ONG-THOMSON
7. 10:24 am - Group 7: ZEZHENH ZHANG and JIAYIN TANG
8. 10:36 am - Group 8: MICHAEL SEN JIE CHAN and BENJAMIN GORMLY CRAINE

Task

You will choose a machine learning model used to solve any learning problem (regression, classification, clustering, etc) for any type of data (table, time series, network, etc). You will read some of the relevant references and compare your algorithm to at least one alternative on a dataset that you have chosen. Finally, you will write a short report (5 pages) describing your model, explaining the experiment you have performed and summarizing the key results.

Each team will make a presentation of their work for the class (max 10 slides, in PDF format). Each team will have 12 minutes (10 minutes for presentation, and 2 minutes for Q & A). Each member of the team should participate by speaking in the presentation for 5 minutes.

You can choose a topic that has been covered in ETC3250 or ETC3555, but you must present new results on that topic. Bonus points will be given for originality, including if you your topic has not been covered in ETC3250 or ETC3555. You will provide the code used for your experiments including comments. It should run without errors. We would like to remind the students that plagiarism will be taken very seriously. See the Monash policy on plagiarism at <https://www.monash.edu/students/academic/policies/academic-integrity>.

Project report and presentation

The data analysis report can be a maximum of 5 pages, and must abide by the section structure described below.

- Section 1: Model

This section presents the learning problem and model you have considered. Some of the questions you could answer include “what are the parameters/hyperparameters?”, “how do you optimize these parameters?”, “explain how these parameters control the complexity/flexibility of the algorithm”, “does your method scale well with large data sets?”, etc.

- Section 2: Experiment

This section describes the data set you have chosen and the experiment you have performed, including a justification of your choices (evaluation metric, optimization procedure, etc). You should also present your model fitting, diagnostics, etc.

- Section 3: Results and Discussion

This includes for example graphs and tables, as well as a discussion of the results.

- Section 4: Conclusion

This includes summary of the findings.

Grading

- Total points: 20
- Report: 13
- Presentation: 7

For your report, you should clearly explain what you have done, using figures to supplement your explanation. Your figures must be of proper size with labeled, readable axes. In general, you should take pride in making your report readable and clear. You will be graded both on *stastical/machine learning content* and *quality of presentation*.

For your presentation, score will be given by the lecturer and other students. All students must be present to evaluate the presentations, and if not points will be deducted from the absent individual's score. Points will also be deducted if the evaluation of presentations is not done seriously.

Deadlines

Do not wait until the last minute. Late submissions will not be accepted.

- September 14, 11:55pm: You have discussed your topic with the lecturer.
- October 14, 11:55pm: Upload to Moodle (i) your project report, one per group, and (ii) the slides for your presentation.
- October 16 and 17: Present your slides in the lecture period.