

# Project 2 Sentiment Analysis

COMP4901K and MATH 4824B  
Fall 2018

## Notes

- Kaggle submission due: October 31 at 23:59.
- Report and code due: November 7 at 23:59. **No late submission is accepted.**

## 1 Content

Project 2 is related to lab 4 and lab 5. The task is the same as lab 4, which is sentiment classification of a document. With a lot of training data, you can train a classifier. Then given a new coming document, the classifier should be able to predict the sentiment of the text ranging from 1 to 5. We have introduced how to build a naive Bayes classifier in lab 4. You can use that classifier. After the submission of that classifier, you will be able to get 60 points of the classifier. Then you can think about how to improve, e.g., changing features, tuning hyper-parameters, and changing classifiers, etc.

The following items provide you some guidelines of the project:

- You should join the in-class kaggle competition and evaluate your models online, which has been introduced in lab 5. The display name of kaggle website should be your **student id**.
- We placed several baselines on the leaderboard. If you can reach any of the baseline, you can get attached scores after submitting your code and report.
- You're supposed to finish this project on your own. Plagiarism and team work is not allowed.
- You can only make 5 submission every day.
- You can use any programming language you like and any third-party libraries.

## 2 Submission

You need to submit two files, code used in competition and report to briefly describe your algorithm.

In the report, you need to include the following points:

- Your name, student\_id, your final rank and scores on leaderboard and the display name of your Kaggle account. (10%)
- What algorithms are you using in this project? (30%)
- How do you conduct parameter tuning? Is there any difference between your local validation and online results? (30%)
- How to run your code? Which third-party libraries are you using? (30%)

### 3 Grading Rubrics

We will follow the following grading rubrics for the final grade of this project.

Grade	Classifier (80%)	Report (15%)	Code (5%)
60%	baseline 1	Submission of the report	Submission of the code
80%	baseline 2	Showing algorithms you used	
90%	baseline 3	Detailed explanation of the algorithms	
100%	baseline 4	Very detailed and insightful analysis	

Table 1: Grading Rubrics.

The 15% of report weight will be applied as a weight for each of the four items shown in Section 2.