

Team name: Outliers

Team members: Yvonne Qiu, Zoey Ye, Jiahuan He, Zenan Wang, Taixi Lu

Date: 4/6/2022

This journal has to be uploaded to 1) a dropbox folder ([link](#)) **AND** 2) posted to your project website every two weeks by 11:59 pm on the second Friday. The first due date is Mar. 4th, Mar. 25th and the third due date is April 7th. When uploading to the dropbox folder, change the file name to "Team_name_MM-DD-2022" where MM = month and DD= day of upload).

Team roles for this report (write down name):

Facilitator(s): Zenan Wang Taixi Lu

Recorder(s): Zoey Ye

Deliverer(s): Jiahuan He

Planner(s): Yvonne Qiu

See last page for description of roles. Obviously one person can take more than one role or there can be more than one person per role or make your own roles!

- 0. Describe briefly what the main goal of your team is (so the peer reviewer has some context). E.g. we are working on image classification for blah de blah. Our goal is blah de blah etc. In the initial part of the semester before your proposal it is ok to put down "we are still coming up with ideas on team project".**

We have gathered tweets data pulled from Twitter, and our main goal is to use the existing tweet posts, date, and location information during the COVID pandemic to predict the risk level of that area and the sentiment of the person who sent those tweets. This is important since we are all under tremendous pressure under COVID, and mental health has become a big problem. We are going to use five models: BERT, logistic regression, KNN, Bayes Naive, Decision tree, and compare results from these models, and assess the difficulty of implementing these models.

- I. **What was done during the report period regarding the project:** If you want to include code include this in the Appendix. Describe what the group did (including contributions of individual team members) with regards to the group project during this report period. Give enough details so I understand what you folks have been doing over the week. Include dates of your meeting(s) and who met on these days.

March 23rd — Taixi has wrapped up the work and explained detailed in the group chat.

April 2nd — the whole group has decided to work on at least one model: Yvonne works on the BERT model, Zoey works on logistic regression, Taixi works on Naïve Bayes, Jiahuan works on decision tree and Zenan works on KNN model.

April 6 — Yvonne, Taixi, Zoey, Jiahuan have gathered to finish up bi-weekly 3 and update the our progress.

April 8 — we are going to have a more detailed zoom meeting, every team member will join and present the code they have written and explain their thoughts.

- II. **What were obstacles faced if any in working on the project?** This could be technical (like not being able to implement or understand particular techniques) or time issues (midterms for other courses etc).
1. One of the obstacles we've encountered was the time issues. Since we have five team members, it was not easy to arrange an ideal time for the meeting. So we schedule each meeting and assign tasks for each member in advance, which maximized our efficiency.
 2. Finding the easiest and most time-saving way to process massive data is one technical problem we met.
 3. When I was trying to utilize the bag of words model to clean the text data for fitting Naive Bayes, a hard question was how to choose the dictionary. What kind of words has the most

valuable predictive information? Also, the implementation of the data structures needed is intensive.

4. We have to learn the whole new BERT model. Although we have slightly talked about the neural network during class, including professor Li Yao's talk, we did not actually learn the use of neural network. And NLP is actually the most important part, so we need to spend a lot of time to learn how to use the model into our project.

5. **What is the plan for the next reporting period including what each team member is planning to work on. Describe goals and potential timelines (" I plan to finish understanding x to see if it can be implemented for our project by Wednesday etc".)**

We plan to finish our models by Friday and prepare to discuss them on our Friday meeting (April 8, 2022). We need to spare some time to beautify our website and refine the homepage content. We will also need to combine all the team member's work together and start to work on some graphic representation. We will schedule the next meeting based on our progress.

While in the biweekly document above you will describe what your team did with regards to the team project (with proper attributions of who did what in the week) there are 4 pre-defined roles. I urge you to have different people do these jobs every week so that you gain experience in each of the jobs. There can also be more than one person per job for example 2 people recording the weekly journal.

Facilitator: Manages the group for this week including setting up times for group members to meet, making sure everyone has a say in the meetings etc.

Recorder: Person in charge of recording the meetings as well as the happenings of the past two weeks and describing what was accomplished in the meeting and writing up this report.

Deliverer: Person in charge of checking the entire report and uploading the file to dropbox folder and project website, as well as the representative of the group getting in touch with the instructor.

Planner: Person in charge of what will be happening next two weeks as well as thinking about longer term goals (what more needs to be done for the project).

Team contact: Person I can email if I see any issues in the biweekly report instead of mass spamming everyone in the team.