Team name: Outliers
Team members: Yvonne Qiu, Zoey Ye, Jiahuan He, Zenan Wang, Taixi Lu
Date: 03/24/2002

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 form 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.

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.

We met on Mar 5 and Mar 18 via Zoom. During the Mar 5th meeting, we finalized the our project proposal and EDA by combining everyone's work together: Jiahuan He was responsible for the description of the dataset; Yvonne Qiu was responsible for the description of our motivations and goals; Taixi Lu and Zenan Wang were responsible for the exploratory data analysis; and Zoey Ye was responsible for coming up with our preliminary idea of what techniques to use in the final report.

In the meeting on Mar 18th, we discussed about what kinds of models we will use for our data, and how to apply them. We decided to use logistic regression, LDA, QDA, KNN, and recurrent neural network for now. Each member will be responsible for applying one or two methods, and one member will be responsible to combine everyone's work together and probably visualize the results for the presentation.

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).

Our main obstacle is the data structure in the tweets data. The data is in natural language. Although we have known how to do LDA, QDA, KNN, and logistic regression on data that has many columns, it seems to be unwise to directly apply those methods in our case. Only one predictor that is a long sentence in English can give poor result. For now, our solution is to come up with new variables (e.g. how many characters in the tweet? How many of those are in uppercase? How many exclamation marks?) and apply ML methods on the enlarged data. We will still looking for better methods to do this. If you have any suggestions, please contact the site maintainer. We really appreciate it.

III. **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 processing the data (e.g. add more variables and extract information from the tweet) for applying LDA, QDA, KNN, and logistics regression this weekend so that we can start our prediction next week. During the process of analyzing the classification models, we will also mark the problems or confusion we have and ask for suggestions during office hours.

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.