

Team name: Anonymous

Team members:

Xinyue Xing, Boyu Liu,
Yingxi Huang, Tianyi Li

Date: 03/04/2022

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): Yingxi Huang

Recorder(s): Tianyi Li

Deliverer(s): Boyu Liu (Team Contact)

Planner(s): Xinyue Xing

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".**

Our dataset consists of 8675 rows and each row is a person's MBTI type (Myers Briggs Type Indicator, a well-constructed personality type system that classifies people into 16 different personality types), and the last 50 posts they had on a forum. We aim to construct a model to predict a person's MBTI type based on their posts.

- 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.

Meeting #1 – 02/23/2022 – All members attended

In our very first group meeting, we mainly shared our topics of interest with each other and searched for dataset together. We eventually came across 5 interesting datasets – Uber pickup address, MBTI text analysis, emotion in text, data scientist salary, Titanic survival prediction. We decided to further explore each of them after the meeting and make our final decision in the next week.

Meeting #2 – 03/02/2022 – All members attended

In this meeting, we first created our team website using Github. We each shared our opinions on the datasets chosen last week. Considering our common interests and limited ability, we finally chose a dataset consisting of the participants' MBTI type and the last 50 posts they had on the forum. We aimed to construct a model to predict a person's MBTI type based on their speech on the Internet.

- 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).

Since we were still new to the field of machine learning, it was difficult for us to properly evaluate the availability and difficulty of a potential project. Regarding some topics of interest such as image recognition, we were concerned whether we were experienced enough to work on such project. On

the other hand, we also had concerns about whether a dataset was large enough to construct a machine learning model.

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”.)

Next week:

We will mainly focus on working on exploratory data analysis and the proposal.

- Xinyue Xing: further explore the data and conduct EDA to collect information such as the distribution of data over the groups and finish data visualization by Tuesday
- Yingxi Huang: learn to deal with text data and do EDA on the text data to get information like the frequency of the words and the pattern of word distribution in each group by Tuesday
- Tianyi Li: read resources on NLP and learn to extract features from text data; list the plan of how to extract features in our project by Tuesday
- Boyu Liu: work on the introduction part of the proposal and set up the frame of the proposal by Tuesday

We are going to meet on Wednesday and discuss what each of us found and combined our findings in our proposal by Thursday

The week after next week: SPRING BREAK!

All of us will start to extract features on our own following the plan created last week. Hopefully we can make some progress on feature construction. We plan to meet on the Sunday of the week (last day of SPRING BREAK), discuss about our results, and together finish setting up the features.

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.