

Reflection Report on Natural Language Processing for Mental Health Risk Prediction

Team 13, The Cognitive Care Crew
Jessica Dawson
Michael Breau
Matthew Curtis
Benjamin Chinnery
Yaruo Tian

1 Changes in Response to Feedback

1.1 SRS and Hazard Analysis

SRS: Our SRS has undergone many changes since the revision 0 submission. Some were from feedback from the TA such as reformatting document structure to be more clear, creating more readable functional requirements with labelling, adding a reflection section which we missed for rev 0, etc... We also made many changes directly because of our expanding of scope and knowledge of natural language processing pipelines increasing so we had a better idea of the functional requirements needed as well as our division into the three different NLP tasks.

Hazard Analysis: For this report we utilized a lot of high quality feedback from our professor and TA. The first major concern we addressed from there feedback was making it explicit what was in scope for this project. There was also the point made that they thought we could come up with more hazards to identify and once we read this, we actually came up with a couple of different new hazards to add that we previously did not think of. Lastly there was the correction we added from the feedback about ensuring we have all terms we used defined and also following their advice to make our recommended action more detailed and less vague. We also implemented one main change based on group 8s feedback, specifically about SR1. We used there feedback to realize that we needed to add clarity to SR1 to include and explicitly mention the idea of not sharing data online or exposing it to third-party sites.

1.2 Design and Design Documentation

The main feedback that we received from supervisors is that we should modularize our components which we added into the design doc.

1.3 VnV Plan and Report

VnV Plan: Our VnV Plan was mostly changed to fit the feedback from the new VnV Report. VnV Report: A good piece of feedback we got from group 8 for our VnV Report was to add something into our automated testing section to mention how we used a linter and Github actions in our implementation because this would fall under this category.

2 Design Iteration (LO11)

2.1 Task 1

How we arrived at our final design and implementation was through a lot of research and talking with the team in Montreal, specifically Diego. Our first version was very bare bones and we had yet to determine any testing metrics or even create a system to test our product. Also we were originally between using a vectorization approach and bert topic modeling approach. After research however, attempts at implementing both approaches, talks with the Montreal team along with analyzing our testing data, we came to the conclusion that it would not be viable to use the bert topic approach. We realized the data results we were given to test on were not comprehensive enough and were not adequately labeled to be used to train data. This brought up to fully committing to our vectorization word embeddings approach! We also considered multiple different word vector databases to use which would act as our basis for which words the system could understand. We originally just the twitter glove database but in the end we decided to go with a crawl vector database. How we came to this was we tested our code with multiple different databases and compared how many words in our feature set were found in the databases and how accurate our results were. At the end of this, crawl ended up being the most encompassing database for our needs and also produced the best results.

2.2 Task 2

For our second natural language processing task related to the binary classification of anorexia our initial model was a simple iteration of grouping documents together and we decided that for the "training" the group with the most amount of positive users would be considered to be the group predicted to have anorexia. The initial model was not able to predict new users which was added later on. Another issue we ran into was that Bertopic has many different parameters but with a long runtime we were not able to test many of them but from the recommendation of one of our supervisors we used grid hyper parameter tuning which allowed us to determine the best parameters for many different fields to be tailored to the training data.

2.3 Task 3

I (Jessica) talk about this a lot in the rev1 version of the VnV Report, for the sake of not repeating what I said there I'll ask you to refer to that document. To summarize though, some literature review to see what directions were out there and constantly consulting with the supervision team to see if the approach I was taking was on the right track. The one interesting bit not included in the VnV Report is how I started to use classes and their state behaviour to cache data in order to speed up reruns of models for stuff like parameter tuning. I realized they could be used for this and began implementing this in a lot of my code to help reduce run times.

3 Design Decisions (LO12)

3.1 Task 1

b Our first large design decision was whether to use the BERT approach or a vectorization word embedding approach. When considering this choice, we had to assess both methods and decide which we deemed more suitable for our project. We had to consider feasibility, time constraints, and the limited resources we had available. We ended up deeming that the BERT topic model may be the best option but we had limited data to use to test and with that, not adequately labeled data either for training purposes. As a result we deemed this strategy as not feasible. We considered trying to label the data to allow training ourselves but we could not ensure accurate labeling and with our time constraint of the capstone this was not feasible either. This left us to go with the vectorization option which fit our needs and abilities perfectly. We were also faced with the design decision of which word vector database to use for our word embeddings. This was an important decision because some words we wanted to use are not in every database. To come to our design decision, the main factor was how encompassing the database was for the words we were looking for (as few words missing in the database as possible) along with the biggest factor which was what were our accuracy test results with each database. This led us to choose crawl for our database.

3.2 Task 2

The primary limitations and constraints influencing the design of this task stemmed from the rules and guidelines of the competition. It was important for the team to try to find the best balance between achieving the highest accuracy possible, but also keeping a good handle on length of operation, resource allocation, and of course repeatability in order to optimize our performance scores. Although the team experimented with various models which boasted all sorts of situational expertise's, the team felt that the BERTopic model was able to strike the best balance between our allocated priorities. A major component of this task is that the model is meant to take information in sequential chunks

in order to imitate real time posting behaviours, what made BERTopic stand out was its ability to update decisions on its users on the fly, without having to retrain its database every time, cutting down on operation time as well as resource usage.

3.3 Task 3

I think I would have liked to plan more of the design process before getting into coding as there was a fair amount of spaghetti going on in my code as a result, I think part of that can be attributed to coding in an area that I started out having very little knowledge of. I think the model I built was interesting, it built upon prior work done by teams at the competition and did something new. I think one of the biggest limitations I ran into was not having a foundation of knowledge in this area, it was an interesting challenge trying to develop an approach and system while still learning how everything NLP fit together, it resulted in going down a number of deadends and some floundering. I think in the future I would like to learn how to better approach this sort of learning process.

4 Economic Considerations (LO23)

The target market for an Natural Language Processing system capable of detecting depression, anorexia and eating disorders is large, namely due to the growing acknowledgement and importance of the mental health problems. To market this product, targeting healthcare professionals, mental health organizations, clinics staff and/or individuals or relatives who seek support would be carried out. The cost factors of a sell-able version would include development, testing and on-going maintenance, which could fall into the moderate to more significant category depending on the complexity and scale of the system. The pricing, most likely would depend on the target market and the value promised. If we plan to further maintain and develop this product into something that could actually be used by real life users, then we could charge a monthly or yearly subscription fee per user registered.

To make a profit of units sold should be larger than the total production and marketing costs. If the project uses the open-source model and would not to be sold to the public, attracting users along the help of community engagement, collaborations with mental health experts, and making the system available for researchers and professionals will be possible. This audience is potentially the most big and diversified because it comprises healthcare providers, people with mental health disorders, researchers, and educators.

5 Reflection on Project Management (LO24)

5.1 How Does Your Project Management Compare to Your Development Plan

After completing the project, our team now understands how lacking in scope we were at the start of the project. Originally we were just planning on creating this one pre diction model but we ended up expanding our project to cover three different natural language processing tasks. In turn a lot of our development plan had to be updated to reflect these changes. For example our team roles had to be changed to suit the expansion of three tasks so we had to all learn the entire natural language processing workflow instead of having members specialize in single skill sets. Overall our team did follow the communication plan well which did not need any changes. The technology we planned on using did also expand quite a bit since we learned of new methods over the course of the project and were constantly trying many different technologies which would have been impossible to predict at the start of the project.

5.2 What Went Well?

In terms of what went well for project in terms of processes and technology was that the team was able to achieve our primary goal of improving upon the prior metric results achieved by past year entries of the research competition.

5.3 What Went Wrong?

We needed to be using a lot more tools for tracking issues and where everyone was at in terms of knowledge. It could've improved team cohesion a lot and helped ensure everyone was on the same page.

5.4 What Would you Do Differently Next Time?

If given the opportunity to work on a similar project which would require the coordination of 2 independent groups, we feel that the best course of action would be to be more strict with our merging timelines in order to foster better results for everyone. Earlier stages of the project's development lost valuable time when there weren't clear boundaries and roles of how the teams would interact. Both teams were so focused on managing their immediate responsibilities that less emphasis was put on the bigger picture, so that when the time came for the systems to start integration both teams realized that valuable time and effort was misplaced either doing overlapped responsibilities, or creating systems that would require some reworking in order to work within the other team's design constraints. Overall, better communication would lead to stronger team synergy, and less misplaced effort that could be better directed at making a more efficient project.