Slide Annotations

Topic modeling

- I chose NMF to create my topics because I like the assumptions it makes about a word or document not having a negative relationship to a particular topic. Especially in the context of mental health where words like depression are comorbid with topics like anxiety or panic. Here are some examples of post belonging to different topics.
- I think it's helpful to contextualize some of the topic names particularly because topics are more of a general theme and a post belonging to suicide may have mentions of family or panic/anxiety. QUE

Understanding the Topics

- So here we have the top words for few topics of interest. Personally I find the language of these topics very interesting particularly some of the quarantine related words in the school topic.
- A common theme in these topics is a desire for advice and support QUE ARROWS.

Providing Support

- So how do we provide support…So one way is to examine trends over time.

School related Posts Graphs

- Here we can see some periodic patterns in post related to school see upward trends along these months.
- Assuming some periodicity with our data we can anticipate trends in the upcoming months for particular topics and potentially allocate support for these topics accordingly.

Topic Volume by hour graph

- We can also look at time of day.
- So, most posts/topics follow a general trend with a higher volume of posts in later evening and early morning hours QUE. due to the hours, one thing to consider is that there is likely not enough support is being provided during these times.
- Another way to provide support is by better understanding the ideas and language behind these posts

Understanding the language.

- After topic modeling I was interested in other approaches. I thought word2vec could help gain insight into the language in my data.
- By comparing a model trained on r/depression to Google's pertained model we can see some linguistic differences as well as some overlaps.

Google's model (similarity to 'self') graph

- The intuition behind picking the word self was because I had selected it to be a stop word, it's also highly context dependent so I thought the comparison could be interesting. With googles model you can see its associated with/closest to equally abstract and innocuous words, that are more inline with ordinary expressions.

Our model (similarity to ‘self’) graph

- Whereas when we look at the expression with our model we can see its associating with less ordinary compound expressions.

Word Similarities “Cycle”

- Cycle was another interesting word because it illustrates overlap with our model. but also some connotative differences…

- With googles model you can see the words have a more positive connotation and also seem more economically driven, whereas ours is associated with words that are more descriptive of an emotional state. QUE

Exploring further

- Overall I think there are relatively simple things that can be done to provide support but I don’t think enough emphasis is put on understanding the language around mental health. That being said I’ve been exploring this visualization tool with r/depression data and comparing it to models trained on very different datasets to find similarities and differences between ordinary expressions in both.