# Mental Wellness Sentiment Analysis

Jiaan Nie, 200327281

## SUMMARY

The purpose of the project is try to determine the current mental wellbeing of a person through input of their text using technique call RNN (Recurrent Neural Network). Initial plan is to create a small prototype that will work on 2 or 3 categories such as depression, anxiety and suicidal thoughts using labeled data that scraped off from websites and clean up the data using panda, then implement a basic RNN model using Tensorflow and test out the result.

## REQUIREMENT

Hardware Requirements:

- i7 CPU
- GTX 1080 CPU

Software Requirement :

- Anaconda
- Tensorflow
- Panda
- PRAW
- WordCloud

## METHODS

First scrape labled data from reddit , use python to lablel and organizes the data and parses into to panda.

Split the data set into:

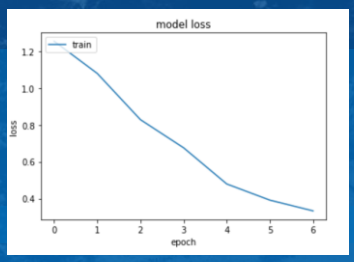
- 2/3 tranning
- 1/3 validation

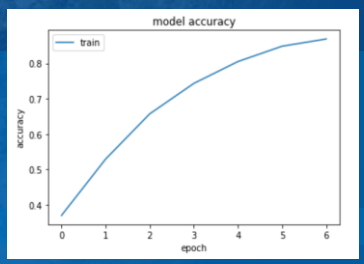
Finally feed training data into the LSTM model and use the validation data set to validate.

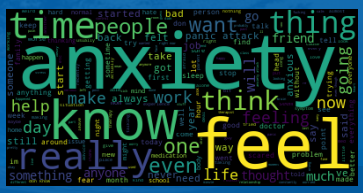
## IMPROVEMENTS

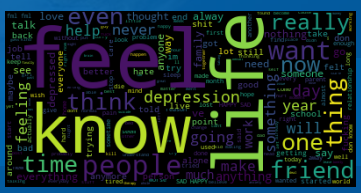
- Tuning Hyper Parameters to improve accuracy.
- Data Sanitization such as fix spelling, remove special characters , emoji etc.
- More labeled data and better quality data
- Testing on different activation functions

## RESULTS

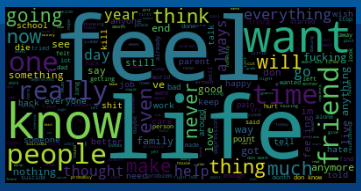







Model: "sequential_6"

| Layer (type) | Output Shape | Param # |
|---|---|---|
| embedding_2 (Embedding) | (None, 3149, 128) | 256000 |
| dropout (Dropout) | (None, 3149, 128) | 0 |
| lstm (LSTM) | (None, 196) | 254800 |
| dense (Dense) | (None, 4) | 788 |

Total params: 511,588
Trainable params: 511,588
Non-trainable params: 0

```
888/1 - 2s - loss: 1.2531 - accuracy: 0.5709
score: 1.26
acc: 0.57
```

## CONCLUSION

Overall the accuracy of the model is not ideal, to improve on the model it will requires better and more labeled data. Sanitizes the data before feeding into the model reduce the noise. The model it self seem having trouble classify between depression and suicide posts due the similarity of contents.