# Progress report

**Overview plan for the project:**

The first step of the project is setting up an environment where it has all the basic developer tools required for the project. The tools that has been chosen are Python 3.7+ and Tensorflow. The next requirement in order to move the project forward are data, more specifically labeled data, due to the topic of the project is very unlikely to obtain clinical data due to privacy concerns, hence it would has to be open data sets. After some research there are two sources where the project can obtain labeled data from, Reddit and <https://www.supportgroups.com> with the total of 18,000 labeled data.

To start it off scrape small amount of data off of Reddit which is about 3,000 samples, then find different deep learning models via online or Github apply the sample data into the models just to test out the result see what we get. Next learn the fundamentals of deep learning via by taking online courses, if it’s possible make changes to the code try to improve the model with the data.

**Completed Tasks:**

1. Scraping small sample of labeled data online, in this task a Python library call PRAW were used in order to scrape data off of Reddit roughly about 3000 sample which contain 4 different categories such as happy, anxiety, depression and suicidal.
2. Organizing the data via another Python library call PANDAS this library allow users to manipulate their data in to a table format so is easy to read and pip into Tensorflow for training the model.
3. Finding different deep learning models via online or Github and apply the sample into the model and see what kind of result you will get. Currently have try it on LSTM (Long Short Term Memory) model, however it seem like the model is not convergent that mean it seem like it stuck in training forever.

**To-do List:**

1. Going to try it out a basic RNN on the sample data and see what kind of result will generate
2. Learning the fundamentals of deep learning fundamentals via taking online courses by Andrew NG
3. If I able get a prototype working, and the project required more data the next option for scraping labeled data is from <https://www.supportgroups.com/> which contain around 15,000 labeled data, however is a bit harder to scrape the data.