

Mental Health Data Science Project

Group Members

Yoni Levine and Farhad Zmarai

- Scraping - Yoni
- Cleaning - Farhad and Yoni
- NLP - Farhad and Yoni
- Modeling - Farhad
- Visualizations - Yoni
- Report - Farhad and Yoni

Project Plan

The idea is to study [takethislife.com](https://www.takethislife.com), a forum dedicated to support of users dealing with depression, and help understand what factors are associated with suicidal ideation.

The website is split into two separate forums, one for depression and one for suicidal ideation. The objective is to see if we can use the text and usernames from the depression forum to predict which users will also post in the forum for suicidal ideation.

We are viewing this as an NLP/classification problem and are considering logistic regression, tree based methods, support vector classifiers, and a Bayesian classifier.

Dataset

The raw dataset will be the post text and usernames, along with other features of interest, scraped from the depression section of the website with a label corresponding to whether or not that user had also posted in the suicide forum.

Due to the lack of mental health resources available, the internet has become a popular venue for people to reach out for help. All this public data provides us with an opportunity to study this population and try to understand what factors may be contributing to the rising suicide rates around the world.

Some specific challenges that we may face with this dataset are as follows:

1. Data science projects often rely on the advice of subject matter experts, this is particularly true in healthcare. We may not have this, but I still think there will be interesting findings from our analysis.

2. Due to the informal nature of forums many words will be misspelled or abbreviated, which may weaken the results of our model.