Project Proposal
DS 5230
Project Group 5

# Detecting Signs of Depression amongst social media users

**Authors**: Kavya Gajjar, Parth Bhatt

## Summary:

The world is facing a plethora of mental disorders, majorly depression. Most of the population is active on social media, and if we can detect depression in early stage, it would be helping the individual getting right support on time. In this project, we are looking to detect the signs of depressions by performing sentiment analysis on social media text. We will be utilizing the twitter data for the same purpose.

## Dataset:

We have decided to utilize the twitter sentiment tweets dataset from Kaggle. The dataset contains 1.6M tweets extracted using the twitter API. The tweets have been annotated (0 = negative, 4 = positive) and they can be used to detect sentiment.

The dataset contains the following 6 fields:

- *target:* the polarity of the tweet (0 = negative, 2 = neutral, 4 = positive)
- *ids:* The id of the tweet (2087)
- *date:* the date of the tweet (Sat May 16 23:58:44 UTC 2009)
- *flag:* The query (lyx). If there is no query, then this value is NO_QUERY.
- *user:* the user that tweeted.
- *text:* the text of the tweet (Lyx is cool)

Currently, we are planning to use the tweets with target values of 0-3 (i.e., removing positive tweets). The mentioned dataset contains the tweets for the year 2009. If we want to utilize the tweets from the recent years, we might have to web scrape the tweets using tweepy/twint.

## Proposed plan:

Perform EDA on the dataset and treat the missing values and outliers to extract a robust and well split dataset for analysis. Preprocess and clean the tweet text, transform into vectors and feed the same into the model. Currently, we are planning to compare the efficacy of two algorithms: LDA and SVM in detecting the sentiment of the tweets fed into the models and comment on the shortcomings/pros of the same.

## References:

- https://www.kaggle.com/kazanova/sentiment140
- https://en.wikipedia.org/wiki/Latent_Dirichlet_allocation
- https://en.wikipedia.org/wiki/Support-vector_machine
- https://www.tweepy.org/
- https://github.com/twintproject/twint