**Project 4 Proposal**

Group name: 3.14

Group members: Dan T Do, Daniel Herfurth, Felix Poley, Lucas Vianna

1. The problem we will be working to solve is training an ML algorithm to recognize and differentiate between real and fake news stories.
2. We will use Sklearn to process and categorize the data. Data will be visualized with Seaborn and Plotly, and we may use other visualization libraries if need be.
3. Sklearn’s passive-aggressive classifier will be the main classification method to categorize the news. Tfidfvectorizer will be used to process the natural language sentiments.
4. Pandas and Matplotlib will be used initially to structure the information and output the plots. We plan to add at least one other library from the list in the future.

The real world implications of misinformation in the press are far reaching and ever present in our society. In order to fight the oncoming tide of misinformation, it is necessary to first be able to identify and sort out fake news with leading, biased language from neutral sources of information. Once a model can be trained to do this effectively, one no longer needs to rely on human interpreters to identify misinformation, thus drastically increasing the speed at which articles can be classified.

Once the model is classified, it then can be used to identify and warn users about fake news on sites like Twitter, Facebook, Reddit, and others.

Date set:

<https://www.kaggle.com/c/fake-news/data?select=train.csv>

<https://www.kaggle.com/c/fake-news/data?select=test.csv>