**Detection of Fake News**

Team member: Saide Tang, Anyang Lu

**Objective**

The fake news is a made-up story with an intention to deceive and mislead readers. The fake news has a significant impact on our social life. To understand the features of fake news and differentiate the fake news from true news could improve readers’ judgement on the accuracy of on-line information. The object of this project is to apply the techniques of natural language processing to detect the fake news.

**Data Description**

Kaggle collects two datasets including fake news and true news. Every dataset gives the information of the text of the news article, the title of the news article, and the subject of an article and the date of the article was posted. We use 4716 true news articles and 2706 fake new articles focus only on political news articles, because these are currently the main target of spammers. The news articles from both fake and true categories happened in the same timeline, specifically in 2016. Each of the articles length is bigger than 200 characters.

<https://www.kaggle.com/clmentbisaillon/fake-and-real-news-dataset>

**Methodology**

Based on the datasets, there are “title” and “text”. We plan to build models with each of them. For the type of input, we would try both TF-IDF and embedding methods. For the model building, we plan to start with simpler methods, including Naive Bayes, N-Grams, Logistic Regression, and a simpler neural network architecture. We would also try to use the LSTM method with more complex architecture. We will finally compare the accuracies of different methods, and provide the explanation for the differences.

**Reference**

Ahmed H, Traore I, Saad S. “Detection of Online Fake News Using N-Gram

Analysis and Machine Learning Techniques”, [International Conference on Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments](https://link.springer.com/conference/isddc)-ISDDC 2017,127-138.