**Predicting which Tweets are about real disasters and which are not.**

During the age of Social Media where we get all the updates and News from social media like Twitter, Facebook, it is very important to differentiate real tweets and fake tweets. We have about 10,000 tweets and are trying to solve which tweets are about real disasters and which one’s are not. The problem we are trying to solve will be useful to flag fake tweets that are not about real disasters. We will be using python regular expressions, Natural Language processing techniques like bag of words, tfidf, word2vec, sense2vec. Supervised learning like logistic regression, Naïve Bayes classifier, SVM, gradient boosting. Unsupervised NLP. Dimensionality reduction techniques like Principal Component Analysis, Singular Value Decomposition. Various visualization techniques like heatmaps, wordcloud. Biggest challenge to solve this problem is converting the data from its original form to the language that computer understands. The dataset is from a Kaggle competition and below is the link for the dataset.

**https://www.kaggle.com/c/nlp-getting-started/data?select=train.csv**