Project Proposal: Text Classification Competition: Twitter Sarcasm Detection

1. **What are the names and NetIDs of all your team members? Who is the captain? The captain will have more administrative duties than team members.**

Name	NetID	
Saroj Khanal	khanal2	Captain
Kevin Choi	gchoi17	
Jiayi Chen	jiayic15	

2. Which competition do you plan to join?

Text Classification Competition: Twitter Sarcasm Detection

3. If you choose the classification competition, are you prepared to learn state-of-the-art neural network classifiers? Name some neural classifiers and deep learning frameworks that you may have heard of. Describe any relevant prior experience with such methods.

Some of the neural classifiers are : ANN (Artificial Neural Network), RNN(Recurrent Neural Network), LSTM (Long Short Term Memory).

ANN is a feed forward neural network where data pass forward from input to output. On the other hand, RNN has feedback loops in the recurrent layer which enables maintaining information in 'memory' over time. However, it can be difficult to train standard RNNs to solve problems that require learning long-term temporal dependencies because the gradient of the loss function decays exponentially with time (called the vanishing gradient problem).

To look into this issue further, LSTM networks are a type of RNN that uses special units in addition to standard units.LSTM units include a 'memory cell' that can maintain information in memory for long periods of time.There are three gates for LSTM: an input gate, an output gate and a forget gate. Gates are used to control when information enters the memory, when it's output, and when it's forgotten.This architecture lets them learn longer-term dependencies. LSTM assigned relatively more important weights on units with longer term to improve general accuracy.

Two of our team members have experiences with Neural Networks and have worked on personal projects to some extent. However, for this project, our team needs to do some research, learn more about the above mentioned classifiers to properly implement to increase the accuracy of the model.

However, before feeding the data into Neural Network, we will need to perform several text processing that we learn on the earlier part of the class such as tokenization, change words to lowercase, remove numerical data, remove stopwords, stemming, lemmatization, vectorization (Bag of Words), TF-IDF, use Word2Vec and etc.

4. Which programming language do you plan to use?

Python (Tensorflow library, Keras API, Gensim, NLTK library, TextBlob, spaCy).