**CS 5316 – Natural Language Processing**

**Assignment 4**

Deadline:

In this assignment, you would develop a deep learning models for paraphrase detection. Two documents are considered paraphrases when they convey the same meaning even though they may be written differently. Supervised paraphrase detection takes as input the two documents and outputs a label of 1 or 0 depending on whether they are paraphrases or not.

Download the Microsoft Research Paraphrase (MSRP) corpus and study its description. Represent each document in the pair as a sequence of word embeddings. Limit the document length to the first 20 words. If a document is longer pad with zeros, and if a document is shorter truncate it. Use GloVe embeddings of size 300.

Train the following deep learning models on the training set and give performance results (accuracy, precision, recall, and F1-score) on the test set. Use Keras as front end with Tensor Flow on the back.

1. Feedforward network with two hidden layers (300 and 150 units). Use ReLU activations for the hidden layers.
2. LSTM network with 300 units followed by a hidden layer containing 150 ReLU units.
3. Same as 2, but use a bi directional LSTM network

Present your results with default settings and with some variations. Discus the impact of the variations you consider.

Submit your report and code.