CS 410 Text Information Systems Course Project Proposal

Team Name:

Virginia DS

Team Members:

Huanzhen Hu NetID: [hh21@illinois.edu](mailto:hh21@illinois.edu) Role: Leader

Ying Zhang NetID: [ying12@illinois.edu](mailto:ying12@illinois.edu) Role: Member

Topic Selected:

Text Classification Competition

Prior Knowledge and Experiences:

We will use Python as the programming language for this project. To achieve the goal, we have collected some information of the cut-of-the-edge deep learning models which already reached SOTA results in previous research, including CNNs, LSTM, BERT and PRADO. We also prepared some tutorials about some popular deep learning frames, including PyTorch and Keras. We already had some practical experience in constructing shadow neutral network classifiers based on Keras framework. Therefore, the main challenges in this project are listed below:

i. Learn new advanced deep learning NLP models, including the basic ideas and improvements.

ii. Deploy those models on provided dataset. Finetune the model if necessary.

Timeline:

Present - Week 10: Clean the data. Study relative papers and articles to get familiar with the theory of classifiers and how to deploy them.

Week 11: Try traditional machine learning techniques with scikit-learn text classification packages.

Week 12 – Week 13: Deploy deep learning models on dataset and finetune the model.

Week 14: Pick out the best model. Write the report and complete project submission.