Final Project Progress Report

Stock Market Tweet Analysis

CS410 – Text Information System

Kasam Dhakal (kdhakal2@illinois.edu)

Nisarg Mistry (nmistry2@illinois.edu)

Parth Shukla (pshukl21@illinois.edu)

Progress report (5%): Each project team is required to submit a short progress report by the **beginning of Week 13 (Nov 14, 2022)** listing the progress made, remaining tasks, any challenges/issues being faced, etc. It will be graded based on completion with respect to addressing the following points: (1) Progress made thus far, (2) Remaining tasks, (3) Any challenges/issues being faced.

Progress made thus far:

- Familiar with the tools (Tweetpy and Natural Language Toolkit)

- Retrieved, cleaned, and explored Twitter data from Kaggle: https://www.kaggle.com/datasets/utkarshxy/stock-markettweets-lexicon-data for Sentiment Analysis.

- Dataset Tokenization

- Created a Baseline Model to generate a dataset by collecting tweets from Tweetpy.

- Analyzed Twitter data using your sentiment analysis model

Remaining tasks:

- Transform the data using TF-IDF Vectorizer.

- Function for Model Evaluation

- Parameter tuning/improving over baseline

- Testing and Improvements.

- Visualize the results of the Twitter sentiment analysis

- Project Code Submission, Documentation and Presentation.

Any challenges/issues being faced:

- The biggest challenges/issue for the sentiment analysis are to examine the large amount of data sets.