Yixing Zheng (yixingz3@illinois.edu) - team captain
Zerui Tian (zeruit2@illinois.edu)

Text Mining on Financial Text for Predicting Stock Prices Status Report

- [x] ~~Finding and processing the dataset - 2 hours~~
- [x] ~~Initial research, brainstorming and discussing which algorithm to use - 2 hours per person~~
- [x] ~~Dividing the task and setting up the environment - 2 hours per person~~
- [ ] Working on the algorithm, including training initial smaller dataset, evaluating model with larger dataset, training with larger dataset then evaluating the model with larger dataset, and documenting the code. We will try multiple machine learning and deep learning models and compare the results of different models. - 17 hours per person
  - [x] ~~Training dataset with one potential model~~
  - [ ] Training dataset with more models and fine-tuning the parameters
  - [ ] Documenting the code
  - [ ] Comparing results and building results charts
- [ ] Administrative tasks, including setting up CMT and GitHub for submission, drafting the project proposal, scheduling meetings for progress updates, drafting progress reports. - 4 hours for team captain
  - [x] ~~Project proposal~~
  - [x] ~~Project status Report~~
  - [x] ~~Peer review other projects~~
  - [ ] Schedule meetings and communicating - continues throughout the project
  - [ ] Managing/preparing codes and slides for final submission
- [ ] Preparing the demo, including the slides and presentation. - 1 hour per person

Overall progress and challenges if any:

- We've completed all the initial work and are working on the algorithm. So far, things are going well, and we just need to spend more time on trying different models and parameters.
- Because we need to try different models, we need to have some understanding of different models. For the transformer-based model, because we have not been exposed to this related content before, it takes some time to learn. And since the transformer-based model has a very large architecture, it is computationally expensive and requires a lot of time to train. This is what we need to consider in the implementation process later.