GROUP PROPOSAL

DATS 6312 – Natural Language Processing
Group 6: Sai Abhishree Pappusetty, Ketan Kapse

We have decided to work on the financial phrasebank dataset, which is a polar sentiment dataset of sentences from financial news. It consists of 4840 English sentences from financial news, all categorized by sentiment determined by multiple human annotators. Financial sentiment analysis differs from general sentiment analysis not only in domain, but also the purpose. The main purpose behind financial sentiment analysis is to guess how the markets will react to the information presented in the text.

Currently, we are planning to utilize Pretrained NLP (Transformers Based Networks) on this dataset, however Recurrent Networks, specifically classical ML could also be used instead. We are planning to use Keras and SKlearn due to their support for the above models, and matplotlib for plotting purposes.

Some of the metrics that may be used in this project include Accuracy, confusion matrices and F1-scores.

The following is a rough schedule for this project:

Task	Date
Analyzing and Data Preprocessing	11/12/2023
Transformer Implementation	11/19/2023
Classical ML Implementation	11/26/2023
Fine-tuning (if required)	12/03/2023
Report and Presentation	12/11/2023