Zach Horton, Virginia Maguire, Tanner Street, Yutong Zhang

**Problem Statement**

Many day traders do not have the skillset to build sophisticated models to help them inform their decisions. The goal of this project is to forecast daily closing stock prices and volatilities using deep learning models and create a chat bot that allows the day trader to interface with the results and receive strategic recommendations.

**Approaches**

We plan to use LSTMs to forecast daily closing prices. The model will make next-day forecasts for each stock after the market closes. We plan to train the model on data up to the end of 2023 and then validate the model with 2024 data. Specifically, for each day in 2024, we will first make a forecast, then integrate that training point into the model before moving to the next day.

We also plan to predict stock volatility using the GARCH model. Although this model does not involve deep learning, it would greatly benefit the chat bot to have volatility information, as this greatly assists in informing strategy. This requires feature engineering, as the closing price needs to be transformed into log returns. To avoid the project’s scope being too large, we will not spend a large amount of time fine-tuning the GARCH models.

Lastly, we plan to use an LLM to create a chatbot the day trader can interact with. We plan to use either GPT4, Gemini, or Llama2, depending on which ones are easier to use for this specific application. We possibly will need to use RAG for the context of the problem and limit the types of responses the chatbot gives.

A stretch goal of this project is to automate this entire process so the model is ready at 5 pm every day for new information input.

**Dataset**

We will use the API to pull stock data from Yahoo Finance. To limit the scope, we are only interested in a time frame starting January 2022 and stocks in the S&P 500. This data will also be used for the volatility predictions after we do additional feature engineering for this model.

The dataset we feed into the LLM will be an output of the LSTM that contains the following columns: the current price, the predicted price, the predicted change in price, the current volatility, the predicted volatility, and the accuracy of the model for that stock over the last 7 days (mean square error and mean directional accuracy).

**Deliverables**

At the end of this model, we will produce:

* LSTMs that predicts closing price
* GARCH model that predicts volatility
* A chatbot/interface to provide information and portfolio recommendations
* (Stretch) Automated process to update Yahoo Finance API pull and models daily.