MVP for Group 26

Team Name: Group 26
Team Members (NetIDs): mgome33, msun55, rahulm3, rishaj2, shijiea2, smoylan2
Team PM (NetIDs): andrejz2

Name your MVP "Group 26 MVP." You are allowed to use bullet points!

What problem is your project trying to solve?
(What audience is it for? How will your project solve the issue?)
The financial markets are complex and dynamic, and traditional trading strategies often struggle to adapt to rapidly changing market conditions. Our group project aims to use text and financial data related to Alphabet stock to predict movements in the stock's price. The goal is to outperform traditional strategies and achieve consistent, profitable results by simulating trading Alphabet stock in various market conditions.

Our project will solve the issue by first gathering relevant data such as daily/hourly stock prices (opening, close, low/high), text related to Alphabet stock (news articles, tweets, official company announcements, regulatory filings, etc), economic indicators (interest rate, employment, economic growth, consumer spending, etc), and financial performance (revenue, market share, profit per share, buybacks, etc). Then, we will train various models (LSTM or other) for stock prices, transformer (pretrained BERT or other) model for text sentiment analysis, and simpler models (regression, SVM or other) for economic indicators (regression) and financial performance (SVM to compare to analyst estimates). With these various models, we will have a final outcome based on an ensemble or hybrid model composed of the results from each of these individual models.

Lastly, our project may be hosted on a web server, using Flask and MySQL as a backend/middleware.

What features make up your MVP (Minimum Viable Product)?
(What features do you hope to accomplish by the end of the semester?)
Our main features include:
- A stock price prediction, used primarily for a buy/sell indicator. The feature will be a function of the output of the models mentioned above, and it will ultimately answer the question: is Alphabet stock a buy or sell today, given tomorrow's predicted price?
  - This will involve a lot of experimenting with different models, tuning parameters, transforming data, and extracting features.
- Automated data gathering and cleaning as daily/hourly stock price, news articles, economic indicators, and filings are published.
- A visualization of the stock's past price movements, as well as any relevant stories/economic indicators/financial statements/regulatory filings released recently.

What are some additional features outside the MVP?
(A list of features that would be nice to add after finishing the MVP.)
It would be nice to host it on a website, and to flesh out/include more data visualizations

Which does the tech stack look like and why did you choose these over alternatives?
(Feel free to discuss with your PM! Examples: React, Python, Java, etc. You do not need to know how to use these right now.)
- Python:
    - Environment:
        - Conda?: already has all relevant packages installed
        - Collab?: access to free GPU for model training; could use for the final model training after all data has been collected and cleaned
    - Libraries:
        - PyTorch: Neural networks, transformers
        - Numpy: matrix manipulation
        - MatPlotLib: visualization
        - Scrapy: web scraping data
        - Sci Kit Learn: regressions, statistical models
- Flask:
    - middleware/backend
- MySQL:
    - Time series data and our text data are both inherently tabular
        - columns can be time (days/hours) and the rows can be the relevant features (open/close/low/high prices, article body/text/title, label aka buy/sell/price point?)

What will the project timeline look like?
(Discuss this with your PM as well! You don't have to stick to it, but this should give you a general guideline for how the project should progress.)
- Find sources to gather data / begin building tools to gather data
- Finish tools that automatically gather data
- Think about models to use / think about parts of the data we want
- Begin cleaning data, labeling data
- Extract features from data / transform data