STOCK MARKET PROJECT:

Basic idea: - run a script every morning to fetch stock prices

- Use historical data using alpha vantage or finance to create a graph based chart of how the stock is performing
- ML part- we predict the future of the stock based on its current and past performance and check whether its feasible to invest in this stock or not.
- ML - part basic classification of each stock and its history (riser, faller, volatile, blue chip)
- Moreover the aim could be the target beginner stock investors who want to start investing but arent sure.
- For closing, we can check if we want to sell the stock or keep it.

Data sets and APIS

- Yahoo finance
- Alphavantage
- https://finance.yahoo.com/quote/%5EGSPC/history?p=%5EGSPC
- https://www.kaggle.com/datasets/andrewmvd/sp-500-stocks/code
- https://algotrading101.com/learn/yahoo-finance-api-guide/

Target Audience

- Beginners have difficulty in understanding market trends, making informed investment decisions, and the risk of financial loss without proper guidance.
- Need to discuss with the TPM as to how we can pitch it for new investors.
- Our target stocks all have high market caps, therefore they are always a safe bet and incredible low chance of losing money.

1. **Choice of dataset:**
   a. https://finance.yahoo.com/quote/%5EGSPC/history?p=%5EGSPC
      i. Overall historical data of the S&P 500
      ii. Has open value, closing values, daily highs, daily lows, adjusted close and volume (total stocks traded)
   b. Yahoo finance API to get the individual stock data and fetch real time data feed

   c. https://www.kaggle.com/datasets/andrewmvd/sp-500-stocks/code
      i. Has individual stock data as above
      ii. In addition it has sector, industry, trading location, market cap and revenue growth
2. **Methodology;**
   a. Data preprocessing.
      i. Data is feasible, the revenue growth, and individual open and close. We can use the data to calculate the volatility of the stock.

ii. Not much data needs to be processed, as all the prices will be in USD already, the dates are pre

b. Machine Learning Model:

- Primary objective is to predict the future performance of stocks based on their current and historical data.
- Classify stocks into categories such as risers, fallers, volatile and blue-chip.

$$\text{Volatility} = \sqrt{\frac{\sum_{i=1}^{n}(P_i - \bar{P})^2}{n}}$$

Where:

- Volatility is the calculated volatility.
- P_i represents each individual price in the set.
- $\bar{P}$ is the average price (mean) of the security's past prices.
- n is the total number of prices in the set.
- $\sum_{i=1}^{n}$ denotes the sum over all individual prices in the set.

-

Below are the models we are looking to use

- **Long Short Term Memory Network (LSTM)-** Type of recurrent neural networks

  PROS:

  a. Ability to learn long term dependencies and trends pattern and cyclic behaviours.
  b. Can handle sequences of varying lengths, like events happening at irregular intervals
  c. Easily handle gaps in data or missing data points

  CONS:

  a. Computationally intensive, however could probably be resolved as we intend to leave it running on a PC every morning
  b. Prone to overfitting therefore need to exercise caution while training large datasets

- **Random Forest Classifier -** Constructing a bunch of decision trees at training time and outputting the class that is the mode of individual trees.

  PROS:

  a. Less likely to overfit and can handle gaps and irregular data
  b. Multiple decision trees will help avoid one incorrect DT from affecting the overall output

  CONS:

  a. Prediction time may be slow with large forests.
  b. Computationally intensive but again could be resolved due to the PC's capabilities .

Other Models we could use are ARIMA(Auto Regressive integrated Moving Average) and SVM(Support vector machines) or GBM(gradient boosting machines)

3. Application

a. Users will input stocks that they are interested in. Will be able to just search the stock they are interested in as there are 503 stocks, and can filter via industry/specialty that they are interested in.
b. Users will receive a projection of the stock (graph extrapolation and value), as well as its classification.
c. We are looking to live plot the graph or update it every 5 minutes depending on how much my PC can handle.
d. We intend to deploy it on netlify and host it.