**Overall Blueprint:**

Python: Grab data from APIs, clean data, store it into SQL or Mongo db. (I prefer Mongo. Fornicate SQL because ORMs suck)

Python: Rescale the data.

Python or Jupyter Notebook (Keras): Feed data into AI. AI will then poop out results into DB.

Python: Run Flask server, create API that allows access to AI results and host website on Heroku or AWS or google cloud.

Javascript and HTML: Retrieve AI results and visually show the results.

**Data Collection:**

Find stock market data since 1980s or 1990s. As detailed as possible.

Clean the data such that for every given **learning time period** (P from now on), you’ll have access to:

52 Week high up till that point

Volume of trade that date or some moving average of volume over some number of days.

Earnings ratio

Dividend yield

Ideally, every single data point we can grab from online will be another input for the NN.

The learning time period is the time range of data that the AI will have access to.

**Data Scaling:**

This data must then be compressed into a format that the neural network can process. This means all data should be turned into some number between -1 to +1. The range should be limited to take outliers into consideration.

For example, if the average trade volume between 1980 to 1981 is 100 million with a range of 50 to 150 million, 150 million will be represented by +1 while 50 million will be represented by 0 or -1. However, there may be outlier days where the trade volume became 500 million. Do we ignore that and count them as 1? Or do we compress the entire scale to account for these outliers? Or do we apply a log to the scaling such that it’s taken into consideration but doesn’t go crazy.

**Generating training data:**

IF NO RNN: Once the data has been cleaned and scaled, it’s time to generate training data. We will need to decide what is the “range” of data we will want the AI to consider. Should the AI look at what happened the last day, last week or last month?

Create training data in the following format:

Decision Date (MetaData), 52week high, 4 week high, Today’s vol, etc etc etc

The last column, aka the test column, should say YES or NO. Aka should the AI have bought or sold the stock on that date?

*This must be decided on how much time we’re giving the AI to think. The idea is “Would it be a good idea to buy or sell right now if the next opportunity we have to buy or sell will be a day, week, or month from now?”*

**AI:**

The AI NN will be a standard feed forward NN that will be fed data from a specific P.

The NN will receive all the data from P, which will all be scaled to some value from 0 to 1 or -1 to 1. The NN will then decide if it’s time to buy or sell.

The AI will output its answer and it’ll be added to a row that will show all the information it was given (before it was scaled), the date of its decision, what its decision was, and how much gains/loss the decision would’ve resulted in.

**Front end:**

What was the average % gains or loss by the AI?

How much money would you end up with if you have given an AI $10,000 since 1980?

Show the performance differences between each AI in some kind of visual comparison.