I did not have much time to work on this, so it is a bit rushed. Sorry!!! ☹

Problem Statement: My objective is to devise a stock market strategy using machine learning techniques. I also wanted to incorporate some NLP if I had time (Still TBD).

Data preprocessing: the data is taken from Kaggle. The link has been put below. The data was mostly already processed, just moved some columns around, and removed most of them. Only interested in the closing price of each day for now.

Feature was simply stock price over 7 days, and label was if it increased by a certain percentage.

3. Machine Learning Model: Initially, I wanted to do a neural network, however, I was advised against it due to its unexplainable nature and inconsistent results.

Right now I’m mostly using some simple classifiers, however it would be better to use Kalman filters as was stated in the feedback for my first deliverable.

4.Overall, the performance of my model is pretty poor, although it usually performs slightly above random. (which I guess is good) I believe the project is still feasible but there is much to do and much to learn.

I’ve put a couple of graphs but I believe there are many aspects that should be improved upon, from the data to the label generation and classifiers I’m using.

5. I believe I might be moving to quantopian for easier backtesting and a more financial approach to the situation. However, there’s a learning curve and I had to push out the deliverable.

The data I have used is from Kaggle :

<https://www.kaggle.com/camnugent/sandp500/version/4>

Most of the code is taken from a YouTube series called Python for finance by sentdex.