Machine Learning for Stock Index Prediction

Shi Kaiwen

Directed Reading Program Presentation

Mentor: Hyun Jong Kim

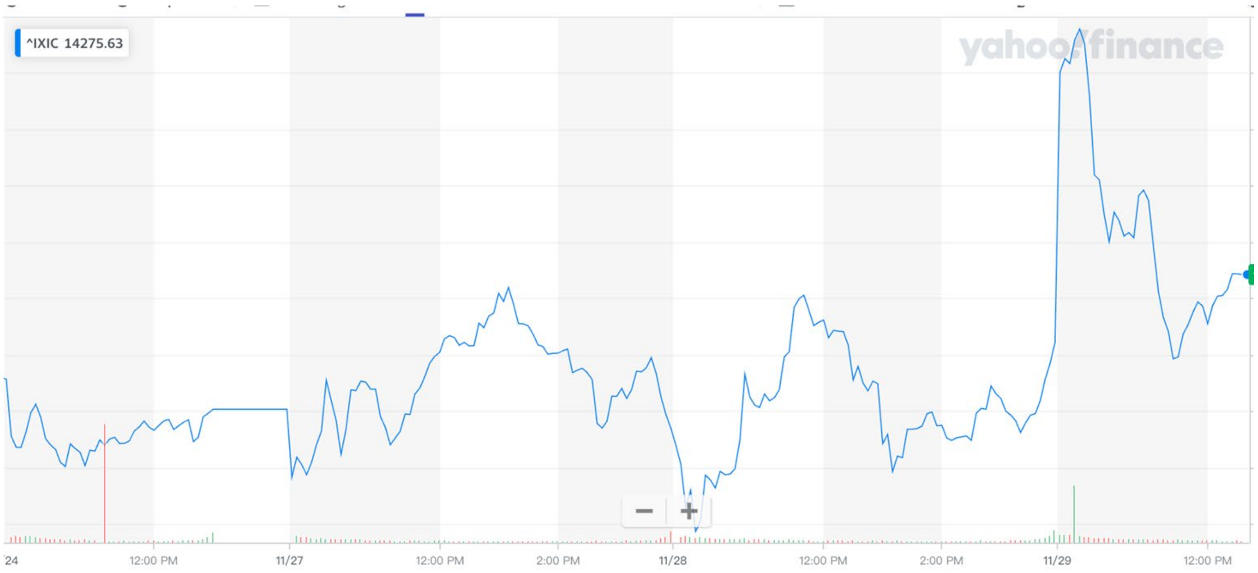
Fall 2023

Agenda

- Initiative
- Data Processing
- Training
- Result
- Next Step

Stock index prediction can be tricky...

Nasdaq index drops today; how will it go tomorrow?



Maybe we can generate some “features”...

- “Trends”: 1 if δ > 0; 0 else

$$\delta = Close[t] - Close[t-1]$$

- Daily Returns (**DR**)

$$DR = \frac{Close[t] - Close[t-1]}{Close[t-1]} \cdot 100\%$$

- Simple Moving Averages (**SMA(n)**):

$$SMA_n = \frac{1}{n}\sum_{i=1}^{n} P_i$$

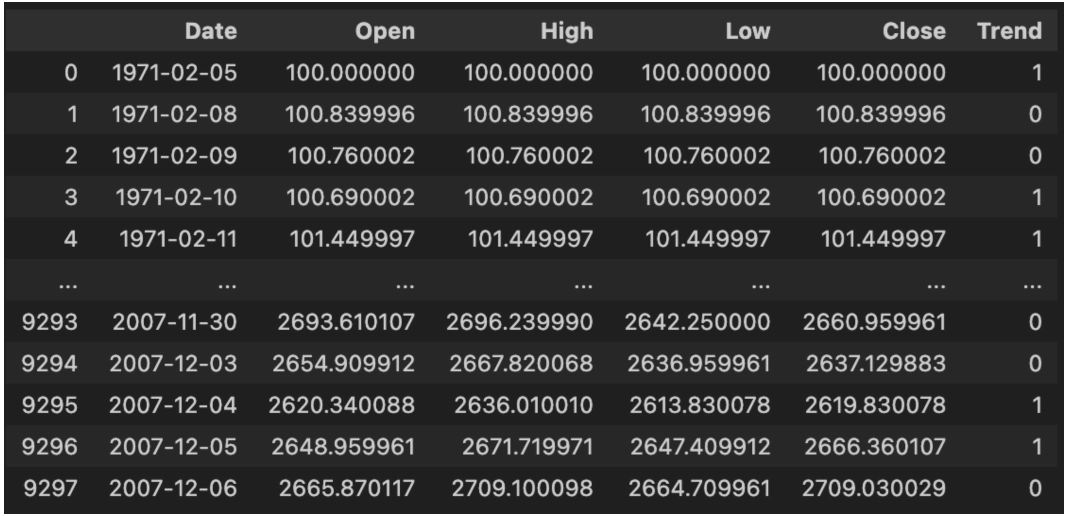
- ...

Start Training!

Approach 1: Bernoulli Naive Bayes

Approach 2: Single - Featured LSTM

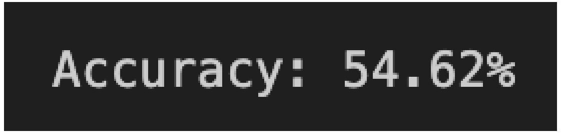
Approach 3: Multi - Featured LSTM



	Date	Open	High	Low	Close	Trend
0	1971-02-05	100.000000	100.000000	100.000000	100.000000	1
1	1971-02-08	100.839996	100.839996	100.839996	100.839996	0
2	1971-02-09	100.760002	100.760002	100.760002	100.760002	0
3	1971-02-10	100.690002	100.690002	100.690002	100.690002	1
4	1971-02-11	101.449997	101.449997	101.449997	101.449997	1
...	...	...	...	...	...	...
9293	2007-11-30	2693.610107	2696.239990	2642.250000	2660.959961	0
9294	2007-12-03	2654.909912	2667.820068	2636.959961	2637.129883	0
9295	2007-12-04	2620.340088	2636.010010	2613.830078	2619.830078	1
9296	2007-12-05	2648.959961	2671.719971	2647.409912	2666.360107	1
9297	2007-12-06	2665.870117	2709.100098	2664.709961	2709.030029	0

Bernoulli did a decent job...

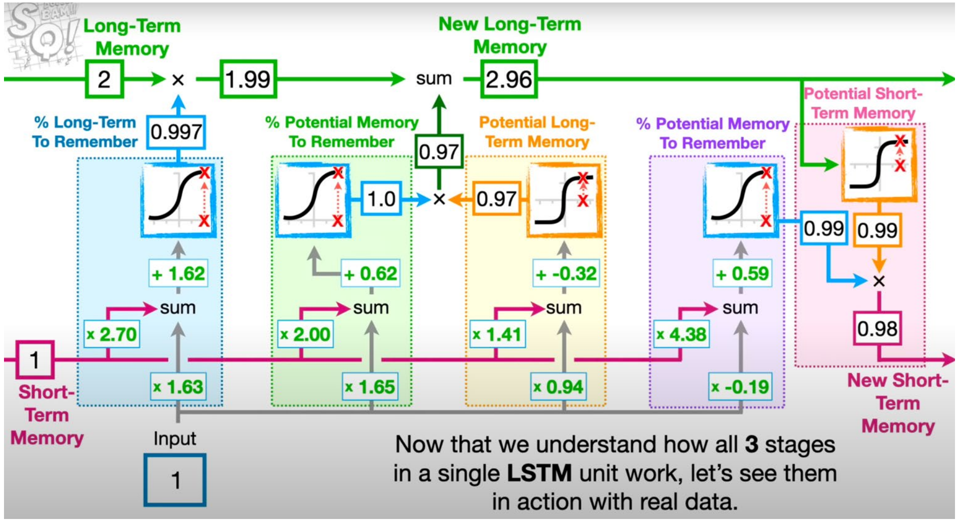
Some Simple Probabilistic thing…



```
Accuracy: 54.62%
```

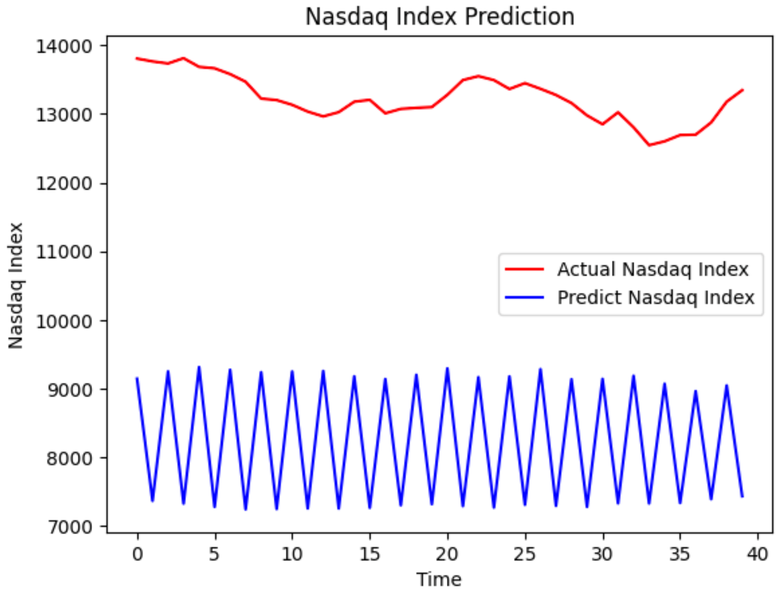
Long Short Term Memories (LSTM)..

- Good for time-series datasets (e.g., stock prices)
- Use 2 things to predict:
  - Long-Term Memories
  - Short-Term Memories



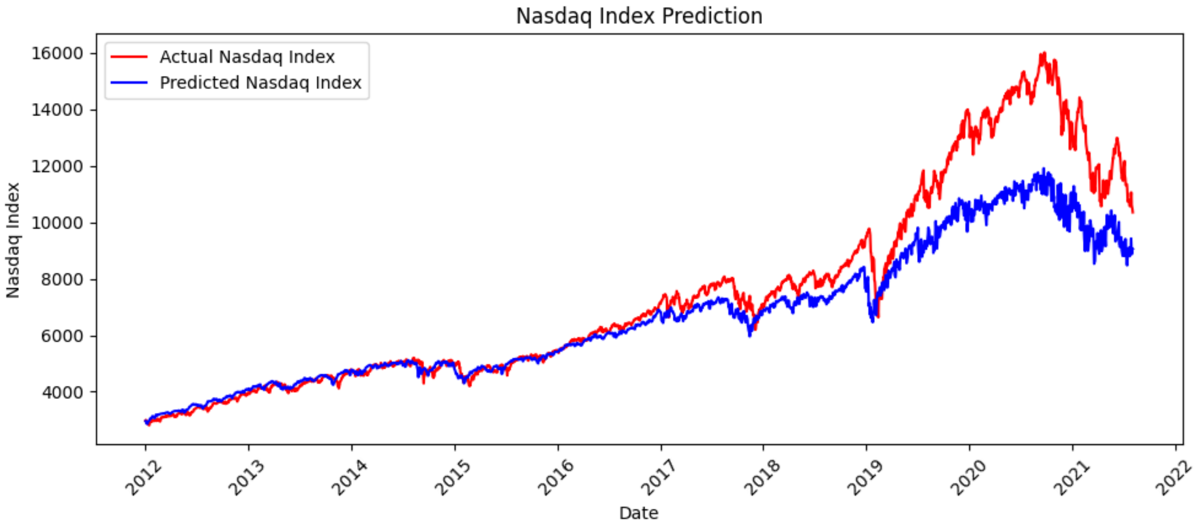
Now that we understand how all **3** stages in a single **LSTM** unit work, let's see them in action with real data.

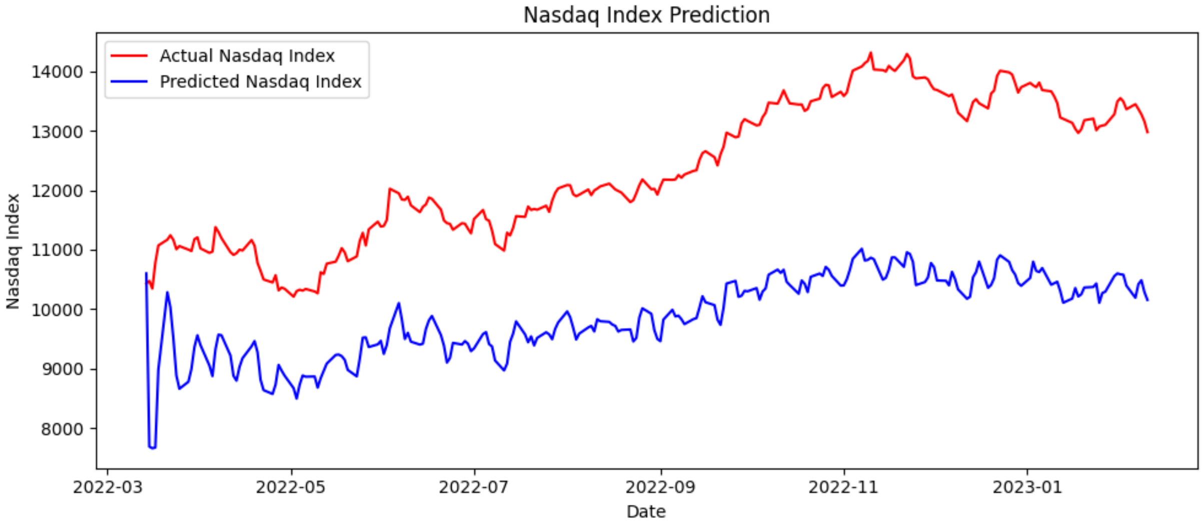
What happened if we just try to predict one month…



But general, LSTM seem to be good

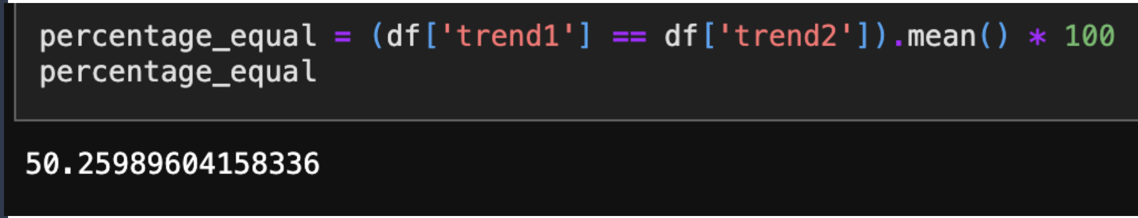
10 year & 1 year prediction





Is the graph the truth?

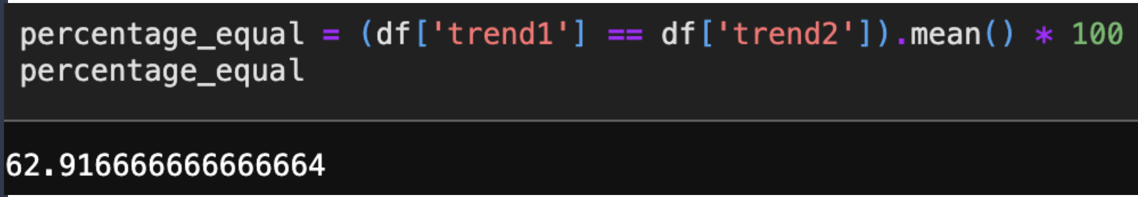
10 years:



```
percentage_equal = (df['trend1'] == df['trend2']).mean() * 100
percentage_equal
```

```
50.25989604158336
```

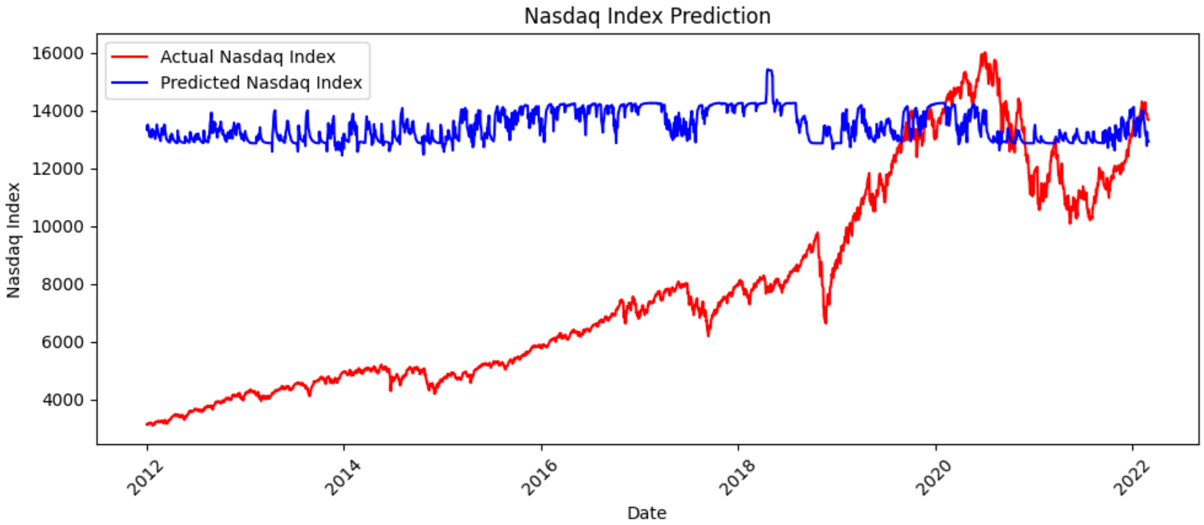
1 year:

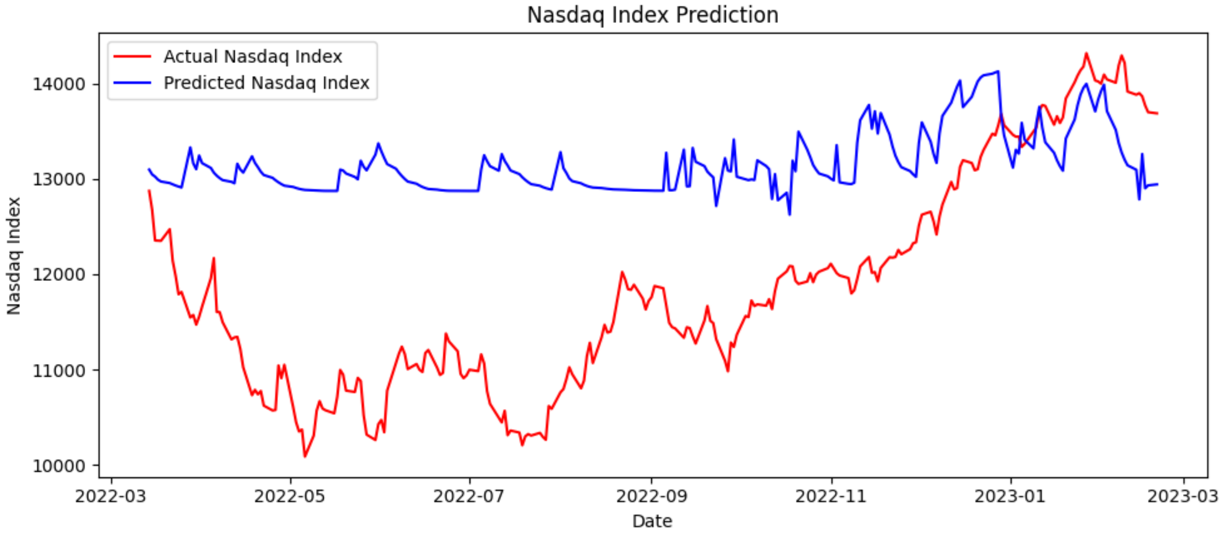


```
percentage_equal = (df['trend1'] == df['trend2']).mean() * 100
percentage_equal
```

```
62.916666666666664
```

What if we add more features...





Conclusion so far:

1. Lot of things to learn when we try to “trade rationally & wisely”
2. Trend Similarity ≠ Prediction Accuracy!
3. Long way to go to make more accurate predictions

Next step...

1. Figure out why multi-feature does not work
2. Seek ways to improve prediction accuracies
3. Use it for actual trading!

Questions?