

REPORT

[Abhinav RaghuvanshiH2]

Label 'Y' and its significance -

0 is when your stock starts to rise/bullish curve/ideal to sell, 1 is when the stock starts to fall/bearish curve/ideal to buy

	A	B	C	D	E	F	G	H	I	J
1	y	x1	x2	x3	x4	HL	OC	HL - Close	OC - Close	y
2	0	3861.22	3861.66	3860.75	3861.38	3861.205	3861.3	-0.175	-0.08	0
3	0	3861.19	3861.38	3860.16	3861.25	3860.77	3861.22	-0.48	-0.03	0
4	0	3861.38	3861.38	3858.47	3858.53	3859.925	3859.955	1.395	1.425	0
5	1	3858.72	3858.72	3857.03	3858.19	3857.875	3858.455	-0.315	0.265	1
6	1	3858.09	3859.16	3858.03	3858.81	3858.595	3858.45	-0.215	-0.36	1
7	1	3858.63	3859.63	3858.28	3859.63	3858.955	3859.13	-0.675	-0.5	1
8	0	3859.53	3859.88	3858.88	3859.66	3859.38	3859.595	-0.28	-0.065	0
9	1	3859.69	3859.72	3858.22	3858.59	3858.97	3859.14	0.38	0.55	1
10	1	3858.72	3860.31	3858.66	3860.19	3859.485	3859.455	-0.705	-0.735	1
11	0	3860.31	3861.81	3860.31	3861.53	3861.06	3860.92	-0.47	-0.61	0
12	0	3861.34	3861.59	3860.41	3860.5	3861	3860.92	0.5	0.42	0
13	0	3860.75	3861	3859.97	3860.13	3860.485	3860.44	0.355	0.31	0
14	1	3860.19	3860.22	3858.97	3859.5	3859.595	3859.845	0.095	0.345	1
15	1	3859.38	3859.97	3858.78	3859.56	3859.375	3859.47	-0.185	-0.09	1
16	1	3859.69	3861.91	3859.25	3861.69	3860.58	3860.69	-1.11	-1	1
17	0	3861.56	3862.5	3861.56	3862.19	3862.03	3861.875	-0.16	-0.315	0
18	0	3862.13	3862.19	3861.13	3861.66	3861.66	3861.895	0	0.235	0
19	0	3861.5	3861.84	3861.22	3861.56	3861.53	3861.53	-0.03	-0.03	0
20	1	3861.44	3861.47	3860.91	3860.94	3861.19	3861.19	0.25	0.25	1
21	1	3861.13	3861.88	3860.72	3861.53	3861.3	3861.33	-0.23	-0.2	1
22	1	3861.56	3863.22	3861.56	3863.03	3862.39	3862.295	-0.64	-0.735	1
23	1	3863.16	3864.75	3862.78	3864.47	3863.765	3863.815	-0.705	-0.655	1
24	1	3864.66	3866.59	3864.66	3866.59	3865.625	3865.625	-0.965	-0.965	1
25	1	3866.47	3869.28	3866.47	3868.38	3867.875	3867.425	-0.505	-0.955	1
26	0	3868.5	3870.72	3868.31	3870.72	3869.515	3869.61	-1.205	-1.11	0

As we can see in the excel data computed via taking average of Opening and closing price and subtracting the closing price from this average. Value tends to rise in +ve value with trailing zeroes which signifies that the average price of the stock is rising as compared to the closing price. As soon as a 1 is encountered, the value pivots from positive to negative hinting towards a fall in stock prices hence trailing 1s lead to a falling/bearish curve.

Until another 0 is encountered which again pivots the value from negative to positive signifying a change from falling prices to rising prices again

And as a binary label (0s and 1s namely) it makes perfect sense for the profit or loss of a trader to be dependent on the OHLC prices .

Above logic can be easily verified with a OHLC graph where at times we find rising prices plotted with a different color than the falling prices and this is the only useful indicator for a trader to draw conclusions from and is hence an important thing to look out for while analyzing data of such type as given un the dataset .

RESEARCH -

[\[A New Methodology to Exploit Predictive Power in \(Open, High, Low, Close\) Data\]](#)

I wasn't able to fully implement this technique but the paper has some very good techniques that one can use to his/her advantage

The accurate prediction of an asset's direction has long been the goal of many. This paper utilises ML technology in the form of Randomised Decision Trees (RDTs) and Long Short Term Memory Recurrent Neural Networks (LSTM RNNs) as a key component in a process for trend detection. It takes advantage of the relative ease of prediction of the mid-price (defined in terms of OHLC candlestick levels) when compared to the traditional close price prediction target.

[\[An Attention-Based LSTM Model for Stock Price Trend Prediction Using Limit Order Books\]](#)

Another very good paper I came across, shows a brilliant technique utilizing LSTM.

Traditional methods for predicting stock price trends are mostly based on the historical OHLC (i.e., open, high, low, and close prices) data. In this paper, another type of stock price data, i.e. limit order books (LOBs), is used. An attention-based LSTM model is applied to predict the trend of a stock market. The effectiveness of the proposed model is validated by experimental results. Stock price trend prediction is a challenging task due to non-stationary and high volatility of the stock prices. To address this problem, in the paper, an attention-based LSTM model is applied to predict the stock price trend by using the LOB data instead of traditional OHLC data. Especially, the LSTM layer is used to learn the temporal relationships in the sequential LOB data. And then, the attention layer is applied to produce a better representation to make the model focus on the salient features related to the stock price trend. Finally, through using the real LOB data and stock price data in the China stock market, experimental results have validated the effectiveness of the proposed model.

METHODOLOGY AND MODEL DESCRIPTION -

I tried multiple models mainly utilizing directly the Opening Closing High & Low prices as features for the models but did not get very good results and there is not necessarily a curve that could fit the graph of the actual stock prices graph which one sees usually on different website but given that the task was binary classification, I got 54.3% accuracy on the validation set (split 20-80) for Multi Layer Perceptron And Logistic Regression and 49.35% from Random forests. But the final model is an ensemble of all the 3 models which was able to achieve a final accuracy of 54.95% on the same validation dataset as can be seen in the .ipynb notebook uploaded on github. The ensemble was weighted voting classifier with weights assigned as follows: 1.5 for Logistic Regression 1 for Multilayer Perceptron and 1 for Random Forests. I have tuned the hyperparameters and achieved the best results for the given configuration.

RESULTS -

I've learned a lot about the financial world from this task which was rather unknown to me before this and have reached a conclusion that predicting stock prices is a fairly difficult task with not much literature available in the AI community and this is the reason ML models usually are not able to fit the actual trend of prices