

Stock price



with Decision Trees & Random Forests

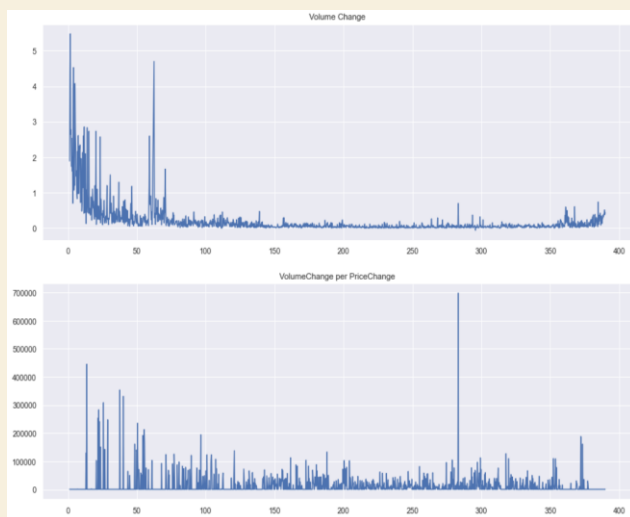
I am crazy about predicting stock prices. To figure out the best way to minimize the loss in our choice, analyzing the movement of the giants such as investment banks, fund managers, and tycoons are very required. A stock price can be manipulated but not trade volume. That is, volume can be compared to a log file containing all the historical facts. Again, only if we can go back in time, we can possibly change the amount of the past trade volume. For these reasons, I gathered the price and volume for a day and calculated subsets from those two variables to find insight from the dataset.

1. Dataset

This data comes from our GitHub repository. We gathered the real-time stock price with its volume by using API-yfinance and then get subsets. US stock market operates for 390 minutes from 9 30 AM to 4 00 PM, the data contains the values checked approximately three times every single minute (every 20 seconds). The price gathered here is an adjusted close price which applied an equity change, stock split as an example.

2. Data Exploration

We conducted a sample data visualization for some interesting insights. The supplementary report contains more detailed visualizations, the process to get the result, and how we verified the result.



3. Confusion matrix

	Actual Positive (Up)	Actual Negative (Down)
Predicted Positive (Up)	84	68
Predicted Negative (Down)	78	232

The confusion matrix shows there are fewer false positives and false negatives than true positives and true negatives.

4. Classification Report

	precision	recall	f1-score	support
0.0	0.52	0.55	0.54	152
1.0	0.77	0.75	0.76	310
accuracy			0.68	462
macro avg	0.65	0.65	0.65	462
weighted avg	0.69	0.68	0.69	462

The accuracy score is about 68%.

5. Summary of Findings

We conducted a decision tree and random forests to predict whether the stock price would go up or down based on the given features. The accuracy measures how often the classifier makes the correct prediction, and the precision tells us what proportion of trials we classified as 'up', actually price-up. Recall tells us what proportion of trials that actually were price-up was classified by us as 'up.' Finally, the weighted average of the precision and recall scores represents the F1 score. The supplementary report will explain the whole process and features applied in detail. Here is my belief:
Trade Signal Insight = Volume Speed + Volume Speed Acceleration + Volume Change per Price Change