

Applying Machine Learning to Stock Market

Statement of Work

Utsav Vanodiya (100804273)

AI Algorithms 1 - AIDI 1002

Course Facilitator – Marcos Bittencourt

Problem Statement

I would like to apply Machine Learning to make better decisions for intraday stock trading. From my past experience, I have noticed that there are strong patterns exist, so where ever patterns exists, we can use Artificial Intelligence to identify those things accurately. However, there are lots of factors affect market. For example, elections, government policies, budgets, international events (trade war, fight, pandemic, changes in oil prices, etc). Here, we need to understand one thing that is patterns still exist in these kind factors too. So, we can also use these factors to make final decisions.

Rational Statement

To make a better trade in intraday, we should follow below golden rules religiously as per the gurus.

1. Never trade against the market trend.
2. Never trade more than your losing capacity.
3. Always put stop loss order and stick with it.

So, if you notice the first point, we can see that we have to go with the flow, not where we want to go. Here, other two points don't need any sort of ML, because those depend on individual's capacity and desire.

To make successful trades, I would like to apply ML to answer below several questions. Those are highly related to the first point, because it's supreme in day-trading. *(Note: In trading, there is a concept called short sell, it means we can sell stocks even if we don't own it, but at the end of the day, we have to buy it to square off the position. This type of strategy is used to make money in downtrend of the market.)*

Questions to be answered by ML models:

1. What will be the market trend for the next day?
 - a. We can categorize market based on change in percentage. For example, above 0.8% = Strongly Positive, between 0.8% to 0.4% is Positive, 0.4% to -0.4% is Mixed, Less than -0.4% as Negative and less than -0.8% as Strongly Negative.
 - b. The point is, either we can perform regression or classification based on the past data. We can try both the approaches to find the trend.
2. What are the best stocks to pick for the next day?
 - a. We know that the market trend is supreme, but after knowing it, we need to identify the stocks which can easily go in that direction quickly.
 - b. The stocks which have high volatility and liquidity are usually considered as the best stocks. However, we have to find the best stocks for positive trend as well as for negative trend.

- c. So, if market is negative, we will go with negative stocks and short sell it. If market is positive, just buy it.
- 3. What are the correct entry and exit points?
 - a. After selecting the stocks, we need to predict the best entry and exit points.
 - b. If we fail on that, we are going to loss money even if market is up/down heavily.
 - c. So, we can apply ML to predict low and high values for the next day.

Data Sources

- 1. Kaggle.com
 - a. <https://www.kaggle.com/hk7797/stock-market-india>
 - b. Above link provides 1 minute dataset of 150 stocks of NSE(National Stock Exchange) and 9 Indexes of NSE from 2017-01-02 to 2020-05-15.
- 2. https://economictimes.indiatimes.com/indices/nifty_50_companies
 - a. From the above link, we can get live data of top 50 stocks of NSE, current Nifty index value, number of advance/decline stocks, volume, last 30 and 365 days change.
 - b. In case of testing our model on live data, we can scrape data from it. I have made a simple python script which scrapes 5 minute data and found that they don't put restriction on that number of calls.
- 3. NSE official website
 - a. <https://www.nseindia.com>
 - b. We can download daily open, high, low, close data from this site.
 - c. So, in case of testing out models after 5th month, we can use this site.

Data Assumptions, Limitations and Constraints

We should only trade on liquid stocks, so only volatility is not important, but liquidity too. So, in this case, I will perform analysis of only top 50 stocks. So, I'm assuming that the provided dataset contains data for all of these top 50 stocks and there will be no missing values for them.

As far as limitations and constraints are concerned, I don't see any of it in this case. It's public dataset. However, at some point in future, we might face problem related to less number of samples, but I'm not sure that whether it will happen or not.

Testing Processes

First of all, in this scenario, even if we get 70% accuracy, it would be great, because even expert traders don't get much accuracy than this, so, we will consider it as minimum requirement.

For model testing

1. Separate dataset on three sets: training, validation and testing.
2. Apply various regression testing methods like MSE, RMSE, MAE, R-square, etc.
3. For classification, we will go for F1-Square and Confusion Matrix.

For final application testing of above model

Precision is the key here, it doesn't matter if our model miss some good trades, but how much precisely it predicts few trades.

$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$

So, at the end, I will consider that based on my trading strategy and with the help of this model, how much successful trade I could have made. So, will try to identify how much precision I could have achieved with this model.

Strategy

Based on market condition, there will be some profit/loss ratio will be defined. Ratios will be highly correlated with the market movement. For example, if market is up by 0.5%, we can expect 1% profit and 0.5% loss. Here, our risk to reward ratio is 1/2.

If model can help to achieve those results, it means trades are successful else fail. After that, precision will be calculated.

Data Acquisition and Understanding

Our dataset contains 1 minute of data for 150 stocks from 2017-01-01 to 2020-08-14. There are data of 9 indices too. Data is from NSE (National Stock Exchange), which is one of the major stock exchange of India.

Exploratory Data Analysis can be seen in Jupyter notebook `eda_of_one_minute_stock_data.ipynb`

Key Points

- We have data of 895 days.
- Data of 150 stocks.
- 9 Indices of NSE.
- Missing data of only one day, which we can see in the notebook.

Correlations

We can see that banking companies and some other companies are moving with market. So, we can leverage this knowledge to make powerful models.

Feature Engineering

We can calculate simple moving averages to improve accuracy of our models. This is one strategy to predict that whether price can go up or down based on breaking certain level. If stock moves more than moving average, we can find an opportunity.

Data Manipulations

- We will consider data of one day as one sample. For example, data from 09:15:00 to 11:00:00 will be considered as features. After 11:00, high and low values will be our output.
- We will rescale data between 0 to 1 for each day for each stock, so for every day, we will apply min max scalar rather than on all values for particular stock. For example, ASIANPAINTS -> 2017-01-01, 2017-01-02, 2017-01-03,... and so on.
- Above example says that we will rescale for each day for each stock, because we are working for intraday trading.