WiDS 2023 REPORT

*TOPIC: Dynamic Stock Market Analysis*

*MENTORS: Siva Madhav Chinta and Rakshan*

Name: Anshu Arora

Roll number: 22b1207

In this project I have learned numerous new things, like how the stock market works, the different technical indicators and also got comfortable with coding in python.

Below is a detailed explanation of the code. Justification for choosing particular technical indicators, choice for machine learning model, threshold for updating position and investment strategy is also elaborated upon.

Started off by have just imported the required python libraries. Tvfeed library helps us get the data of the stock, and TA-lib helps in doing technical analysis

Imputer is used to fill in any missing data, using the strategy mean- that is it fills the missing data by taking the mean of the values around it

The most common and popular indicators are the following, and I have justified why I am taking these indicators

EMI\_200 - Calculates the average price of a security over the last 200 periods. Emphasizes recent prices more than older ones, giving greater weight to the most recent data.

MACD (Moving Average Convergence Divergence) - Consists of two moving averages: the MACD line (the difference between a short-term and a long-term EMA) and the signal line (EMA of the MACD line).

RSI (Relative Strength Index) - Measures the magnitude of recent price changes (involves the number of days the stock closed up and down) to evaluate overbought or oversold conditions.

AROON (Aroon Indicator) - Consists of two lines, Aroon Up and Aroon Down, measuring time since the highest and lowest closing prices, respectively.

BOLLINGER BANDS - Consists of a middle band (typically a simple moving average) and two standard deviation bands above and below. Bands expand or contract based on volatility, helping identify potential overbought or oversold conditions. Hence we will only consider Bollinger bands up and Bollinger bands low.

OBV\_LAST (On-Balance Volume) - Adds or subtracts the day's volume to a cumulative total based on whether the price closes higher or lower.

SAR\_NOW (Parabolic Stop and Reverse) - Provides potential entry and exit points by plotting points on a chart based on price direction. It is a complicated calculation different from the rest of the indicators hence we will have it as a feature

ADX (Average Directional Index) - Measures the strength of a trend, whether up or down using the closing and opening price data

As we can see, each of these indicators has different mathematical derivations and is based on different parameters, so together seem to form a mutually exclusive collectively exhaustive list of technical indicators we can use.

Then we are going through the data for each of the niftyfifty stocks and and are defining lists X\_train, Y\_train, Y\_train\_pos and gain.

X\_train list helps in keeping track of the dataframes (with the indicators) and Y\_train is used to train the model to find the gain

Based on the gain (we are calculating percentage gain of the closing price of the stock and the next day opening price) we set our ‘position’ – that is if we want to sell, buy or hold the stock

Then I have added the dataframes of technical analysis so that we use this information while we train our model

I have not added bbands\_mid and ppo as they have high correlation – they are not bringing any independent insights, hence it is better to not include them. This makes our model more accurate.

Here we are just creating a list (for each stock, since we are still running in the main loop) with the index of when the trading day starts and when it ends (these are added to separate lists)

The last line of code is just to make sure the day is complete (we have opening and closing data for that day)

Also, we are considering the opening time as 9:25, though the market opens at 9 because the first few mins involve a lot of fluctuation in the price. Similarly for the closing time.

Here we are checking if there is no missing data for that day, and that the 5 columns we dropped have gotten removed in our new data frame

After this we append the values after flattening in X\_train. This will be used for training purposes

Then we are calculating the gain

Here we are normalizing the data- that is bringing all the values from 0 to 1 based on the minimum and maximum values.

Now finally we are training our model! The reason to use Random Forrest regressor is that it is a tree based decision making model that is good at capturing the fluctuations in data (this is important for stock market prediction).

We are first shuffling the X\_train and y\_train so that no underlying patterns due to the sequence of data is captured

Now we are finding the position – that is if we want to sell, hold or buy the stock based on the gain. We are appending our position in y\_pred\_pos

Now to run our code at 3:15pm

In the above code snippets, we are basically running our model on real data (that we get from TV-data feed) and are calculating the predicted gain.

However this time our position is on a greater threshold of gain (0.2)

Then we are implementing the model to see the predicted gains in real time.

There are two different strategies to generate profit, and below is the explanation

the loop iterates through the top 10 symbols and calculates the cost of purchasing each symbol. A fee that is 0.0005 times the total cost is deducted from the total profit.

Method 1- Here we check the cost of purchasing for each symbol and check if we have enough balance to purchase it, and we continue doing this iteratively till balance is exhausted.

Method 2- Here we are trying to distribute the capital (10,000) equally amongst the top 10 predicted stocks

As we can see, the profit that arises from method 2 is higher, and even across other simulations that I performed, method 2 consistently did better

A possible reason for this is because in method 2 we split our capital equally among different stocks and that hedges our risk. One must keep in mind that our model is not perfect and may also cause losses on some days, so dividing the risk between many stocks is a good method to generate profits.