Algorithmic Trading Bot Final Report

Name: Ayush Singhania

Roll: 20171031

Supervisor: Prof. Kavita Vemuri

Introduction

Speculation in stock markets is believed by many to be akin to gambling and winning in stock markets is thought of as pure luck. But such a viewpoint may not be entirely true as over the time, several algorithms have been developed and are currently used by many reputed trading firms for their primary trade decision making process. Do these algorithms really work? If so, what are the circumstances when the algorithms may falter and what are the circumstances that the algorithms predict perfectly? These are the main questions that this project intends to answer through implementation, simulation and experimentation.

This report will describe the entire journey of this project in brief. We will begin with the targets that were supposed to be achieved by the end of the project, followed by the project findings and developments in a timeline fashion and ending in a brief conclusion of the overall project.

Aim of the Project

1. To research and learn about the financial markets and the banking system, so as to get an in-depth perspective of how the markets function and the parameters that influence the stock prices.
2. To build an algorithmic recommendation system that will be trained on the opening and closing stock prices of a basket of listed company stocks, to make predictions on when to buy and sell the stocks for maximum profit.

Project Progress Timeline

Learning Phase

The first few weeks were utilized by researching and learning about the fundamentals of the banking system, the role of governments and central banks in regulating the market. The main tasks completed were:

1. Learned about the fundamentals of banking. The major concepts that I learned are :
    a. ADR (American Depository Receipts)
    b. Capital Account Convertibility
    c. Capital Gains and Indexation
    d. CRR (Cash Reserve Ratio) or Liquidity Ratio
    e. Repo Rate and Reverse Repo Rate
    f. SLR (Statutory Liquidity ratio)
    g. Credit Spread
    h. Fiscal Deficit
    i. Revenue Deficit
    j. Trade Deficit
    k. Fiscal Stimulus
    l. Market Stabilization Scheme(MSS)
    m. Monetary and Fiscal Policy

2. Learned about the various financial instruments that are used by financial analysts, stock traders and investment banks; their advantages and the risks involved in each. The major concepts that were covered are:
    a. Arbitrage
    b. Commercial paper
    c. Treasury Bill(T-Bill)
    d. Certificates of deposit(CD)
    e. Credit Default Swap(CDS)
    f. Dividend Yield and related calculations
    g. Hedge Funds and their working
    h. Mutual Funds
    i. Debentures
    j. Mark to Marketing Accounting Process
    k. Money Market
    l. Oil Bonds

m. PTC(Pass Through Certificate)
n. PE (Price to Earnings) ratio
o. PEG (Price Earnings to Growth) ratio
p. PPP(Purchasing Power Parity)
q. Sharpe Ratio
r. Information Ratio
s. REIT (Real Estate investment Trust)
t. ROCE (Return On Capital Employed)
u. RONW (Return On Net Worth) or ROE (Return On Equity)
v. Speculative Attack
w. Stagflation vs. Inflation
x. Market Capitalization
y. Velocity of Money
z. Yield Curve

3. Learned about how to understand financial statements, the jargons of accounting and how accountants can bend the rules of accounting to suit their purposes. The major topics covered were:

   a. GAAP and FASB(Financial Accounting Standards Board)
   b. Cash Accounting vs. Accrual Accounting
   c. Balance Sheet:
      i. Types of Assets
      ii. Types of Liabilities
      iii. Types of Equity
      iv. Making T accounts and Double Entry Ledgers
      v. Statement of Owner's Equity
   d. Income Statement :
      i. Gross Margins
      ii. Operating Income vs. EBIT (Earnings before Interest and Tax)
      iii. EBITDA
      iv. Net Income
      v. Depreciation vs. Amortization
   e. Statement of Cash Flows:
      i. Leveraged Buyouts
      ii. Operating Activities
      iii. Investing Activities
      iv. Financing Activities

f. Liquidity Ratios
g. Capitalization Ratios
    i. Financial Leverage
    ii. Debt to Capital
h. Activity Ratios
    i. Assets Turnover
    ii. Inventory Turnover
    iii. Day Scales in Inventory
i. Profitability Ratios:
    i. ROS
    ii. ROE
    iii. RONW
j. Du Pont Chart
k. Sales Price variance
l. Sales Volume variance
m. Total Sales variance
n. Purchase Price variance
o. Labor or Material Efficiency variance
p. Cost Accounting Basics
q. Decision Tree analysis
r. Cash Flow Analysis

4. Learned about the fundamentals of stock markets and how they work.

5. Learned about reading Japanese Candlestick Patterns. The most commonly used patterns include :
    a. Marubozu : Bullish and Bearish
    b. Spinning Top : Bullish and Bearish
    c. Doji : Bullish and Bearish
    d. Paper Umbrella : Bullish and Bearish
    e. Hammer : Bullish and Bearish
    f. Hanging Man : Bullish and Bearish
    g. Shooting Star : Bullish and Bearish
    h. Engulfing : Bullish and Bearish
    i. Piercing pattern : Bullish and Bearish
    j. Dark Cloud Cover : Bullish and Bearish
    k. Morning Star
    l. Evening Star

6. Learned about charts and related concepts.

7. Learned about support and resistance construction.

8. Learned about Technical Analysis. The major concepts include:
    a. Moving Average
    b. Crossover Systems
    c. RSI (Relative Strength Index)
    d. MACD (Moving Average Convergence Divergence)
    e. Signal Lines
    f. Bollinger Bands
    g. Dow Theory

9. Learned about several stock selection procedures, their merits and demerits.

Implementation and Testing of Individual Algos

After learning about the required financial concepts and the various algorithms used in technical analysis, I started implementing various algorithms from scratch and tested them with stock prices taken directly from BSE's website. After iterating through the merits and demerits of the various algorithms, I chose 4 commonly used algorithms to be used for the final bot. The chosen algorithms were:

a. Moving Averages
b. Support and Resistance
c. Candlestick Patterns
d. Bollinger Bands

Now, we will look into the chosen algorithms individually and in detail:

A. Moving Averages

1. **Algorithm:** The Moving Averages (MA) algorithm studies the average trend in the movement of the stock prices over a given window of time. It smoothens out the fluctuations in the window giving us an idea of the general trend of that stock. In this algorithm we use moving averages of two different window sizes for comparing the long and the short term trend in the movement of the price. This is done because of

the economic theory which states the market has two trends – a long term and a short term trend superimposed on each other. The MA with the shorter window acts a measure of the short term trend and the one with the longer window is used as a measure of the long term trend.

The buy-sell decision is based on the assumption that when the short-term trend line crosses over the long-term trend line, it is expected that the price is about to rally up and thus this is the time to buy the stock and when the short-term line crosses below the long-term trend line, it is time to sell the stock.

Instead of using simple buy-sell operations, I have used the market standard of long-short operations. Going long on a stock is equivalent to buying the stock. Shorting a stock is different from simply selling a stock. It means borrowing the stock when the price is high and selling it off on the market, then buying it again when the price has fallen and returning those stocks to the our lender. This allows us to profit from the change in the price of the stock between the selling and buying back the stock.

2. **Implementation Details:** For construction of MA based prediction, the most important parameter to be decided is the window size for the short-term and long-term trends. A generally accepted ratio of 1:4 is used in the market that is the longer trend window is 4 times the shorter trend window. The window size however mainly depends on two factors:
    a. The time frame of trading, whether it is intraday (buying and selling on the same day) or swing trading (the buying and selling times are a few weeks or months apart). Generally shorter windows are more profitable for shorter time frames as they cover the fluctuations better.
    b. The general market volatility at that time. Higher volatility warrants a longer window to smooth out the unnecessary noise.

For the purpose of this project I decided to work on swing trading rather than intraday trading, thus the time frame of trading will be around two to three months. To make the algo more robust to changes in the type of stock used for prediction and the changing volatility of the market, I decided to simulate trading the stock of interest on historical data by varying the window sizes and then using the window size with the maximum profit.

The stock that was chosen for running the simulation was Housing Development Finance Corporation (HDFC) Ltd. and the stock price data was collected directly from the BSE website. The collected data included the Open, Low, High, Close price of the stock from 1st Jan 2015 to 31st Dec 2018, i.e. 1043 trading days. The closing prices were used for the algorithm and are plotted below:

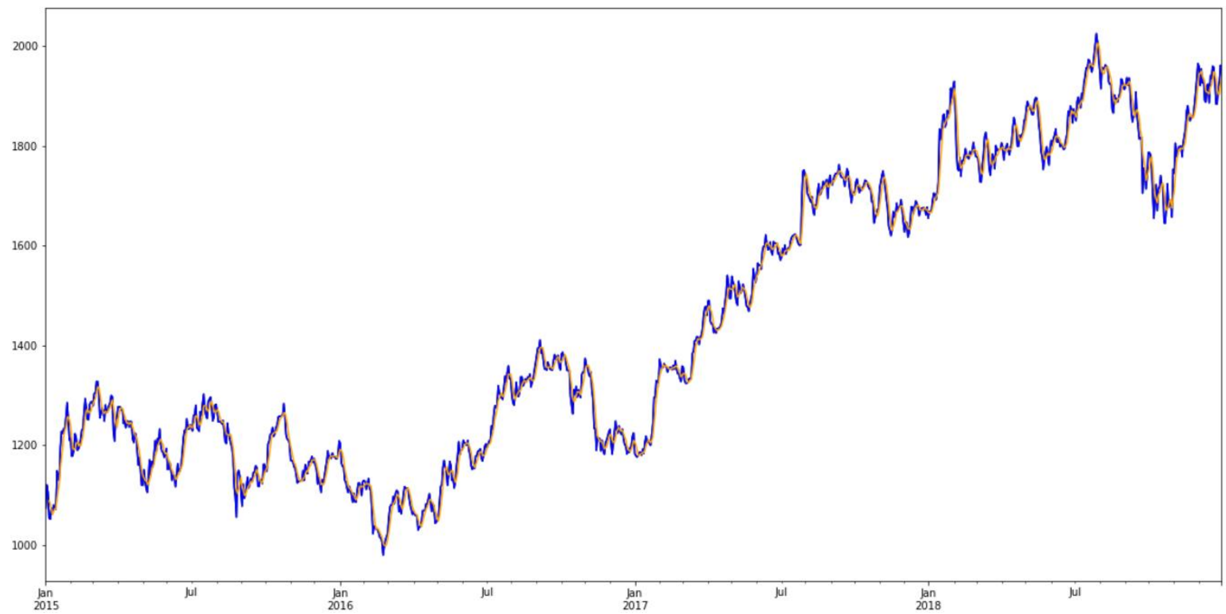
The short-trend and long-trend lines for a 50-200 day window size are plotted as follows:

The blue line corresponds to the 50 day MA and the orange to the 200 day MA. As is

evident from the plot, the short-trend line follows the fluctuations of the price much closely than the longer one.
Based on the buying strategy discussed above the periods for going long and short are depicted in green and red respectively:

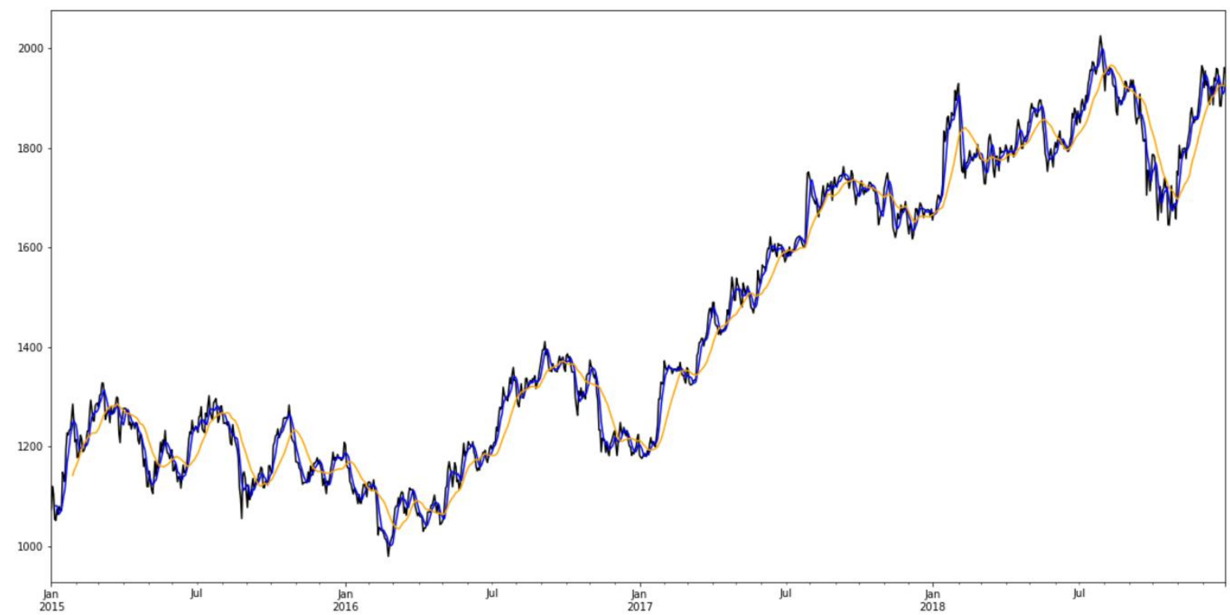
After examining that my algo was being able to plot the MA trend lines properly, I simulated the same process by varying the window sizes to determine which window size yielded the maximum profit. The short trend window sizes were iterated over 1, 5, 10, 20 and 50 and the longer window was kept at 4 times the shorter one. The calculation of the profit was done by using the long-short strategy mentioned above with the transaction days being the days when the shorter trend line crossed over or below the longer trend line. The trend lines, the long and short periods and the points of transaction are shown in the following plots for each of the window sizes. Also mentioned are the net profit or loss incurred and the accuracy of the algorithm in predicting the future trend by choosing that particular window size.

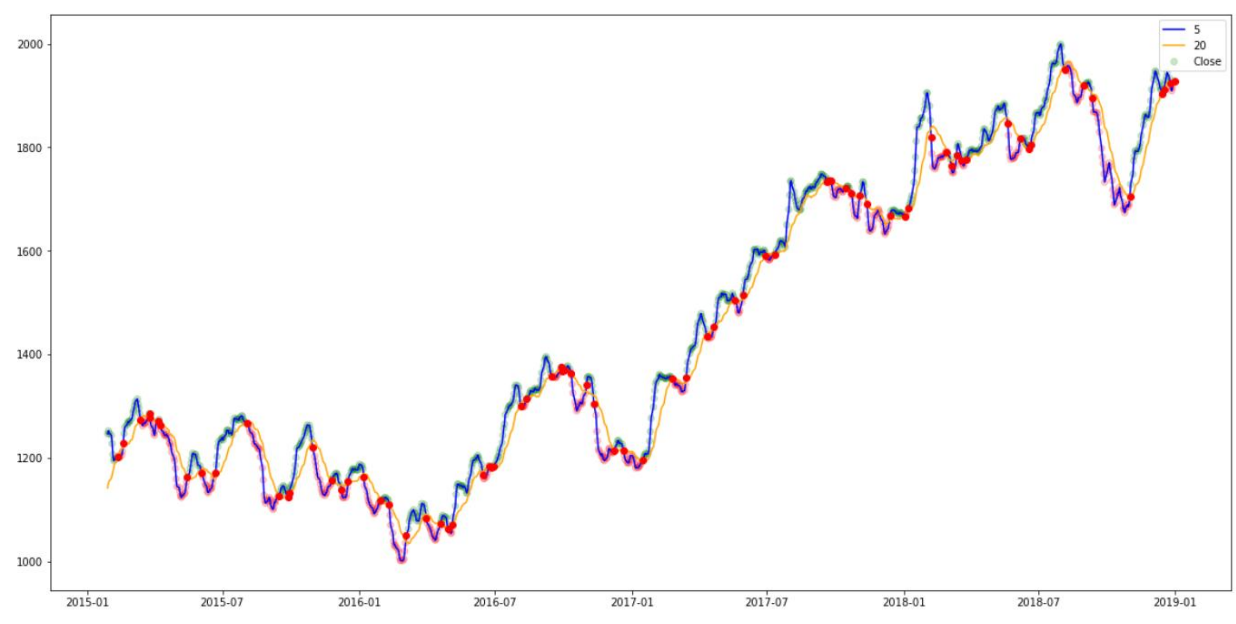
a. Shorter Window : 1 day Longer Window: 4 days



Net Profit = 188.14 base points
Accuracy = 51.39%

b. Shorter Window : 5 days   Longer Window : 20 days

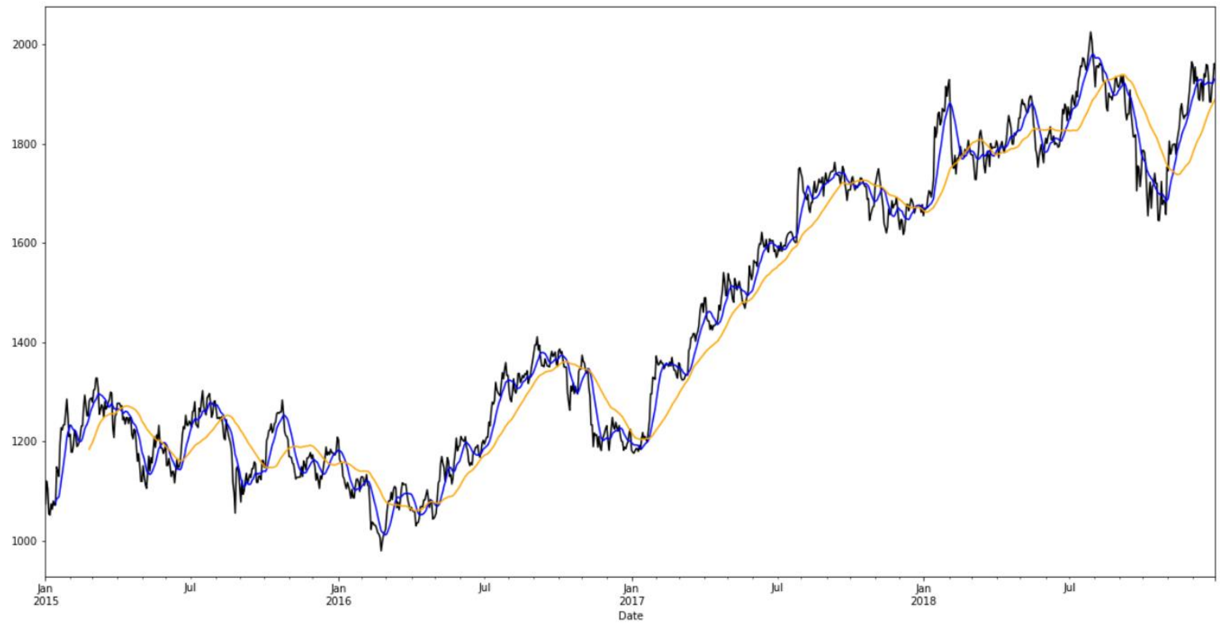


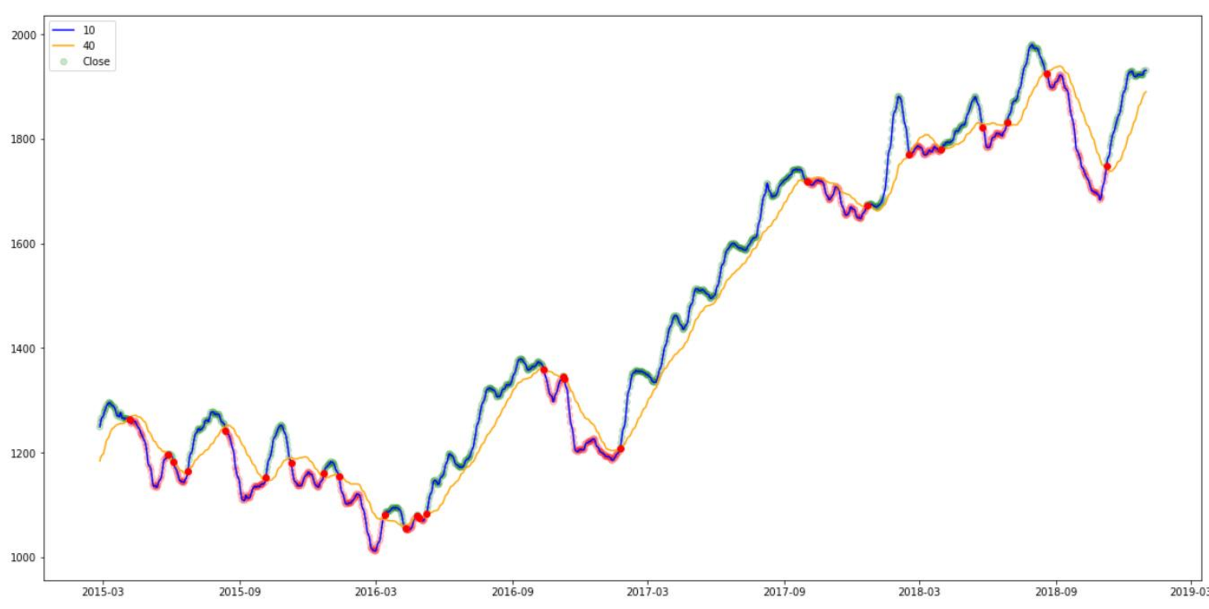


Net Profit = -1046.7 base points

Accuracy = 55.8%

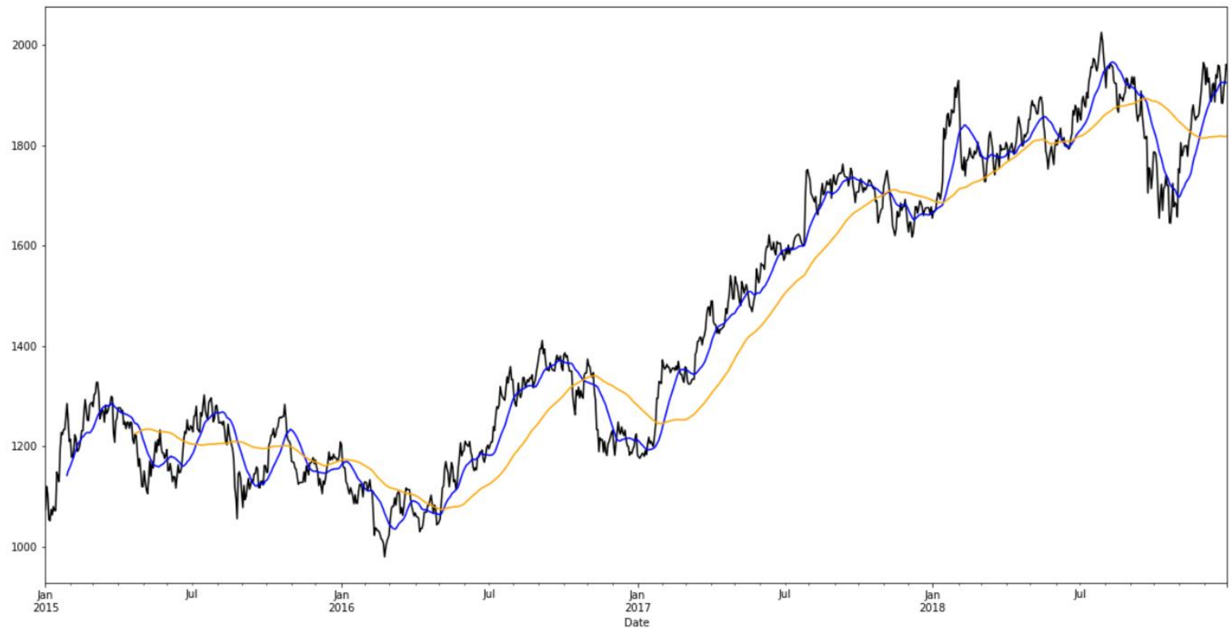
c. Shorter Window: 10 days Longer Window: 40 days

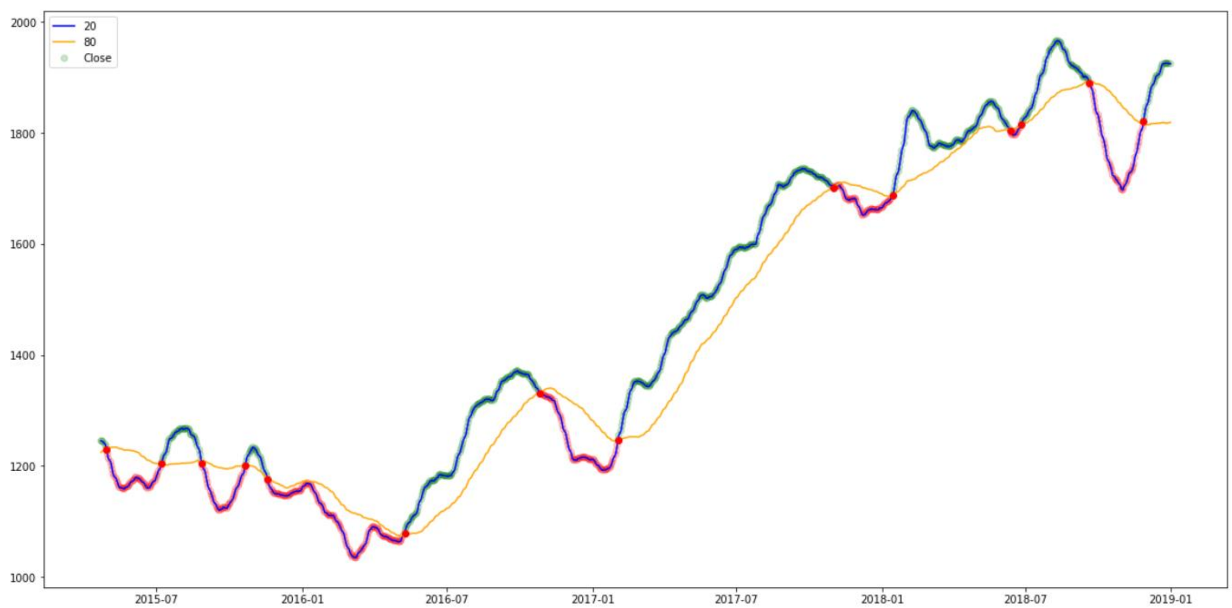




Net Profit = 264.95 bp
Accuracy = 52.34%

d. Shorter Window: 20 days Longer Window: 80 days

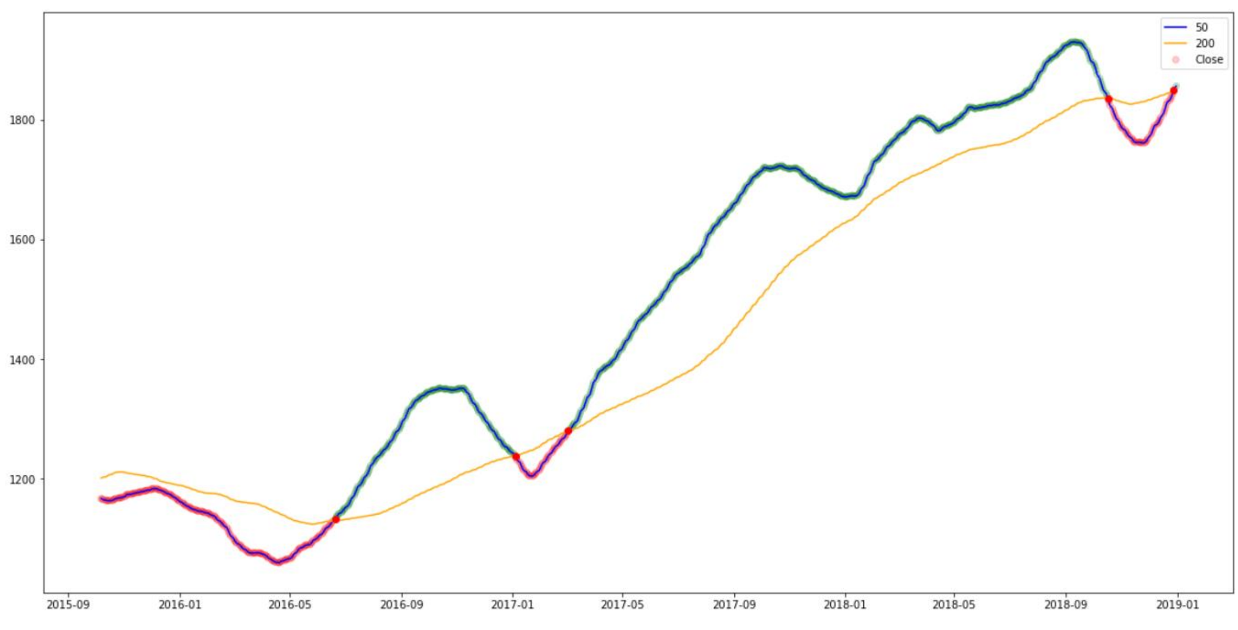




Net profit = -439.67 bp

Accuracy = 54.6%

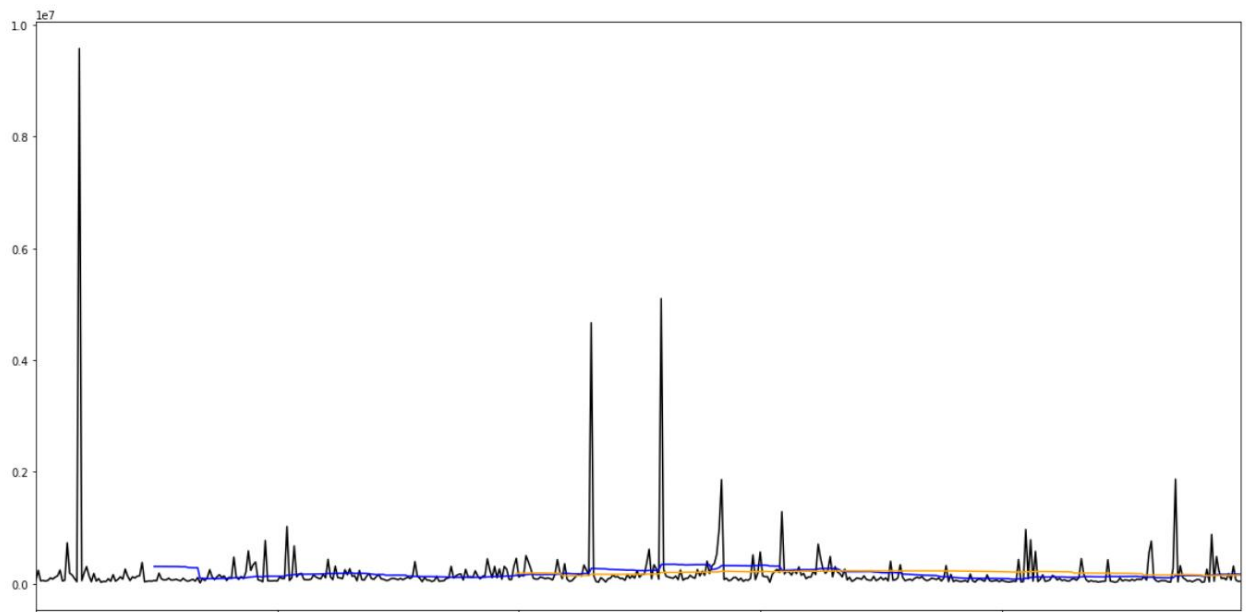
e. Shorter Window: 50 days Longer Window: 200 days



Net profit = -7.3 bp
Accuracy = 52.55%

The above analysis was repeated for the stock prices of Kotak Mahindra Bank. The price data for 500 trading days was collected directly from BSE's website. The reason for selecting Kotak Mahindra was that the daily percentage change in the prices of HDFC Ltd. was less than 1% for the entire data, but the same for Kotak Mahindra was on certain days exceeding even 300%. Thus, the main aim was to test the algorithm for both a low and a high volatility stock and quantify the efficacy of the algo.

The maximum profit was observed for a window size of 50 for the short term trend and 200 for the long term trend. However, even in this situation the algo reported a net loss of 727881 base points and with an accuracy of 49%.



The blue line corresponds to the 50 day MA and the orange one corresponds to the 200 day MA.

3. **Observations and Inference:** As is evident from the above simulation results, the MA algorithm provides a decent trading decision for stocks with low volatility i.e. stocks which do not show extreme and sudden change of prices on a day to day basis. However, in situations where there is some extraordinary event that causes a sudden change in the stock price (which may even correct itself in a couple of days), the MA algorithm fails to provide a timely decision for avoiding losses.
   The main reason behind this shortcoming is the inherent delay in the trend lines as compared to the stock price. Thus it takes the trend lines some time before it incorporates all the fluctuations in the actual stock price.
   Thus, the MA algorithm is a good algorithm to use in a low volatility market.

B. **Support and Resistance**

1. **Algorithm:** This algorithm finds certain price regions in the stock's price graph which act as a lower and upper bound to the fluctuations of the stock price. The price level which acts as a lower bound is referred to as "Support" and the upper bound is referred to as "Resistance". Basically, a support level supports the rise in the stock price and a resistance level resists the rise in the price. Thus, by finding the support and resistance levels in a price chart, one can make predictions as to whether the price in the coming future will increase or decrease.
If the current price is near to the support region, then there is a high chance that the price will rise in the near future and if it is near the resistance region, then it is most likely to fall. This is the main strategy used for buying and selling using Support and Resistance Algorithm.

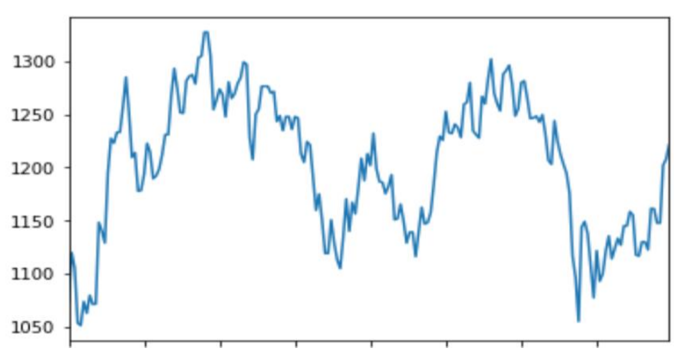
2. **Implementation Details:** The most important step before using the Support and Resistance algorithm is to define exactly what the characteristics are that are desirable in a good Support or Resistance level. There is no fixed set of characteristics that make a region good or bad for Support or Resistance, but there are certain properties that are widely used by market analysts, which include:

a. The more the number of times the price touches a region and bounces up (falls down), the better is the region suited to be a Support (Resistance) level.

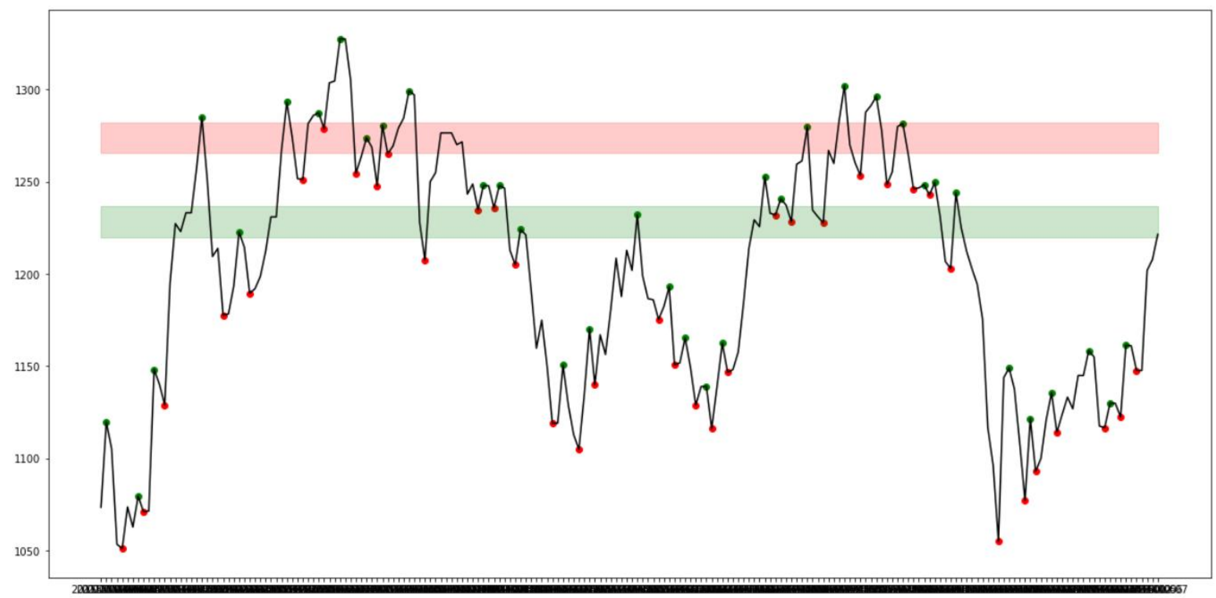
b. If the price has touch a level in recent times, then that level is better than the one that the price touched a long time ago.

c. The lower (higher) the region, the better it is for being a Support (Resistance) level.

For testing the algorithm, closing price data for HDFC Ltd. from 1st Jan 2015 to 7th Oct 2015, i.e. 200 trading days was collected and used.
The data is shown as follows:

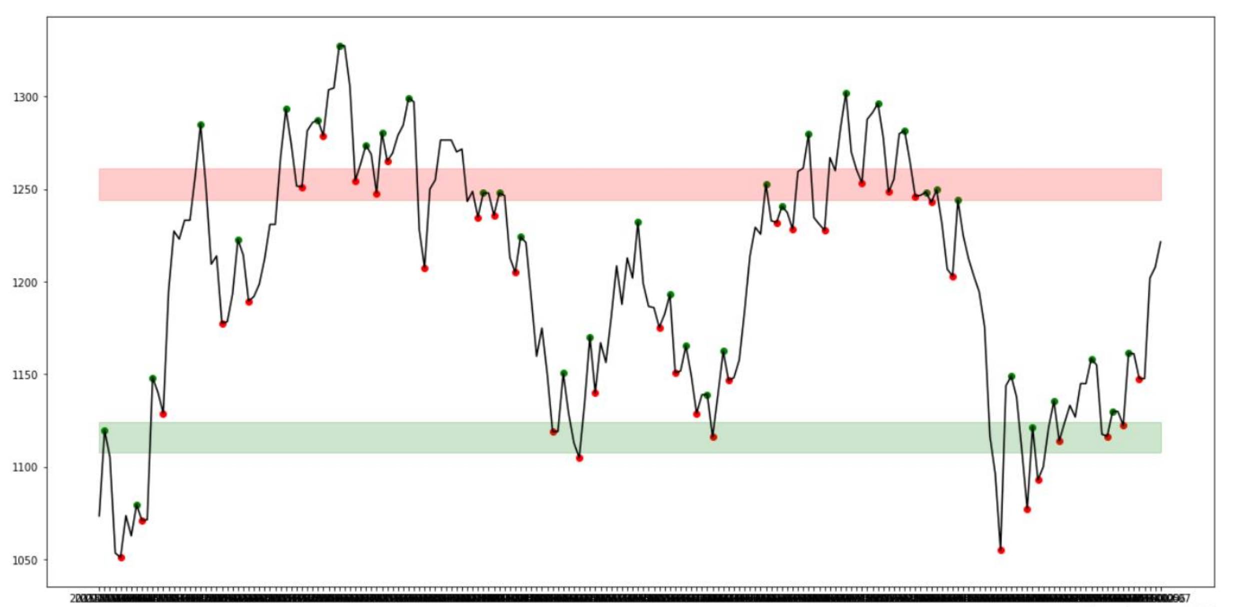


To find a good Support and Resistance level, I first found regions based solely on the number of times the prices touches it and bounces up (falls down) [Criteria 'a' mentioned above]. The result was:



The Support and Resistance levels are shown with green and red colors respectively.

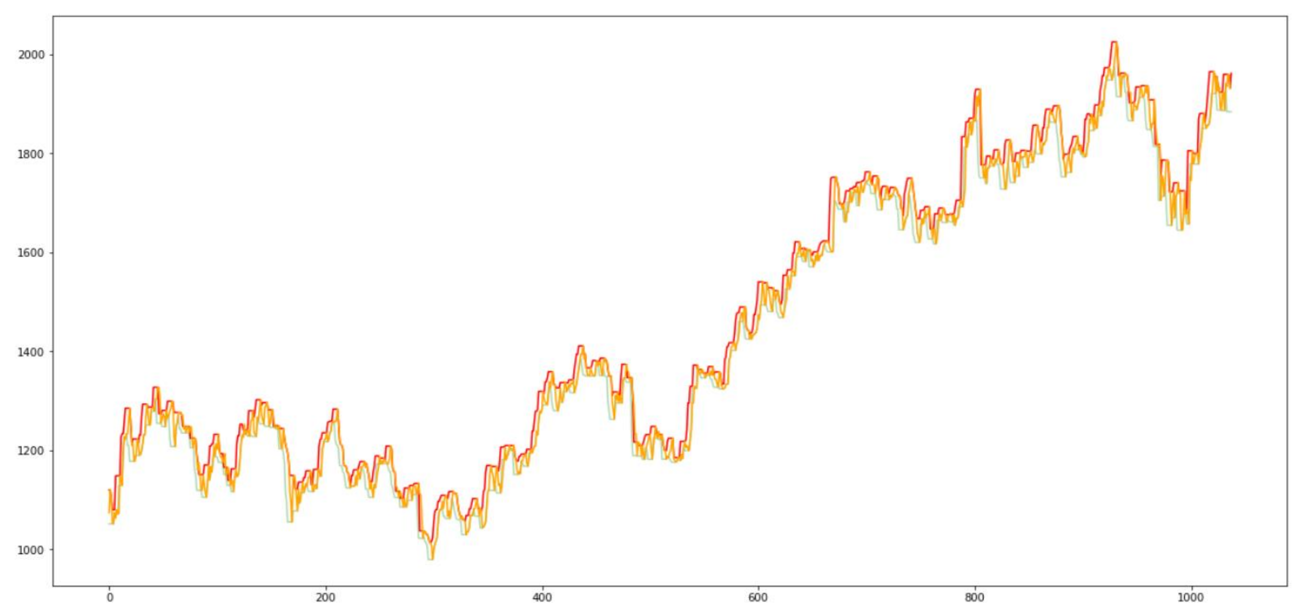
As is evident, this identification has room for improvement. Therefore, I also included Criteria 'b' and 'c' into my algo and the result was:



As is evident from the above chart, the Support and Resistance levels do provide a pretty good lower and upper bound on the price fluctuations. The above result proves that the algorithm is able to find good Support and Resistance levels. Once the algorithm was being able to identify the Support and Resistance levels, it was time to test it out on the complete dataset. The closing price data for HDFC Ltd. from 1st Jan 2015 to 31st Dec 2018, i.e. 1043 trading days (the same dataset used for MA testing) was used for testing.

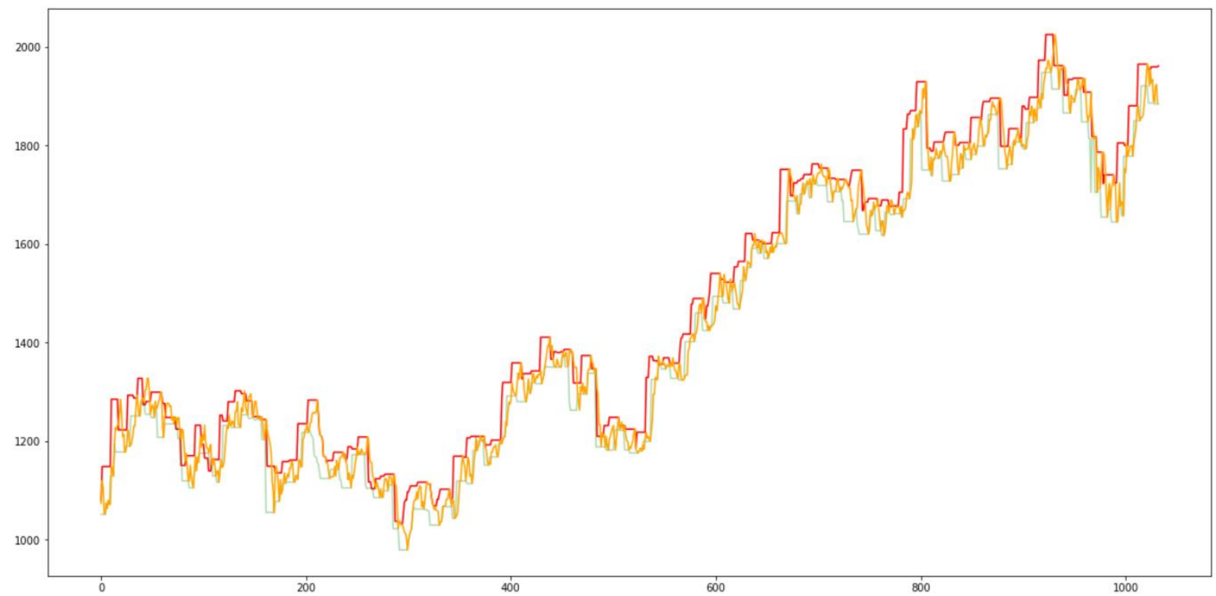
Now, to find the Support and Resistance levels, we need to specify the time frame that should be considered for the calculation. Similar to the MA algorithm, I simulated this algo as well by varying the size of the time window under consideration to decide the best window of operation. The algorithm was tested for window sizes of 5,10,20,50 and 100 days. The support and resistance levels along with the net profit for each of the window are shown in the following charts. The Support levels are colored green, Resistance in red and the closing prices in orange.

a. Window = 5 days



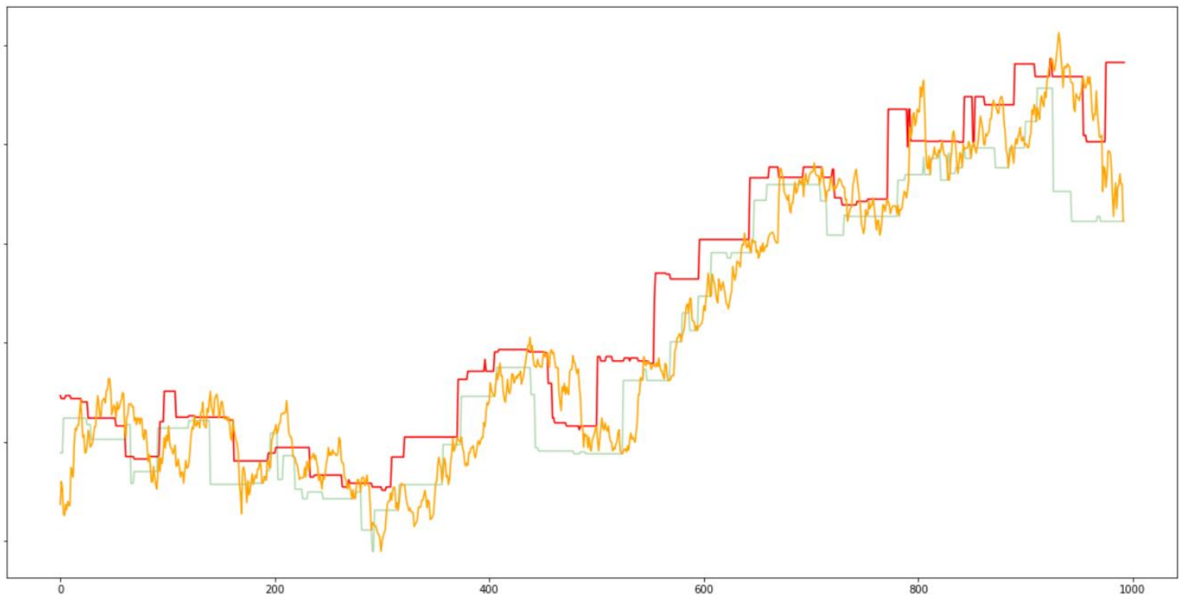
Net Profit = 725.1 bp

b. Window = 10 days Net Profit = 1215.37 bp



c. Window = 20 Net Profit = 240.17

d. Window = 50 days Net Profit = -271.56



e. Window = 100 days Net Profit = 465.35 bp

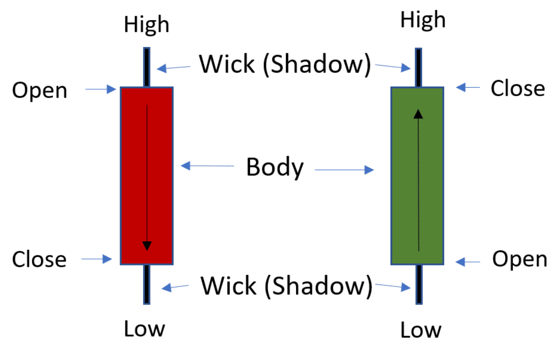
3. **Observations and Inference:** The simulation results showed that the maximum profit was generated for a window size of 10 days and generating a profit of 1215.37 bp. This shows that a very long window is not as useful as current affairs are much more like to be predicted using recent developments than something that happened an year ago. Also, a very small

window is not that useful as it cannot take into account events that happened even a few days ago. Thus, Support and Resistance algorithm with a balanced window size is able to consistently generate profits.

C. Candlestick Patterns

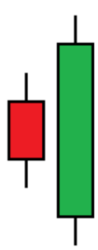
1. **Algorithm:** Candlestick patterns are one of the most prominent charting techniques used by almost all market analysts to predict upcoming trends. The concept is to find certain special patterns in the candlestick price charts and predict the future trend based on the type of the pattern observed.

   The structure of a candlestick is shown as follows:



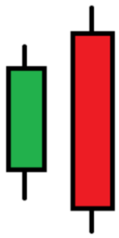
   There are several candlestick patterns which are used; I chose the following six to be used for my algo:
   a. Bullish Engulfing
   b. Bearish Engulfing
   c. Bullish Harami
   d. Bearish Harami
   e. Evening Star
   f. Morning Star

a. **Bullish Engulfing:**



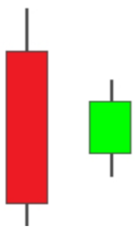
This pattern is characterized by a bullish (green) candle completely 'engulfing' the previous day's bearish (red) candle. This pattern when appears in a falling or bearish market, signifies a possible reversal of the market and the rallying up of the market in the near future.

b. **Bearish Engulfing:**



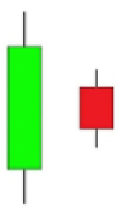
This is the complete opposite of the previous pattern. Here, a bearish candle completely engulfs a bullish candle. When this pattern appears in a rising or bull market, then there is a high chance of market reversal and bear market in the near future.

c. **Bullish Harami:**



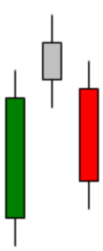
This pattern consists of a small bullish candle preceeded by a bigger bearish candle which completely covers it. When this pattern appears in a bearish market, then it is a strong signal that the trader's sentiments are changing and there is going to be a bullish market in the near future.

d. **Bearish Harami :**



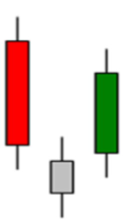
This pattern is the opposite of the previous pattern. Here a small bearish candle is covered by a bullish candle. When this pattern appears in a bullish market, then there is a high chance of market reversal and a bear market in the coming future.

e. **Evening Star :**



This pattern signifies the change of the market trend from bullish to bearish. The middle candle can be bullish or bearish and thus is shown in grey.

f. **Morning Star :**



This pattern is the opposite of the previous one and signifies the market reversal form a bearish to a bullish market.

2. **Implementation Details:** The algorithm was written which could identify the above discussed candlestick patterns using the Open, High, Low and Close price data of a stock. The stock used

for testing the algorithm is of HDFC Ltd. from 1st Jan 2015 to 31st Dec 2018(same dataset used fot the previus two algorithms).

The results along with the accuracy of prediction for all the patterns as shown below:

a. **Bullish Engulfing**

The green points correspond to the points which the algorithm identified with the specific pattern and was correct and the red ones correspond to the points identified with the pattern but the market trend did not follow the prediction.

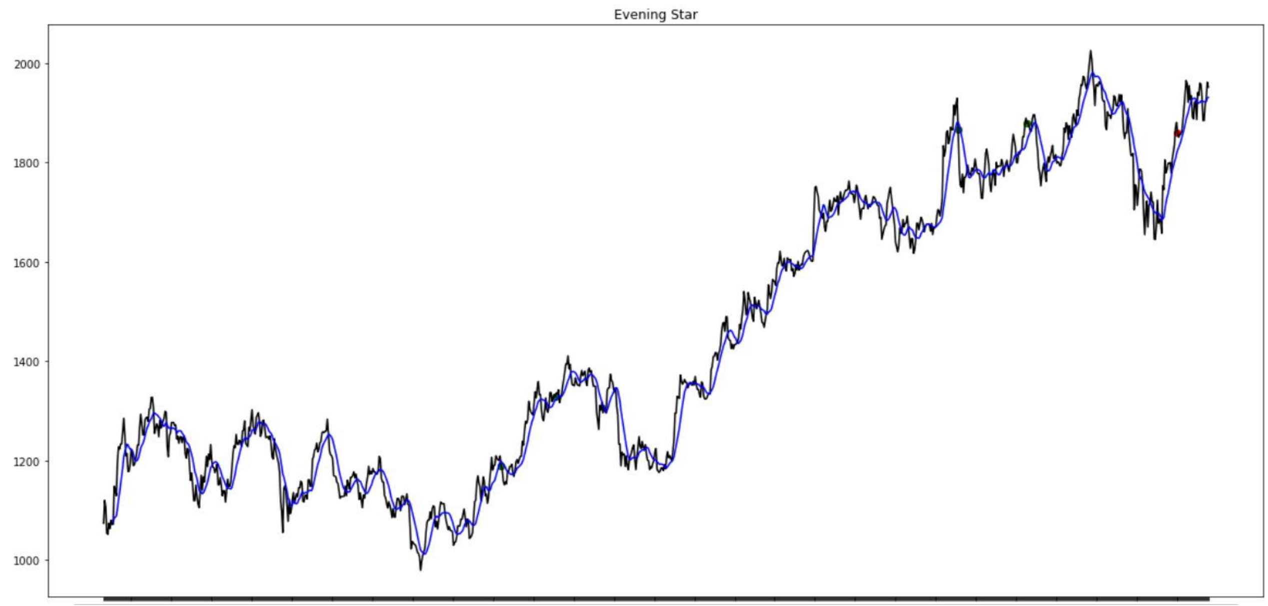
Accuracy for Bullish Engulfing = 72.7%

b. **Bearish Engulfing:** Accuracy = 69.5%

c. **Bullish Harami :** Accuracy = 72.7%

d. **Bearish Harami :** Accuracy = 76%

e. **Evening Star :** Accuracy = 80%



f. **Morning Star :** Accuracy = 71.1%

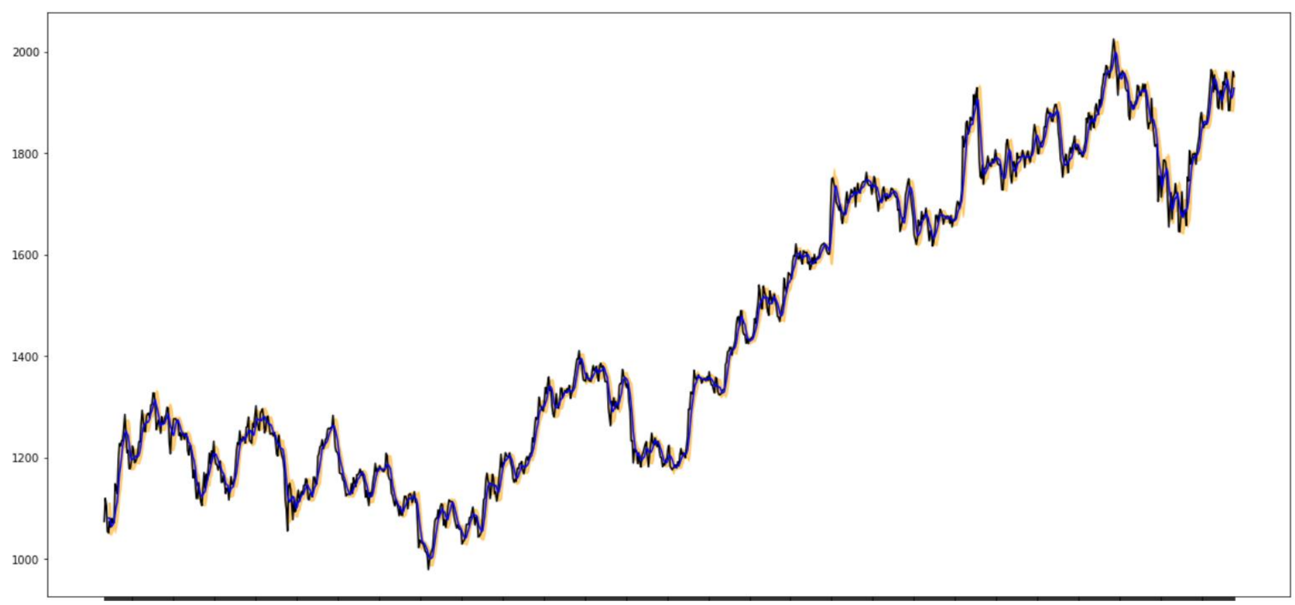
Observations and Inference: The chosen candlestick patterns provide really good prediction accuracy and thus were chosen to be added to the final prediction bot.

D. **<u>Bollinger Bands:</u>**

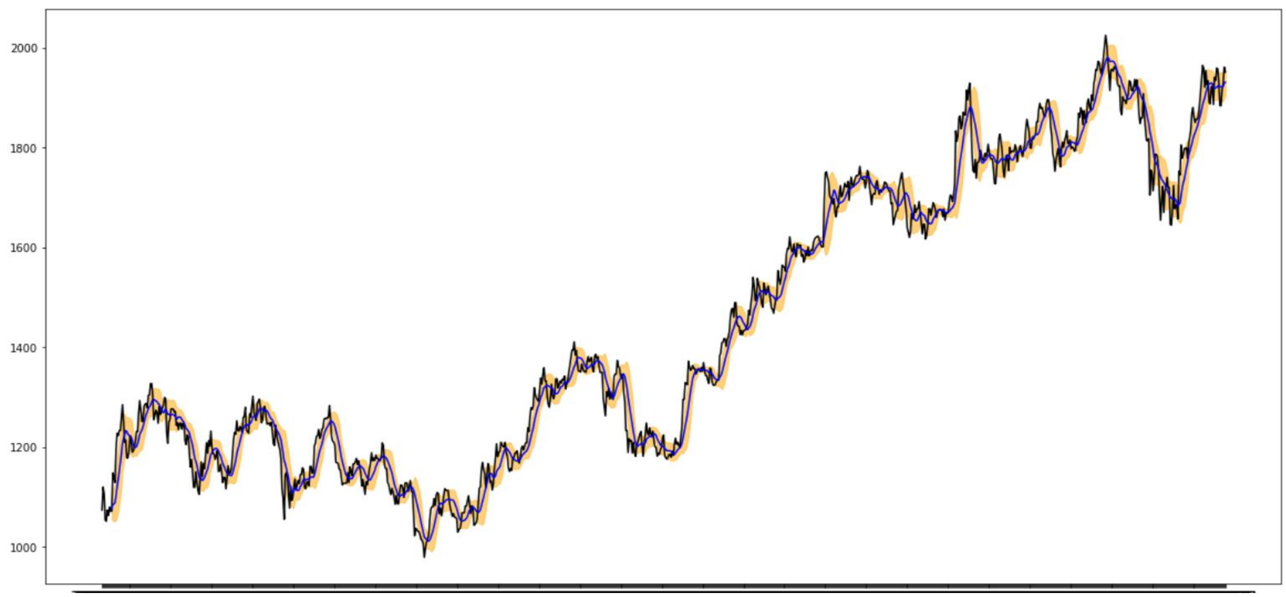
1. Algorithm: In the MA algorithm, the main economic assumption was that the stock prices follow a trend and are supposed to be equal to the value predicted by the trend. However, this assumption does not take into account the inherent volatility of the stock. This drawback can be remedied using the Bollinger Band Algorithm. It is similar to the MA algorithm but here the stock price is supposed be lie close to the trend line rather than on the trend line. This takes into account the volatility of the stock and says that the stock is supposed to lie within the mean +/- one standard deviation.
Whenever the stock price becomes more than mean + std. deviation, it is time to short the stock and when it goes below mean – std. deviation, it is time to buy the stock.

2. **Implementation Details:** After implementing the algorithm, I tested it using the HDFC Ltd. dataset for 1043 trading days. Just like the MA algorithm, the main parameter to be decided in this algorithm is the window size which will be considered for finding the mean and the volatility. Thus, the testing iterated the algorithm by varying the window sizes as 5,10,20,50 and 100 days. The results along with the net profit for each window size are shown below. The blue line corresponds to the MA trend line and the orange band signifies the Bollinger Band.

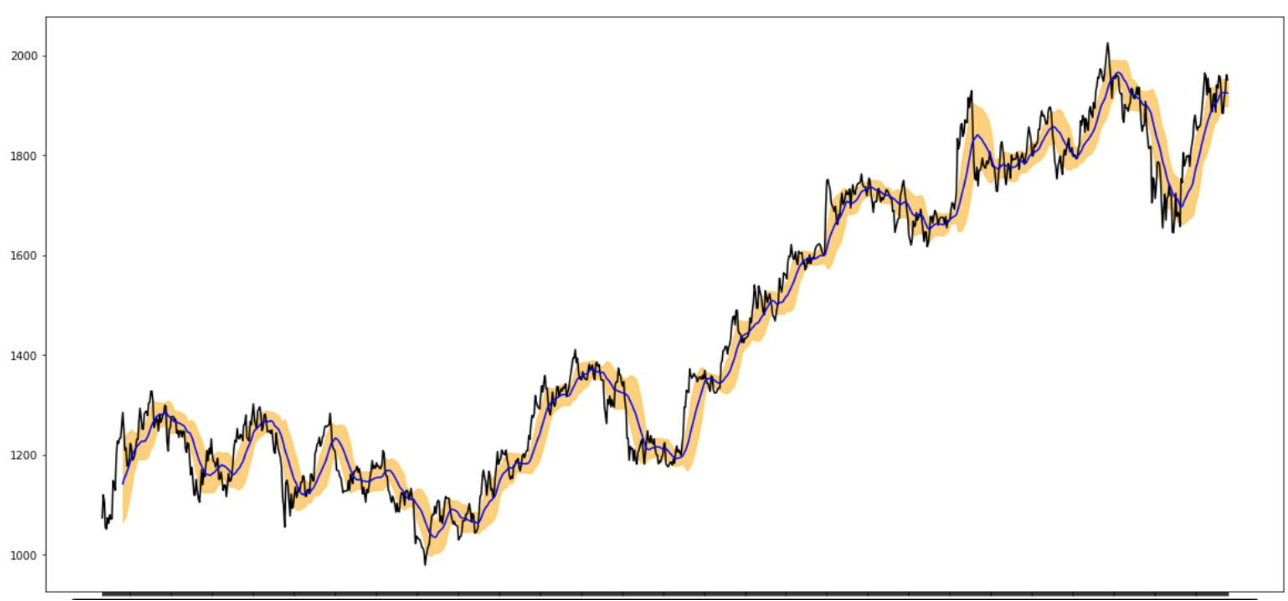
a. Window = 5 days   Net Profit = -401 bp



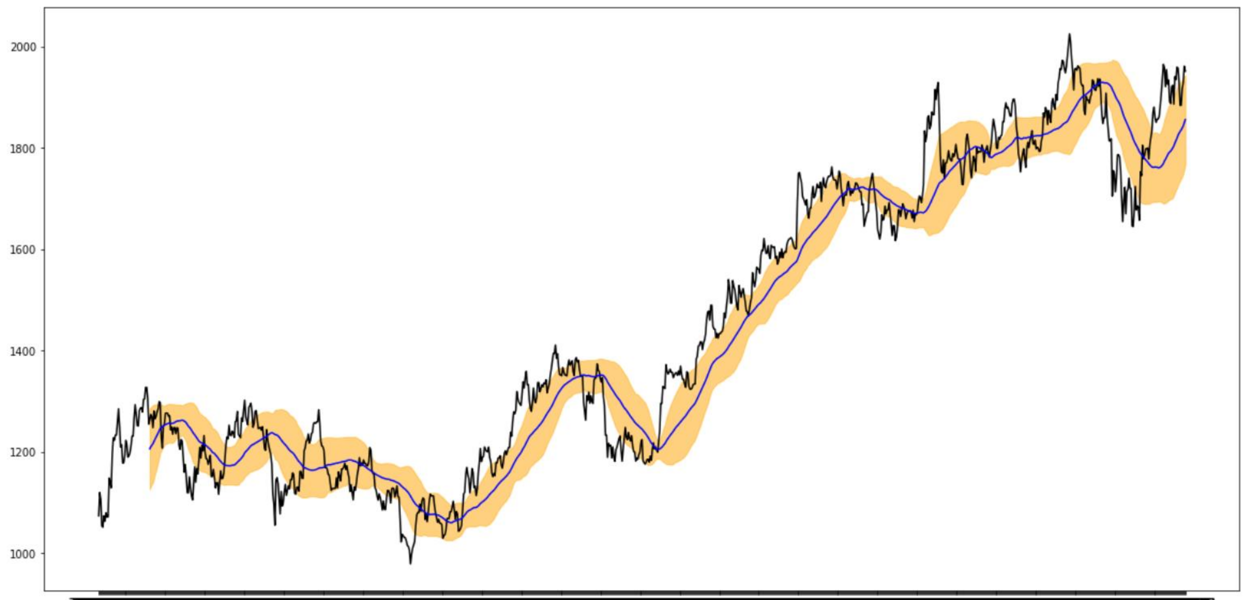
b. Window = 10 days   Net profit = -281 bp



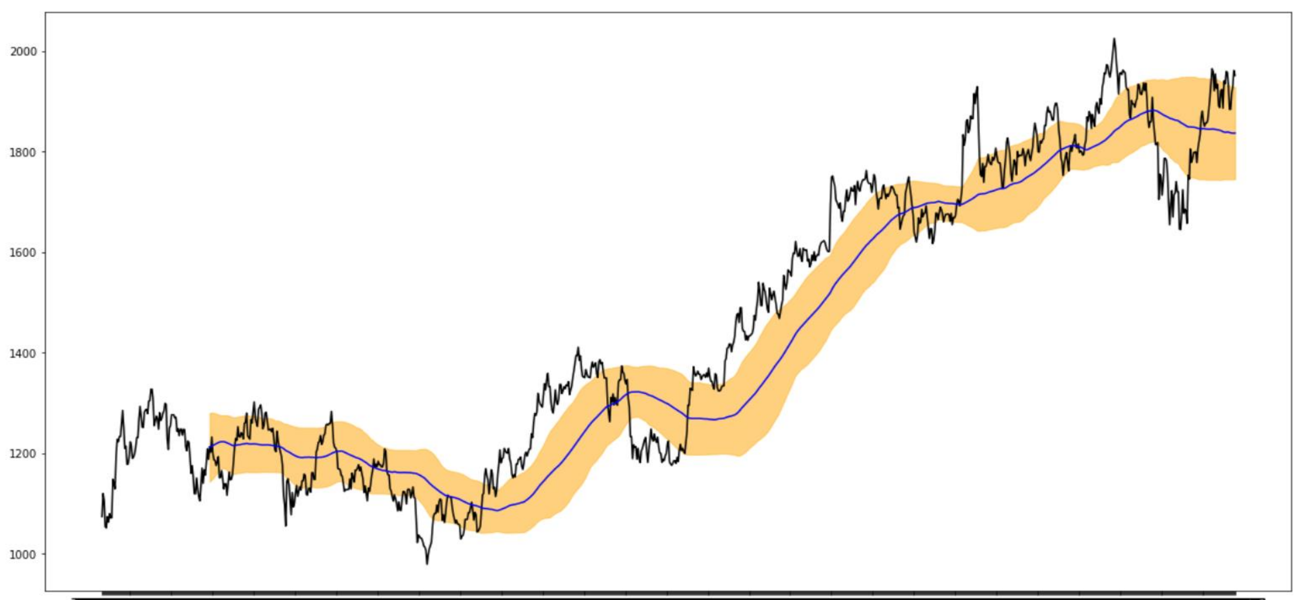
c. Window = 20 days   Net Profit = -287 bp



d. Window = 50 days   Net Profit = 465.6 bp



e. Window = 100 days   Net Profit = 1031.55 bp



The experiment was rerun on a smaller dataset which was the first 200s trading days of this dataset. This was done to see the window size is affected by the size of the data under consideration. The profits recorded for each of the window sizes is recorded below:

Window Size	Net Profit(bp)
5	155.87
10	83.66
20	24.85
50	271.23
100	134.76

The maximum profit was observed for a window size of 50.
The algorithm was rerun on data from Kotak Mahindra Bank. The results are recorded below:

Window Size	Net Profit(bp)
5	26128651
10	33484095
20	23016919
50	11301939
100	5417552

3. **Observations and Inference:** The algorithm generated really good profits for both HDFC Ltd. and Kotak Mahindra Bank. The performance was much better than the MA algorithm thus proving the importance of including volatility into the prediction.

Implementation and Testing of the Final Bot

After the implementation and testing of the individual algorithms was over and the algorithms were giving good results, the algorithms had to be integrated so as to make the final prediction bot. To do so, I applied both bagging and boosting methods and compared the results. Bagging a set of algorithms means combining the results of the individual algorithms giving equal weightage to each algorithm in the final decision process. Whereas, in boosting, we give certain algorithms priority over certain other ones. The final bot was first made by bagging the individual algorithms and tested on the data from a basket of banking stocks listed on the BSE. The accuracies for each of the dataset were recorded. To set a benchmark for comparison, I also implemented a control bot which would randomly decide whether to buy, short or hold the stock. As expected the accuracy of the random bot varied from 45% to 51% over the various datasets. This can be considered equivalent to a novice trader with no experience in stock trading. Now, the results for the different banking stocks are tabulated below:

Stock Name	Prediction Accuracy
SBI	71%
HDFC Ltd.	65.7%
Kotak Mahindra Bank	90.5%
IndusInd Bank	70.7%
ICICI Bank	64.1%

Thus the prediction bot was giving a really high accuracy as compared to a novice trader.

To further improve the results, I applied Boosting to the prediction bot and the results were:

Stock Name	Prediction Accuracy
SBI	74.5%
HDFC Ltd.	66.5%
Kotak Mahindra Bank	90.5%
IndusInd Bank	74.7%
ICICI Bank	65.6%

Thus, the final prediction bot was completed and tested.

Conclusion

We started with the question, do the prediction algorithms actually work or is speculation in the stock market equivalent to gambling. Now, we can safely assert that prediction algorithms do work and speculation in the stock market is a lot more than gambling. They may not provide accurate results under all circumstances but they are able to generate consistent profits under most conditions. Each algorithm has its own merits and demerits but if we combine certain algorithms together then we are able to generate predictions with a really good accuracy.

In this project, I started with the basics of finance and economics, learnt about the financial algorithms used in the markets, converted them into computer algorithms, implemented and finally tested those algorithms. By making certain observations along the way and modifying the algorithms to suit the project's needs, I was finally able to combine the algorithms together and make a stock prediction bot which gave impressive results on a basket of banking stocks.

In the end, this project was able to meet all the aims and objectives that were proposed in the beginning and has been able to provide decent prediction accuracies over a variety of test cases.

Overall, this project has been a great learning experience for me. I got to learn a lot about the financial markets and economics. I also had the opportunity to discuss a lot of new ideas with my supervisor which helped in giving me new perspectives of thinking about the various strategies used in the market and also how the investor's sentiments can be affected by certain events and effect that it has on the market. Also, implementing the algorithms and tweaking the parameters gave me deeper insights as to how the parameters shape the prediction outcome and how these parameters can be used to fine tune our predictions to a variety of market conditions. It also allowed me to explore a lot about the world of finance and behavioral economics.

In conclusion, doing this project was a great experience and I hope to continue working on similar projects in the future.
