**Stock Market Prediction Analysis**

Course: PGCBAMD

Name of the faculty: Prof. Pitabas Mohanty

Area: Financial Analytics

|  |
| --- |
| **Submitted by**: Mr. Rajib Mandal Roll No.:- TA17002  Name of center: -Camac Street(Kolkata) |

# Executive Summary

Stock market prediction is the act of trying to determine the future value of a company stock or other financial instrument traded on a financial exchange. The successful prediction of a stock's future price will maximize investor’s gains. This project proposes a machine learning model to predict stock market price. The proposed algorithm integrates time-series model known as Long Short-Term Memory to predict the daily stock prices. Proposed model is based on the study of stocks historical data and technical indicators

# Project Objective

For that goal, a prediction model was built, and a series of experiments were executed and theirs results analyzed against a number of metrics.

We use an online learning algorithm that utilizes a kind of recurrent neural network (RNN) called Long Short Term Memory (LSTM), where the weights are adjusted for individual data points using stochastic gradient descent. This will provide more accurate results when compared to existing stock price prediction algorithms. The network is trained and evaluated for accuracy with various sizes of data, and the results are tabulated.

# Data Set Information

The daily stock price data of state-owned oil companies (IOC, ONGC, GAIL, BPCL, HINDPETRO) spans from 2015 to the end 2018 is collected from National Stock Exchange of India (https://nseindia.com/products/content/equities/equities/eq\_security.htm). The dataset consists of 4915 observations and 15 features.

The definitions for all the data attributes given below.

|  |  |  |
| --- | --- | --- |
| **#** | **Name** | **Description** |
| 1 | Symbol | Symbol for each company given by NSE. |
| 2 | Series | Series Symbol for each security given by NSE; e.g. "EQ" for common stock, "N1" for first debenture issue, "W" for warrants, etc. Once a symbol and a series have been specified, a security is uniquely known |
| 3 | Date | This gives the Trade date. |
| 4 | Prev Close | The previous day Closing Price |
| 5 | Open Price | The opening price of the day. |
| 6 | High Price | The highest traded price of the day. |
| 7 | Low Price | The lowest traded price of the day. |
| 8 | Last Price | The last traded price of the day. In general, this need not be equal to the official closing price because the official closing price is calculated using a variety of rules (e.g. averaging of trades over the last 30 minutes), etc |
| 9 | Close Price | This is the official closing price reported by NSE. |
| 10 | Average Price | Daily average price of the stock |
| 11 | Total Traded Quantity | The number of shares traded in the day. |
| 12 | Turnover |  |
| 13 | No. of Trades | Number of trades which took place in the normal market (i.e. excluding trades in the auction market. The average trade size is field 8 divided by field 10 (in number of shares) or field 9 divided by field 10 (in rupees). |
| 14 | Deliverable Qty |  |
| 15 | % Dly Qt to Traded Qty |  |

# TECHNICAL ANALYSIS

Technical analysis is the study of market action using past prices and trading volumes for the purpose of forecasting future price trends. Technical analysis assumes that stock prices move in trends, and that the information which affects prices enters the market over a finite period of time, not instantaneously.

Four technical indicators are calculated from the raw datasets

**Relative Strength Index (RSI):** A technical momentum indicator that compares the magnitude of recent gains to recent losses in an attempt to determine overbought and oversold conditions of an asset. The formula for computing the Relative Strength Index is as follows.

RSI = 100 - [100 / (1+RS)]

The RSI oscillates between zero and 100. Traditionally the RSI is considered overbought when above 70 and oversold when below 30. Signals can be generated by looking for divergences and failure swings. RSI can also be used to identify the general trend.

From the below graph(**Figure-1)** we can find that

The stock IOC overbought during March 2015, May 2016, September 2017 since RSI is more than 70.

The stock IOC oversold during January 2017, May 2018, December 2018 since RSI is less than30.

The stock BPCL overbought during January 2015, April 2015, October 2017, since RSI is more than 70.

The stock BPCL oversold during August 2016, June 2017, JULY 2017 Jan 2018 since RSI is less than 30 .

The stock ONGC overbought during February 2015, April 2016 and May 2018 since RSI is more than 70.

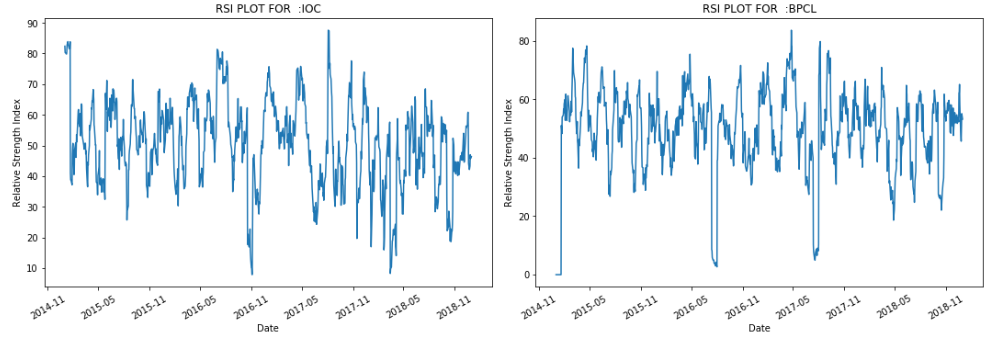
The stock ONGC oversold during December 2016, January 2016, October 2016, October 2018 since RSI is less than 30

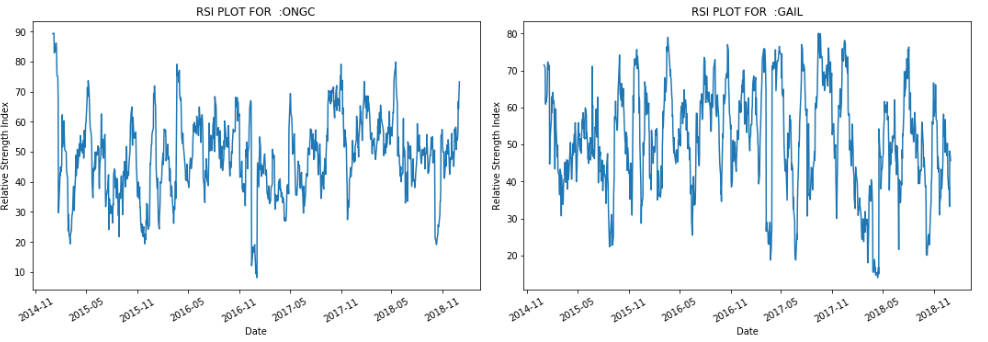
The stock GAIL overbought during November 2015, April 2016, September 2017, November 2017 since RSI is more than 70.

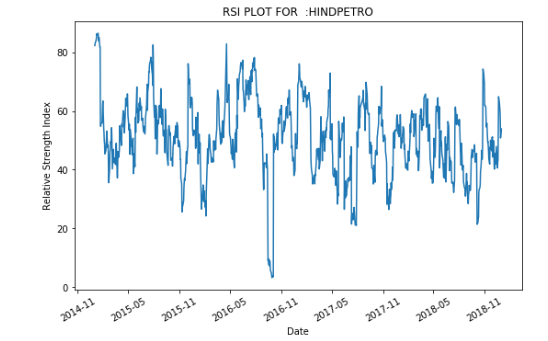
The stock GAIL oversold during June 2015, June 2016, January 2017, June 2017, February-May 2018 since RSI is less than 30 .

The stock HINDPETRO overbought during June-September 2015, July 2016, September 2017 since RSI is more than 70.

The stock HINDPETRO oversold during September-October 2016, October 2018 since RSI is less than 30.







**Figure-1**

**Moving Average Convergence/Divergence (MACD):** This function calculates difference between a short and a long term moving average for a field. The formulas for calculating MACD and its signal as follows.

MACD = [0.075\*EMA of Closing prices] – [0.015\*EMA of Closing prices]

[0.15\*EMA of closing prices]

Signal Line = 0.2\*EMA of MACD

The MACD is calculated by subtracting the 26-period EMA from the 12-period EMA

MACD crossing above zero is considered bullish, while crossing below zero is bearish. Secondly, when MACD turns up from below zero it is considered bullish. When it turns down from above zero it is considered bearish.

There is bullish trend during Aug 2016 for **IOC** and bearish trend during November 2016 and May 2018.

There is bullish trend during April 2016 for **BPCL** and bearish trend during August 2016 and July 2017.

There is bullish trend during February 2015 for **ONGC** and bearish trend during December 2016

There is bullish trend during October 2016 for **GAIL** and bearish trend during April 2017.

There is bullish trend during July 2015 and June 2016 for **HINDPETRO**  and bearish trend during October 2016 and April 2016.







**Figure-2**

**Money Flow Index (MFI):**

It is an oscillator that uses both price and volume to measure buying and selling pressure. It measures the strength of the money flowing in and out of a security. Money Flow is positive when typical price rises (buying pressure) and negative when the typical price declines (selling pressure

The indicator is calculated using 14 periods of data. An MFI reading above 80 is considered [overbought](https://www.investopedia.com/terms/o/overbought.asp) and an MFI reading below 20 is considered [oversold](https://www.investopedia.com/terms/o/oversold.asp).

The formula for MFI is as follows:

Money Flow (MF) = Typical Price \* Volume.

Money Ratio (MR) = (Positive MF / Negative MF).

MFI = 100 – (100/ (1+MR)).

The stock IOC overbought during March 2015, May 2015, April 2017 since MFI is more than 80.

The stock IOC oversold during June 2015, January 2016, April 2018, Aug 2018 since MFI is less than 20.

The stock BPCL overbought during October 2015, August 2016, June 2016, November 2016, August 2017, October 2017, since MFI is more than 80.

The stock BPCL oversold during June 2017, January 2018 since MFI is less than 20

The stock ONGC overbought during September 2016, since MFI is more than 80.

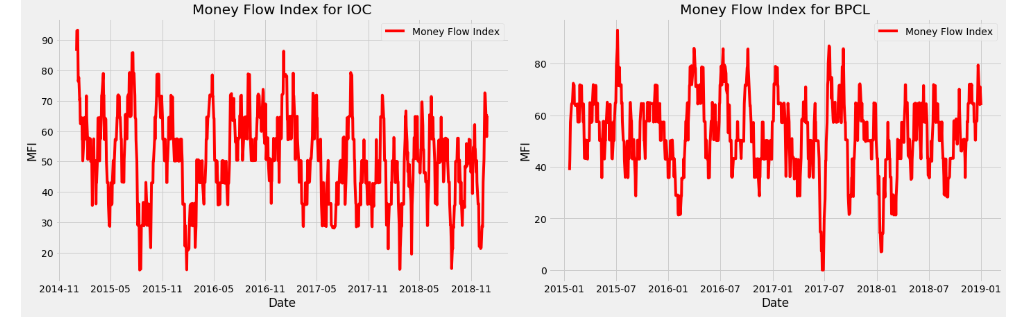
The stock ONGC oversold during July 2015, June 2017, July 2018 since MFI is less than 20

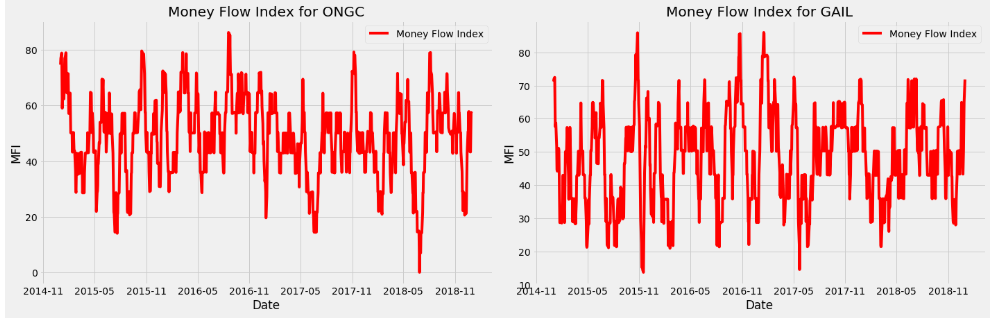
The stock GAIL overbought during November 2015, November 2016, February 2017 since MFI is more than 80.

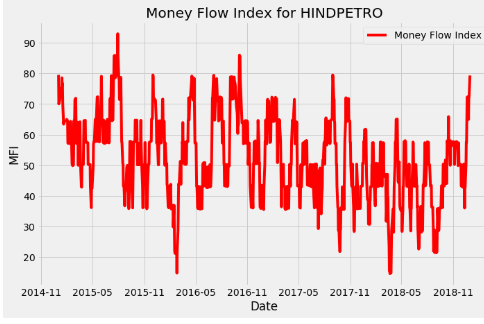
The stock GAIL oversold during November 2015, May 2017 since MFI is less than 20

The stock HINDPETRO overbought during June-September 2015, October 2016 since MFI is more than 80.

The stock HINDPETRO oversold during March 2016, April 2018 since MFI is less than 20







**Figure-3**

**Bollinger Bands** – Uses a simple moving average and plots two lines two standard deviations above and below it to form a range. This is used by traders using a mean reversion strategy where price moving above or below the bands is “stretched” and potentially expected to revert back inside the bands

Calculation: For calculation of Bollinger Bands, the following variables are required:   
  
a) Time Period denoted -- ‘N’   
  
b) Standard Deviation value -- ‘s’   
  
c) Three Bollinger bands or lines where:   
  
1. Moving Average Line or Middle Band for ‘N’ period MA (N). Refer average ‘Moving Average’ concept for calculation  
  
2. Upper Band or line wherein MA line is shifted up by price standard deviation for ‘N’ period multiplied by SD measure value ‘D’ (MA + D(s))   
  
3. Lower Band or line where in MA line is shifted below by price standard deviation for ‘N’ period multiplied by SD measure value ‘D’ (MA – D(s))

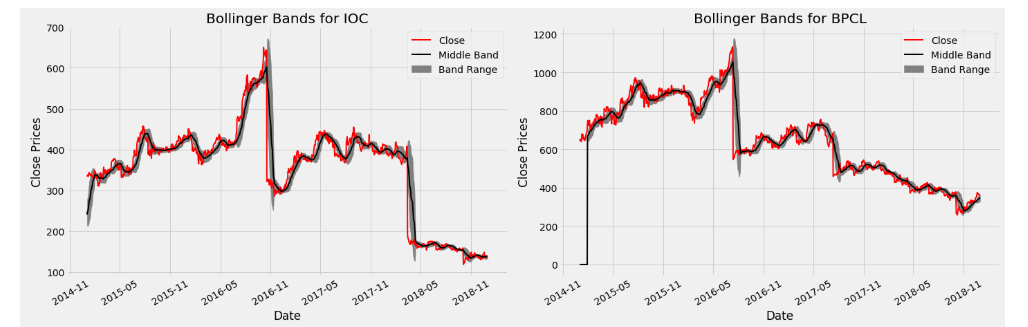
The upper and lower bands for the below graphs are calculated using two standard deviations +/- from a 20-day simple moving average.

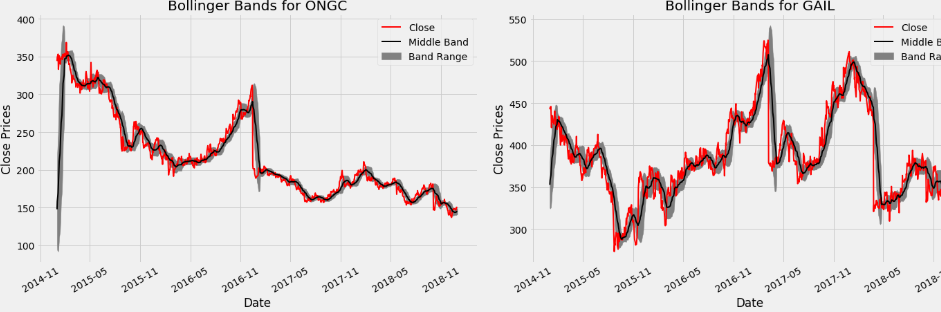
 The black line is middle band showing 20-day simple MA

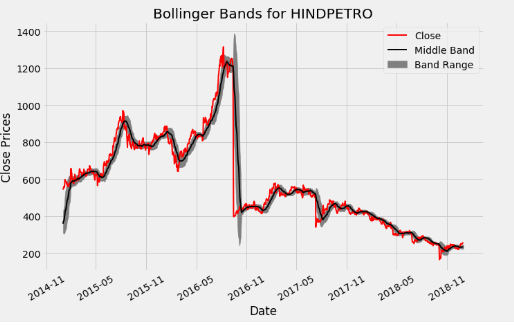
Since the bands are contracting, there are chances of sharp price changes as volatility drops

 When the price moves near the upper band, that shows an overbought market, and when the prices are nearer to the lower band, that signals an oversold market

From the below graph we can find that all the stock oversold during oct 2018 and overbought June 2016







Lowest Price)] \* 100

**Figure-4**

# Data Transformation

1. Following fields (Series,Prev Close,,Average Price,Total Traded Quantity, Turnover, No. of Trades,Deliverable Qty,% Dly Qt to Traded Qty) are dropped out during analysis
2. Scaling the data ('Open Price','High Price', 'Low Price','Last Price','Close Price') to the [-1, +1] range

# Construction of prediction model

We propose an online learning algorithm for predicting the end-of-day price of a given stock with the help of Long Short-Term Memory (LSTM), a type of Recurrent Neural Network (RNN).

The input data is split into training and test datasets; LSTM model will be fit on the training dataset, and the accuracy of the fit will be evaluated on the test dataset.

The LSTM network is constructed with one input layer having 256 neurons, 2 hidden layers , and one output layer (with one neuron).

After fitting the model on the training dataset, hyper-parameter tuning is done to choose the optimal values of following parameters

**Kernel\_initializer**: This defines the starting values for the weights of the different neurons in the hidden layer. We have defined this to be ‘uniform’, which means that the weights will be initialized with values from a uniform distribution

**Activation function**: in a neural network, the activation function of a node defines the output of that node as a weighted sum of inputs.

Activation function=

ReLU (Rectified Linear Unit) activation functions were tested to optimize the prediction model. ReLU has the following formula

**Y=max(0,x)**

**Batch size** : Number of samples that must be processed by the model before updating the weights of the parameters .Here batch size is 512

**Epoch** : A complete pass through the given dataset by the training algorithm .Here the epoch is 90

**Dropout**: A technique where randomly selected neurons are ignored during training i.e., they are “dropped out” randomly. Here the value is 0.3

**Cost function**: a sum of loss functions over the training set. An example is the Mean Squared Error (MSE), which is mathematically explained as follows

**Root Mean Square Error (RMSE)**: measure of the difference between values predicted by a model and the values actually observed. It is calculated by taking the summation of the squares of the differences between the predicted value and actual value, and dividing it by the number of samples. It is mathematically expressed as follows:

/N

Finally, we compile the classifier by passing the following arguments:

**Optimizer**: When building the LSTM model we used the optimizer Adam because of its high performance and fast convergence compared to other alternative optimizer and it was recommended to use it as default. When using the optimizer Adam we set the decay to 0. 3.

**Loss**: This defines the loss to be optimized during the training period. We define this loss to be the mean squared error.

**Metrics**: This defines the list of metrics to be evaluated by the model during the testing and training phase. We have chosen **RMSE** as our evaluation metric.

# Stock prediction algorithm using LSTM

Input: historical price data

Output: Prediction for stock prices based on stock price variation

1. Start
2. Stock data is taken and stored in numpy array of three dimension(N,W.F)

Where

N=Number of Training Sequence

W=Sequence Length

F=Number of feature of each sequence

1. A network structure is built with [l,a,b,l] dimensions ,where l is the input layer, a neurons in next layer, b neurons in the subsequent layer, and a single layer with a linear activation function
2. Train the constructed network on the data
3. Use the output of the last layer as prediction of the next time step
4. Repeat steps 4 and 5 until optimal convergence reached
5. Obtain predictions by providing test data as input to the network
6. Evaluate accuracy by comparing prediction with actual date

In general, smaller the RMSE value, greater the accuracy of the predictions made.

# Predictions and accuracy

Once the LSTM model is fit to the training data, it is used to predict the end-of-day stock price of an arbitrary stock.

The accuracy of the prediction model is estimated robustly using the RMSE (Root Mean Squared Error) metric. This is due to the fact that neural networks in general (including LSTM) tend to give different results with different starting conditions on the same data.

We then repeat the model construction and prediction several times (with different starting conditions) and then take the average RMSE as an indication of how well our configuration would be expected to perform on unseen real-world stock data. That is, we will compare our predictions with actual trends in stock price movement that can be inferred from historical data.

# Train and Test Score

Below table shows the average MSE and RMSE values after evaluating the training and

backtesting.

|  |  |  |
| --- | --- | --- |
|  | MSE | RMSE |
| Train Score | 0.00051 | 0.02 |
| Test Score | 0.00016 | 0.01 |

The mean square error for test set is 0.00016 which is very low and good as well.

# MSE Loss

Below graph shows the training and test loss of our LSTM model with optimal hyperparameter

setting, where the dropout is set to 30% and the decay of the optimizer Adam is set

to 0.3. From the figure we can see that, the MSE loss decrease for larger epoch values .

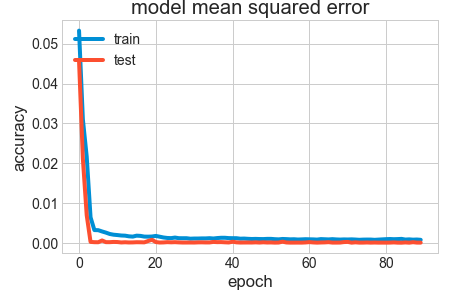


Figure-5

# Results

We deformalize the predicted closing price that we receive during the backtest with our LSTM model,

we can see the LSTM model’s prediction for the adjusted closing price (red line) compared to the actual adjusted closing price (blue line) over time

Our tests for accuracy using different statistical measures such as the MSE loss and RMSE

scores.

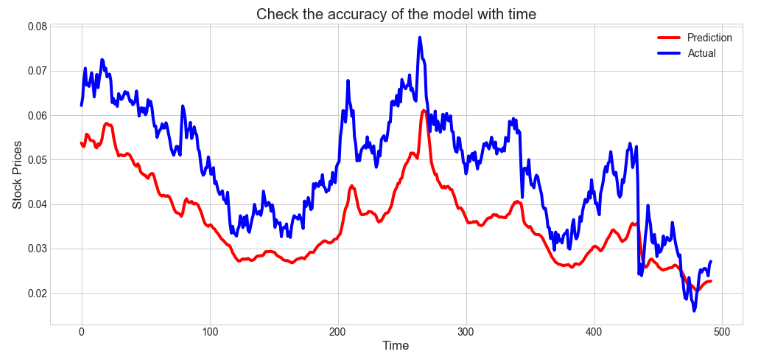


Figure-6

Using LSTM model would not help to minimize the error during volatile periods of the index because

the volatility happens when e.g. during financial crisis or if some negative news concerning

the index comes out to the public. Then the LSTM will not be able to predict with high accuracy. It is quite difficult to predict the adjusted closing price solely based on the open, high, low and closing price as input parameters when building a deep learning model. Financial markets are very complex and there are many factors involved in a stock/index price movement e.g. macro events, market noise, investor sentiment etc. also play an important part of the stock/index price.

# CONCLUSION AND FUTURE WORK

Stock markets are very hard to monitor and require plenty of context when trying to interpret the movement and predict prices.

At its core, the stock market is a reflection of human emotions. Pure number crunching and analysis have their limitations; a possible extension of this stock prediction system would be to augment it with a news feed analysis from social media platforms such as Twitter, where emotions are gauged from the articles. This sentiment analysis can be linked with the LSTM to better train weights and further improve accuracy.

**ANNEXTURE**

1. Code

FINANCIAL\_ANALYTICS\_PROJET-TA17002.ipynb

1. Data File (stock\_price\_2.CSV)