Stock Trading Advisor using Machine Learning

Sazid Rahman Simanto
ECE
North South University
Dhaka, Bangladesh
rsazid99@gmail.com

Shafkat Waheed
ECE
North South University
Dhaka, Bangladesh
shafkat.waheed@northsouth.edu

Mohammad Adib Khan
ECE
North South University
Dhaka, Bangladesh
adib.khan254@gmail.com

***Abstract*— Predicting about a stock trend whether to buy it or not is a very challenging task. An exact prediction may yield profits for investors. Analyzing the stock market data, it's very difficult to build a model which gives a proper prediction. In this study we will develop three models using Random Forest (RN), Support Vector Machine (SVM), Multi Layered Perceptron(MLP) and analyze their performance. The final result we are trying to achieve is to predict 1% profit price of a stock in the next 30 days from when it is bought.**

Keywords—Dhaka Stock Exchange, Support Vector Machine, Random Forest, Deep Neural Network, Profit Maximization, Optimal Stock Selling Price, Stock Trading, Multi Layered Perceptron

I. INTRODUCTION

The stock market is an emerging sector in any country of the world. Many people are directly related to this sector. It is important for those people to gain insight about the market trend. Along with development of the stock market, forecasting stock price has become an important topic among the people. Trend forecasting becomes an essential topic for stockholders, investors and the authority that are related to the stock market business. Predicting stock price is regarded as a challenging task, because stock markets are essentially a non-linear, nonparametric, noisy and deterministically chaotic system . Trend of a market depends on many things like liquid money, stocks, human behavior, news related to stock market etc. All this together controls the behavior of trends in a stock market. The behavior of trend can be analyzed by using technical tools, parametric pricing methods or combination of these methods . Many machine learning techniques have been used in recent time to predict the stock price. Neural network and support vector machine are the most usable among those. At present deep learning is making quite the impact in machine learning and its use in predicting stock price is not properly utilized. The stock market has many sectors starting from cement, bank, pharmaceuticals and much more, among all these sectors bank sectors has the least amount of noise due to its functionality and common factors of influence. The stock market of bank's hold a similarity in its change of trend as the external factors influence all the banks more or less equally. This special property of bank show a pattern that reduces the non-linear, nonparametric, noisy and deterministic chaotic situation of stock market. We used the special nature of this bank stock to predict price more accurately and deterministically.

II. RELATED WORK

Stock Market prediction is a challenging task but various work has been done in this field due to its huge prevailing data. The use of machine learning in this field is daunting as the prediction tends to fade due to various factors influencing the market. Algorithms such as NN and SVM are manly used as they tend to show promising results[1].NN is very efficient in detecting stock price and various NN structure are used by researchers to detect stock price[2][15].These algorithms work better if the data is preprocessed according to the problem to reduce noise to give better results[3].Recurrent Convolutional Neural Network was also used to detect stock price by Bo Xu[15]. Deep learning is a new popular concept in the realm of machine learning, Huy D. Huynh used deep neural network to predict stock price to produce better results [16].The amount of work done in this field is very diverse and we plan to do use three algorithms to predict stock price using the Dhaka stock exchange data. We targeted the bank sectors as the factors influencing all the bank are similar and the chance of having noise in data was very less.

III. DESCRIPTION OF DATA & PREPROCESSING

Stock market data of Dhaka Stock Exchange (DSE) was collected through LankaBangla online portal [3]. Data collection period extends from 1999 to 2018. Each records consist of the 7 columns - stock's ticker name, date of trade, opening price, daily high (price), daily low (price), closing price and daily trade volume. A sample data is provided in figure 1.

Stock	Date	Open	High	Low	Close	Volume
DSEGEN	03/01/2007	1609	1609`	1583	1583	3245459
DSEGEN	04/01/2007	1583	1589	1583	1589	2993592

Table 1: Dhaka Stock Exchange Data

We first segregated the market trend. People express their appreciation about a stock in the market through price. Price of a stock is comparable to the voting result of an election. We derived the market trend which is comparable to election poll.

A set of ten technical analysis (TA) indicators were used for determining the market trend of a stock. We decided not to take take all indicators as it doesn't give the perfect outcomes in all scenarios, some of them are good for long term predictions whereas some works better at short term predictions. A few of those indicators works better in an emerging market where as a few fails in that sector. These are the reason for choosing a set of ten indicators so that a better accuracy could be obtained. The indicators used along with the way they are interpreted has been described below.

1) Simple Moving Average (SMA)

SMA is an arithmetic moving average calculated by adding recent closing prices and then dividing that by the number of time periods in the calculation average. SMA is useful for working with long term market trends[5].

2) Exponential Moving Average (EMA)

EMA is a type of moving average that places a greater weight and significance on the most recent data points. EMA react more significantly to recent price changes than a SMA, which applies an equal weight to all observations in the period. EMA is useful for working with short to midterm market trends[6].

3) EMA Cross

EMA Cross is a popular trading strategy of entering the market when faster EMA (short period) crosses slower EMA (long period), that's when traders enter at the direction of the crossing. This strategy can work very well on trending market, the results are not very good on choppy markets[7].

4) Bollinger Band (BB) - Upper & Lower

BB is plotted two standard deviations away from a simple moving average. Standard deviation is a measure of volatility, when the markets become more volatile, the bands widen; during less volatile periods, the bands contract. It is believed that the closer the prices move to the upper band, the more overbought the market, and the closer the prices move to the lower band, the more oversold the market. Approximately 90% of price action occurs between the two bands[8].

5) Relative Strength Index (RSI)

RSI is a momentum indicator that measures the magnitude of recent price changes to analyze overbought or oversold conditions. It is primarily used to attempt to identify overbought or oversold conditions in the trading of an asset. RSI values range from 0 to 100. Traditional interpretation and usage of the RSI is that RSI values of 70 or above indicate that the asset is becoming overbought or overvalued, and therefore may be primed for a trend reversal or corrective pullback in price. On the other side of RSI values, an RSI reading of 30 or below is commonly interpreted as indicating an oversold or undervalued condition that may signal a trend change or corrective price reversal to the upside[9].

6) Double Exponential Moving Average (DEMA)

DEMA is a fast-acting moving average that is more responsive to market changes than a traditional moving average. It was developed in an attempt to create a calculation that eliminated some of the lag associated with traditional moving averages. The DEMA can be used as a stand-alone indicator and can be incorporated into other technical analysis tools whose logic are based on moving averages. DEMA is very useful for working with short term market trends[11].

7) Weighted Moving Averages (WMA)

WMA assign a heavier weighting to more current data points since they are more relevant than data points in the distant past. The sum of the weighting should add up to 1 (or 100%). In the case of the SMA, the weightings are equally distributed, which is why SMA can't react as fast as WMA to very recent market trends[12].

8) Volume Weighted Moving Average (VWMA)

VWMA is similar to the simple moving average; however, the VWMA places more emphasis on the volume recorded for each period. The VWMA has the ability to help discover emerging trends, identify existing ones and signal the end of a move[13].

VWMA can be used to identify the following signals:

i. Discovering emerging trends - If the VWMA switches below the SMA, this implies a bearish move is on the horizon. This could lead to a weakening in the bullish trend or an outright reversal, when the price is able to break through both the VWMA and the SMA a bearish trend is confirmed and vice versa.

ii. Identifying current tends - if VWMA is between the chart and the simple moving average, then it is a signal for a trending market. Note that sometimes the VWMA will test the SMA as a support and resistance, depending on the primary direction of the security. These tests can be considered as an implication of a potential trend reversal.

iii. Detecting the end of a trend - this signal is pretty much the same as discovering emerging trends. The difference is to look for a contrary signal to the primary trend.

9) Money Flow Index (MFI)

MFI is a momentum indicator that measures the inflow and outflow of money into a security over a specific period of time. The MFI uses a stock's price and volume to measure trading pressure. Since, MFI adds trading volume to the relative strength index (RSI), it's sometimes referred to as volume-weighted RSI[14].

Parameters used for the calculation of the indicator's are provided in table 1 shown below.

Indicator	**Parameters**

SMA	Short Period: 50 days Long Period: 200 days
EMA	Short Period: 12 days Long Period: 26 days
BB L **&** **BB H**	Period: 50 days Standard Deviation: 2.1
RSI	Period: 14 days
DEMA	Short Period: 21 days Long Period: 55 days
WMA	Short Period: 8 days Long Period: 66 days
VWMA	Period: 50 days
MFI	Period: 14 days

Table 2: Time Period of Indicator Calculation

For calculating the values of the indicator, the opening, high, low and closing price was averaged and used in the indicator formulae in order to smooth out any sudden price fluctuation in a day. This is reflected in "Our Price Selection" column of the data table.

We used feature Scaling to scale our data for making our gradient descent more faster and accurate.

IV. ANALYSIS & OUTPUT

We have selected ten features for training with machine learning algorithms. The prediction will be to give a buy signal only if within the next 30 days timeframe, at least for 10 days the stock's price should hold a minimum of 10% or more value than the current price. An example would be suppose stock X's price is Tk.100 right now, by analyzing the indicators it was figured out that for at least 30 days within the next 30 days, the price of stock X will remain at Tk.130 or above (meaning 1% profit or more). If such a condition can be predicted only then a buy signal should be produced. We have selected a minimum of 30 days to hold the profit because indicators most of the time works better with longer averages therefore, taking a short period could give raise to false positive buy signal in a crashing market with just some small upward fluctuations. Figure 2 shows a small section of the table of data with indicator values and the actual outputs which will be used for training the machine learning algorithms.

EMA	SMA	Bbl	Bb	rsi	Dem	WM	VM	MFI	B
-0.7	17	7	-3.12	43.61	-3.3	-1.7	1.13	44.6	0
7.59	30.51	125.95	31.11	69.01	9.21	22.9	8.22	49.4	1
-1.88	-7.05	7.42	-6.98	32.53	-1.3	-5.6	0.45	42.1	0

Table 3: Preprocessed Data

V. METHODOLOGY

We used three algorithms to give our prediction and the algorithms we chose are Multi Layered Perceptron (MLP), Support Vector Machine (SVM), Deep Neural Network (RNN).The problem is a classification problem and these three algorithms showed promising results for the works that was done before

1) Multi Layered Perceptron(MLP)- A multilayer perceptron (MLP) is a feedforward artificial neural network that generates a set of outputs from a set of inputs. An MLP is characterized by several layers of input nodes connected as a directed graph between the input and output layers. MLP uses backpropagation for training the network. MLP is a deep learning method.

A) Hidden Layer: We used relu function as the activation function for this algorithm. Relu function is actually called Rectifier function.

B) Output Layer**:** We used sigmoid function in the output layer as activation function.We have used binary cross entropy as our loss function and Adam optimizer to minimize the loss function.

2) Support Vector Machine(SVM)- A support vector machine (SVM) is machine learning algorithm that analyzes data for classification and regression analysis. SVM is a supervised learning method that looks at data and sorts it into one of two categories. An SVM outputs a map of the sorted data with the margins between the two as far apart as possible.

A) Polynomial kernel- polynomial kernel is used to produce a boundary line of complex shape, polynomial kernel is mainly used in image processing

$$k(Xi, Xj) = (Xi, Xj + 1)^d$$

B) Sigmoid Kernel- It is a kernel that we use as proxy for neural network.

$$k(x, y) = tanh(ax^T y + c)$$

C) Gaussian Kernel- It is a general-purpose kernel; used when there is no prior knowledge about the data. Equation is:

$$k(x, y) = exp(-||x - y||^2/2\sigma^2)$$

3) Random Forest(RF)- A random forest is a data construct applied to machine learning that develops large numbers of random decision trees analyzing sets of variables. This type of algorithm helps to enhance the ways that technologies analyze complex data.

We used 95% of the data as training set, 2.5% of the data as validation set and 2.5% of the data as test set for cross validation for MLP and we used 97% of the data as training set and 3% of the data as test set for SVM and RF. We used Accuracy to evaluate the performance of proposed models. Computation of these evaluation measures requires estimating True Positive (TP), False Positive (FP), True Negative (TN) and False 557 Negative (FN).

Accuracy = TP + TN /TP +FP + TN + FN

We used these methodologies to build our model for detecting stock trend from out data.

VI. RESULT

We have applied three of the methodologies on our data set and on the first try we got poor results due to taking lower samples and high profit rate. We decided to use all the data of bank and we achieved greater accuracy by using this huge data set. We also lowered the profit rate to 1% from 5% producing better accuracy in detecting trend

We applied MLP on the data set for our first methodology and the result we obtained are not very promising.

Accuracy(Train) = 0.7795

Accuracy(Validation) = 0.7770

Accuracy(Test) = 0.7884

The accuracy was the least among the other methodology because of the data we have and the type of NN we used.

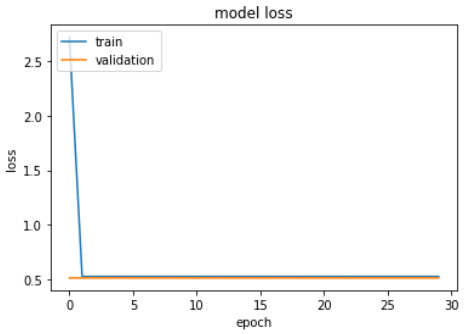

Figure V.1: Graph of Loss for MLP

We see here with each iteration the validation loss and training loss has reduced with each epoch. Recurrent Neural Network is a better NN in case of handling stock data[15].

We applied svm and with three kernels, predicted the trend from the data and we got the following result

SVM	**Polynomial`**	**Gaussian**	**Sigmoid**
Precision	0.66	0.81	0.66
Recall	0.64	0.80	0.65
F-1 Score	0.65	0.74	0.65
Support	7805	7805	7805
Accuracy	0.639	0.804	0.653

Table V.2: Confusion Matrix for Svm

We applied RF to our data with tree size ranging from one to eight.

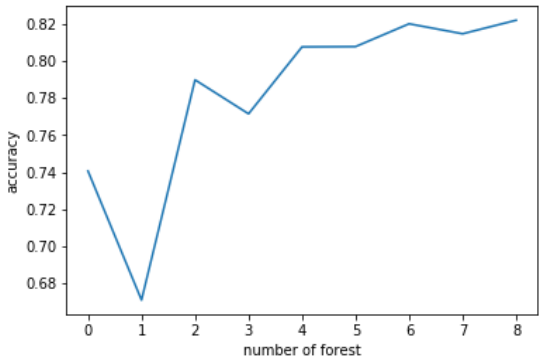

Figure V.3: Tree Size Accuracy

We achieved the best result using Random forest and we used eight types of tree with tree size eight, we achieved the best result and tree size one giving the worst.

	RF
Precession	0.81
Recall	0.83
F1- score	0.80
Support	7805
Accuracy	0.825

Table V.3: Confusion Matrix for RF

We also figured out which features are the most important and which are least from RF. The second feature EMA and VWMA were the most important feature according to RF.

VII. CONCLUSION

We have used all the methodologies in predicting the trend. We believe we can do better with a different neural network instead of MLP. SVM and Random forest showed great results to both being above 80%. We plan to increase the accuracy by adding few features and tweaking the algorithms to more suite our problem.

REFERENCES

[1] Bank.S, Khan.K & Anwer.M. *Hybrid Machine Learning Technique for Forecasting Dhaka Stock Market Timing Decisions*. Retrieved from Computational Intelligence and Neuroscience,feb 19. 2014

[2] Pate.J, Shah.S & Thakkar.P .*Predicting stock and stock price index movement using Trend 4 Deterministic Data Preparation and machine learning techniques*. Retrieved from Computer Science & Engineering Department, Institute of Technology, Nirma University, Ahmedabad, Gujarat, India, 2014

[3] LankaBangla,*LankaBanglaFinancialPortal*.http://lankabd.com/market/indices-sectors

[4] Miah M..B.A , Hossain.Z, Hossain.A, Islam.M,*Price Prediction of Stock Market using Hybrid Model of Artificial Intelligence* ,International Journal of Computer Applications (0975 – 8887) Volume 111 – No 3, February 2015.

[5] Investopedia,*Terms*.https://www.investopedia.com/terms/s/sma.asp

[6] Investopedia,*Terms*.https://www.investopedia.com/terms/e/ema.asp

[7] StrategyQuant,*Tutorial*.https://www.investopedia.com/terms/e/ema.asp

[8] Investopedia,*Terms*.https://www.investopedia.com/terms/b/bollingerbands.asp

[9] Investopedia,*Terms*.https://www.investopedia.com/terms/r/rsi.asp

[10] Investopedia,*Terms*.https://www.investopedia.com/terms/m/macd.asp

[11] Investopedia,*Terms*.https://www.investopedia.com/terms/d/double-exponential-moving-average.asp

[12] Investopedia,*Ask*.https://www.investopedia.com/ask/answers/071414/whats-difference-between-moving-average-and-weighted-moving-average.asp

[13] TradingSim,*DayTrading*.https://tradingsim.com/blog/volume-weighted-moving-average/

[14] Investopedia,*Terms*.https://www.investopedia.com/terms/m/mfi.asp

[15] Xu.B, Zhang.D, Zhang.S, Li.H, and Lin.H, *Stock Market Trend Prediction Using Recurrent Convolutional Neural Networks.Dalian University of Technology, Dalian, China*

[16] Huy D. Huynh.H, Dang.L.M and Duong.D *A New Model for Stock Price Movements Prediction Using Deep Neural Network*