

Prediction of Gold Stock Market using Hybrid Approach

Kashyap Kitchlu, Shubham Kumar Singh
Department of Computer Science and Engineering,
BNM Institute of Technology,
Bangalore, India.

Abstract—Gold has been associated with one of the ancient trades and over the past years, Machine Learning has become an important subject in the technological field. The fusion of this modern technology with an essential commodity such as gold, can lead us to important insights in the stock market. Moreover the sentimental affinity towards gold has been sturdily guiding the fluctuations in the stock market, making it ever so more important to present an approach which consists of prediction based on historical data and sentimental analysis.

Keywords: ANN, LSTM, SVM, NSE, OHLC.

I. INTRODUCTION

Stock market has been studied in several ways in which multiple factors are determined like gold, oil, sentiments, social movement that have an effect on general stocks. Artificial neural networks (ANNs) are a unit computing systems mistily impressed by the biological neural networks that represent animal brains. Such systems learn i.e. increasingly improve performance on tasks by considering examples, usually while not task-specific programming. ANN is predicated on a set of connected units or nodes referred to as artificial neurons that may be a simplified version of biological neurons in associate animal brain. Every association between artificial neurons will transmit an indication from one to a different. The unreal somatic cell that receives the signal will method it, then signal artificial neurons connected to it. Artificial neurons are a unit organized in layers. Totally different layers could perform different forms of transformations on their inputs. Signals travel from the primary (input), to the last (output) layer, presumably traversing the layers multiple times. Presently, everywhere around the globe, huge quantity of investment units are dealt by the Stock Markets. Nationwide financial set-up are a unit sturdily connected and closely inclined to the accomplishment of their Stock Markets. It is the foremost fashionable investment theme promising, albeit some risks and is a very important part of the economy of the country. In addition, securities market platform is just too approachable by each planned investor through conjoint analysis. But in this digital world, the aim of any capitalist within the stock market is to earn higher returns. Therefore, the matter becomes: for a given market history, verify the instant of buying/entry or selling/exit price that generates profitable revenue. By taking all the above issues into cognizance, we realize additional returns by market prediction. We have a solution which proposes a way that incorporates 3 subtasks: (1) predicting stock worth of gold by accumulating sentiment of stock capitalist through Google feedback survey and (2) prediction of stock worth of gold by

set of metrics exploiting the historical information of gold; (3) merging the results from each ways and prediction of ultimate stock worth of gold and result of alternative commodities on gold value.

II. LITERATURE REVIEW

Neelima Budhani et al proposed a feed forward neural network with back propagation training algorithm for prediction. Prediction methods include Technical analysis and Fundamental analysis. Feed forward network feeds input from one layer to other layer (e.g hidden layer) in one direction and at each intermediate the input value coming from previous layer is multiplied by weights and summed up before sending to output layer then output is taken from output node and with this Back propagation algorithm is used to reduce the error by feeding the output of model as input after calculating the percentage of error, so in this way NN model learn itself for new data samples.

Jiahong Li et al proposed a new method for stock market prediction which adopts the Long Short Term Memory (LSTM) neural network and incorporates investor sentiment and market factor to improve forecasting performances. Naive Bayes method is used on investor sentiment data taken from news article and twitter. This model included three phases. In the first phase Naive Bayes method to classify data set into one of the three categories: positive, negative or neutral. In the second phase, an investor sentiment index is constructed to measure the daily mood of stock market. In the third phase, a Long Short-Term Memory model is used to test the hypothesis that the prediction accuracy of stock market prediction models can be improved by including measurements of investor sentiment.

Vivek Rajput and Sarika bobde had proposed hybrid model which includes sentiment analysis and clustering algorithm. This methodology had given two outputs set i.e. one from sentiment analysis and another from clustering based prediction with respect to some specialized parameters of stock exchange. Output from both the model is accumulated processed, analyzed based on value of technical indicators for each stock and final prediction is provided. Model A uses data from different sources such as twitter and news articles then undergoes through various phases such as: parsing, filtering, tokenization etc. Model B uses clustering algorithm on historical data and on the basis of that analysis it forms three data sets. Accuracy was better than past proposed model (<65%).

III. PROPOSED SYSTEM

There were various models that were implemented before, like HMM, SVM, Artificial Neural Networks etc. over different models, none of the models considered essential commodities as a major front and sentiment affiliations from experts in the field directly. The key difference between other models and this model is that the method of implementation is incremental learning. The connection of past data and the sentiments together is in an incremental form. The most important part of the starting phase is the system design.

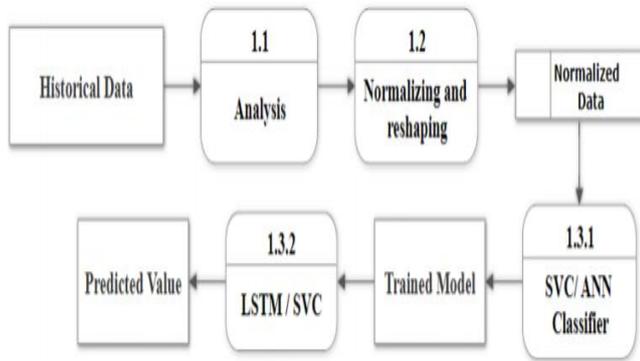


Fig 3.1: Data flow chart of model

IV. METHODOLOGY

This section tells us how we wish to proceed after collecting data. The various stages that are going to be implemented are given below:-

- 1) Input
- 2) Segregation
- 3) Analysis
- 4) Classification
- 5) Prediction
- 6) Result

A. Data Collection

The first step is to collect data. Dataset is a collection of existing information coded in suitable form for usage and processing. Our dataset is the historical data of gold stock value in csv form which is taken from National Stock Exchange (NSE). The csv file contains daily Open High Low Close (OHLC) data for the stock of RELIANCE GOLD on NSE for the time period from 1st January 2007 to 4th May 2018.

B. Segregation

• Segregation of Historical Data

In the second stage, we segregated data by creating various excel sheets and categorize the data on the basis of technical indicators. As we are predicting the value of Gold stock on daily basis, we segregated the data based on days.

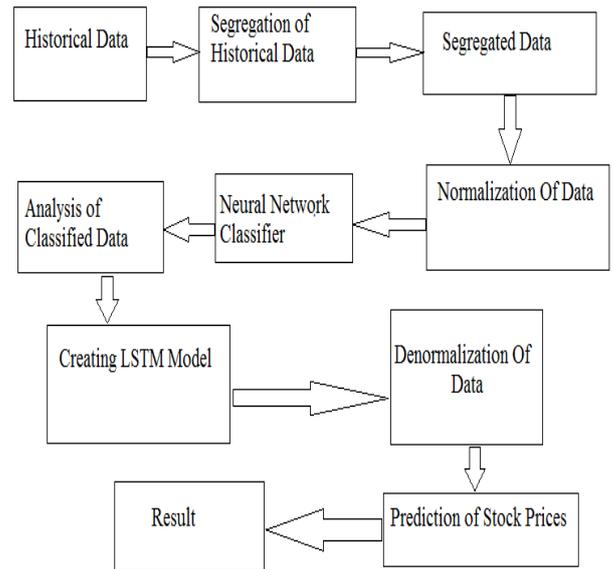


Fig 4.1: Block diagram for Neural Network Model

• Segregation of Feedback

Feedback survey results are segregated on the basis of whether user is a stock market investor or not and is stored in csv format.

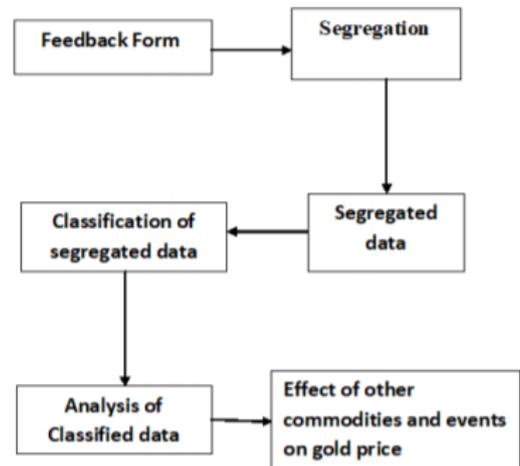


Fig 4.2: Block diagram for Feedback data

C. ANN Model

• Importing the Datasets Module

Dataset is imported which is stored in the csv format file. This is done using the pandas library, and the data is stored in a data frame named df. We chose only the OHLC data from this dataset, which also contained the date, Adjusted Close and Volume data.

```
def get_stock_data(normalize=True):
    prices=pd.read_csv('data1.csv', header=0)
    df = pd.DataFrame(prices)\n",
    df.drop(df.columns[[0,1,3,6]], axis=1, inplace=True)
    df.to_csv('data.csv', mode='w',float_format='%0.3f')
```

• *Building Neural Network*

To build the Neural Network, we have imported essential functions like sequential() method from the keras models library that is used to build the Network sequentially and dense() method from keras.layers library which is used to build the layers in the Neural Network. We instantiate the Sequential() function into the variable model. This variable will then be used to build the layers in the network. To add layers into our Model, we make use of the add() function.

```
def build_model2(layers, neurons, d):
    model = Sequential()
    model.add(LSTM(neurons[0],input_shape=(layers[1],layers[0]),return_sequences=True))
    model.add(Dropout(d))
    model.add(LSTM(neurons[1],input_shape=(layers[1],layers[0]),return_sequences=False))
    model.add(Dropout(d))
    model.add(Dense(neurons[2],kernel_initializer="uniform",activation=relu))
    model.add(Dense(neurons[3],kernel_initializer="uniform",activation=Sigmoid))
    model.compile(loss='mse',optimizer='adam', metrics=['accuracy'])
    return model
```

• *Sign up form*

This form is constructed for users to fill in their details and create their account. Whenever they want to access this page again, they can login and can find their previous information saved. They can create secure password with which they get authority to access particular account. This makes every account safe and prevent unauthorized access

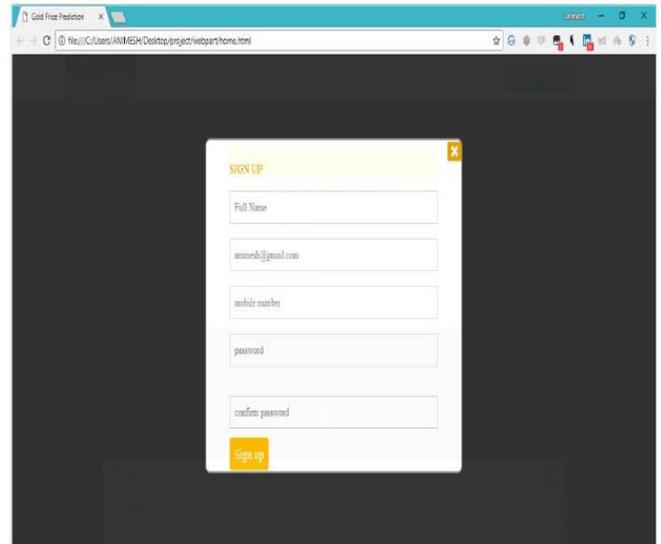


Fig 4.3: Sign Up form

• *Algorithm for feedback form*

The user is allowed to enter his/her name and whether he/she has ever invested in stock market in a “Yes” or “No” format. The answers to the queries are then retrieved and analyzed. This is done by downloading the user feedback in csv format. Answers for a question are passed as a query to the piece of code written in ‘python’ to check whether the feedback given by the user is appropriate or not. This is summarized to check the effect of other commodities and events on price of gold.

```
Algorithm for feedback form
survey=pd.read_csv('survey.csv', header=0)
df = {0:survey['Have you ever invested in stock market?'],
      1:survey['Effect of increase in crude oil price on gold price?'],
      2:survey['Effect of coming elections on gold price?'],
      3:survey['Effect of festival like diwali,akshaya tritiya?'],
      4:survey['Effect of increase in Nifty and Sensex on gold price?'],
      5:survey['Effect of decrement in US dollar value on gold price?']}
d = pd.DataFrame(df)
c=0, oil=0
ele=0
fest=0, ns=0, usd=0, l=len(d)
for i in range(l):
    if d[0][i]=="Yes":
        c=c+1, oil=oil+d[1][i], ele=ele+d[2][i], fest=fest+d[3][i], ns=ns+d[4][i]
        c=c*5, oil=(oil*100)/c, ele=(ele*100)/c, fest=(fest*100)/c, ns=(ns*100)/c,
        usd=(usd*100)/c
```

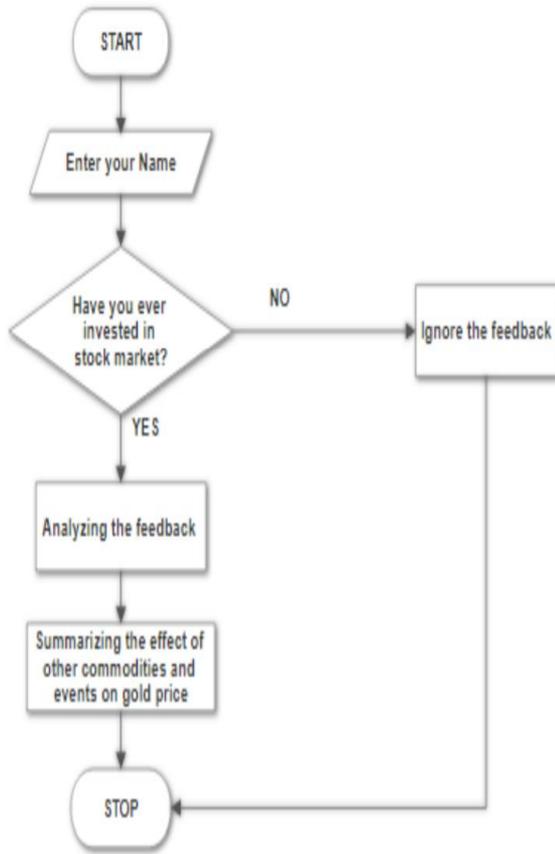


Fig 4.4: Flow chart for feedback by user

D. PREDICTION

- LSTM Model Score

Train and Test score is calculated by comparing the actual values using Root Mean Square Error

```

Train Score: 0.00013 MSE (0.01 RMSE)
Test Score: 0.00004 MSE (0.01 RMSE)

(0.00013450330063929732, 4.281678143442595e-05)
    
```

Fig 4.5: LSTM Model Score

- Graph of Actual vs Predicted Value

The graph has been plotted for the predicted and actual value of RELIANCE GOLDBEES stock value during testing phase.

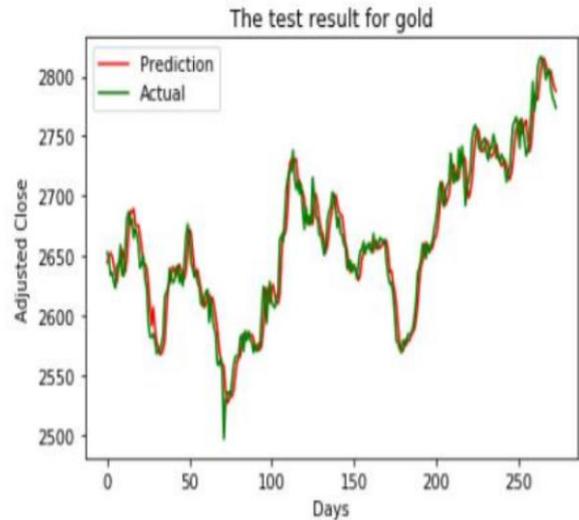


Fig 4.6: Snapshot of Plot of Predicted and Actual value

In this, the normalized data is divided into two parts, 90% for training purpose and rest 10% for testing purpose of the model.

- Expert Advice

The expert gives advice in the form of percentage relative to how other events are going to affect the increment in gold price. Other events can be election, festival or a change in price of other commodities. Also, dollar price has major role in change in price of gold.

```

expert advise on gold price increment as percentage:
Due to increase in crude oil price: 72.41379310344827 %
Due to election: 66.20689655172414 %
Due to festival: 80.0 %
Due to increase in Nifty and Sensex: 46.206896551724135 %
Due to dollar price: 55.86206896551724 %
    
```

Fig 4.7: Snapshot of Expert Advice

V. RESULT

The proposed algorithm i.e. LSTM has been implemented using 'python'. The performance of proposed approach has been studied using different kinds of measures like confusion matrix.. The proposed model gives prediction for gold stock value for each day and for the next day. The master work of this application is to guide the user who is investing in stock market so as to get maximum profit.

Comparison between Actual and Predicted value

Actual Value	Predicted Value
2652.784	2638.667
2646.985	2646.572
2633.925	2648.517
2635.777	2645.126
2627.59	2638.071
2623.205	2629.857
2642.112	2625.107
2640.163	2646.98
2658.875	2644.484
2634.364	2653.959
2633.34	2643.007

Table 1: Snapshot of Actual and predicted values for 10 days

In below table actual and predicted values are compared and error rate for all predicted values is shown using RMSE and MAE.

Next day prediction	Actual value	MAE	RMSE
2784.166	2774.95	11.6588	14.5474

Table 2: Result with Performance Metrics

CONCLUSION

For our project we have used already existing systems as benchmark to evaluate the prediction correctness and Error. Support Vector Machine and Hidden Markov Model are used as benchmark for comparative study with our LSTM Model. Comparative Study shows us LSTM MODEL is a better model to predict the Gold stock value of RELIANCE GOLDBEES.

REFERENCES

- [1] C. Xiang and W. M. Fu , “Predicting the Stock Market using Multiple Models”, 1-4244-0342-1/06.
- [2] Neelima Budhani, Dr. C.K.Jha and Sandeep K. Budhani, “Prediction of stock Market Using Artificial Neural Network”, 978-1-4673-9120-7/14,1-8.
- [3] Vivek Rajput and Sarika Bobde , ”Stock Market Prediction Using Hybrid Approach”, 978-1-5090-1666-2/16,82-86.
- [4] Kashyap Kitchlu, Shubham Kumar singh, “Prediction of Gold Stock Market Using HMM Approach”
- [5] Jiahong Li, Hui Bu*, Junjie Wu , “Sentiment-Aware stock market prediction: A deep Learning,978-1-5090-6371-0/17
- [6] Maryam Farshchian and Majid Vafaei Jahan, “Stock Market prediction Using Hidden Markov Model”, 978-1-4673-9762-9/15,473-477.