

Prediction of Gold Stock Market Using HMM Approach

Kashyap Kitchlu¹, Shubham Kumar Singh², Dr. Vimuktha Evangeleen Salis³

^{1,2}Computer Science & Engineering, B.N.M Institute of Technology, Bengaluru, Karnataka, India

³Associate Professor, Department of Computer Science & Engineering, B.N.M Institute of Technology, Bengaluru, Karnataka, India

Abstract – Machine learning has been extensively studied for its potential in prediction of financial markets. Nowadays stock market is one of the major sources of raising resources for India and is acting as a key driver for economic growth of the country. The stock market can be eccentric and may be deemed capricious seeing that various different commodities and factors affect the economic index. Hence, the need arises for an intelligent tool for stock prediction. The objective of this paper is to construct a model to predict stock value movement of a particular commodity i.e. gold to predict National Stock Exchange (NSE).

Key Words: NSE, Python, HMM, Pandas, RSME, MAE

1. INTRODUCTION

Over the past some years, machine learning has become an important subject in the trending technological field. It is a subfield of computer science and artificial intelligence that focuses on the design of systems that can learn and make decisions and predictions based on data, instead of explicitly programmed instructions. Overall it enables computers to act and make data-driven decisions rather than being explicitly programmed to carry out a certain task. One of the important uses of machine learning has been the growth of its potential in predicting the stock market. There are various factors that affect the stock market such as the core of a country's economic strength, the rise of unlikely political scenarios, the influence of investor sentiment through newspapers and media sources etc. We isolate the other factors, considering mainly the stock data present for the commodity. Oil market is predominantly made responsible for market volatility but other commodities e.g. corn, gold and cotton have, more often than not, their influence on daily stock prices. It is seen that gold indirectly follows the movements of the market while oil does it directly. It is historical fact that gold has been viewed as a 'counter cyclic asset', which means it gains value during market downturns. Investors flock to the precious metal as and when outlook of the equity market

seems bleak, because it is present throughout the world and is perceived as a universal currency holding high intrinsic value. On the other hand, gold in India has been traditionally and culturally important to people because of their emotions deeply connected with this commodity. In literature, various publications have had a concerted effort on predicting stock market as a general theme but none of them considered gold as an essential commodity. Hence a need arises to implement a model to predict gold value.

In technology, Hidden Markova Model is the one in which we observe a sequence of emissions, but do not know the sequence of states the model went through to generate the emissions. Analysis of hidden Markov models seek to recover the sequence of states from the observed data. Hidden Markov model was proved appropriate for modelling the dynamic system and can be beneficial for finding unknown parameters of hmm. Thus we implement it for our study.

2. LITERATURE REVIEW

Neelima Budhani et al[2] 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.

Maryam Farshchian and Majid Vafaei Jahan's[3] proposed Hidden Markov Model was proved appropriate for modelling the dynamic system and it has been shown how Baum-Welch and Progressive algorithm can be beneficial for finding unknown parameters of HMM. In

implementation phase Baum-Welch was used to train model to learn progressively and incrementally and calculates error possibility after each transition. In their research they have compared their predicted values with actual data for different industries and got maximum 81% accuracy Poonam Somani et al [4] have shown how Hidden Markov Model is used to locate patterns from past data sets that matched the current days stock price behavior. Baum Welch algorithm was used to train the data set for developing model. Input features were Open, Close, High, Low and stock data was stored in 3-d vector form in training phase to calculate the fractional change in testing phase for analysis purpose. Once the model was trained and testing was done, result analysis and accuracy was done by using Mean Absolute Percentage Error (MAPE).

Jiahong Li et al [6] 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 Naïve 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.

3. PROPOSED SYSTEM

The Hidden Markov Model has been widely used in financial mathematics area to predict stock prices (Hassan and Nath 2005; Nobakht et al. 2012; Nguyen 2014). Hence, we try to implement HMM model on National Stock Exchange to predict market values based on open, low, high and adjacent close.

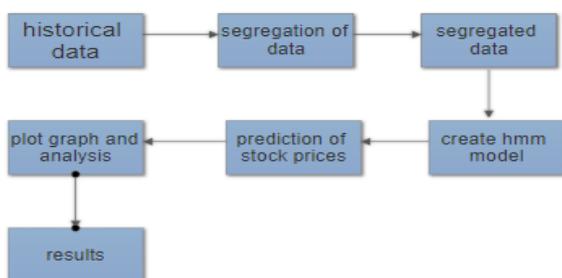


Fig-1: Proposed model of hmm

Hidden Markov Model (HMM) is a statistical Markov model in which the system being modelled is assumed to be a Markov process with unobserved (i.e. *hidden*) states. In simpler Markov models (like a Markov chain), the state is directly visible to the observer, and therefore the state transition probabilities are the only parameters, while in the hidden Markov model, the state is not directly visible, but the output (in the form of data or tokens), dependent on the state, is visible. Each state has a probability distribution over the possible output tokens. Therefore, the sequence of tokens generated by an HMM gives some information about the sequence of states. A hidden Markov model can be considered a generalization of a mixture model where the hidden variables (or latent variables), which control the mixture component to be selected for each observation, are related through a Markov process rather than independent of each other.

4. METHODOLOGY

A HMM is a non-deterministic stochastic Finite State Automata model. The basic structure of a HMM consists of a connected set of states, $S = (S1, S2, \dots, Sn)$. We use first order HMMs, where the state of a system at a particular time t is only dependent on the state of the system at the immediate previous time point, i.e., $P(St|St-1, St-2, \dots, S1) = P(St|St-1)$.

The various stages that were implemented are given below: -

- 1) Data collection,
- 2) segregation,
- 3) data preparation,
- 4) hmm model,
- 5) prediction
- 6) analysis
- 7) results

4.1 Data Collection

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

	open	low	high	Adj Close	
0	1000		940	1105	1000
1	957.8		940.05	957.8	947.8
2	948		944.1	948.5	942.39
3	950		946	951.9	945.67
4	950		945	951	950.15
5	945		940	945	947.18

Fig-2: Data Collection

4.2 Segregation

This is the second stage in which we segregate data by creating various excel sheets and making the data categorized on the basis of technical indicators. As we are predicting the value of Gold stock value on daily basis, we are segregating the data based on days.

4.3 Data preparation

Dataset is imported which is stored in the .csv format. This is done using the pandas library, and the data is stored in a data frame named df. We then drop the missing values in the dataset using the dropna () function. We choose only the OHLC data from this dataset, which would also contain the date, Adjusted Close and Volume data. We will be building our input features by using only the OHLC values. In this part of the code, we will split our input and output variables to create the test and train datasets. This is done by creating a variable called split, which is defined to be the integer value of 0.8 times the length of the dataset

4.4 HMM Model

We build a HMM instance by passing the parameters described above to the constructor.

```
import numpy as np
from hmmlearn import hmm
np.random.seed(42)

model = hmm.GaussianHMM(n_components=3, covariance_type="full")
model.startprob_ = np.array([0.6, 0.3, 0.1])
model.transmat_ = np.array([[0.7, 0.2, 0.1],
                             [0.3, 0.5, 0.2],
                             [0.3, 0.3, 0.4]])
model.means_ = np.array([[0.0, 0.0], [3.0, -3.0], [5.0, 10.0]])
model.covars_ = np.tile(np.identity(2), (3, 1, 1))
X, Z = model.sample(100)
```

4.5 Prediction

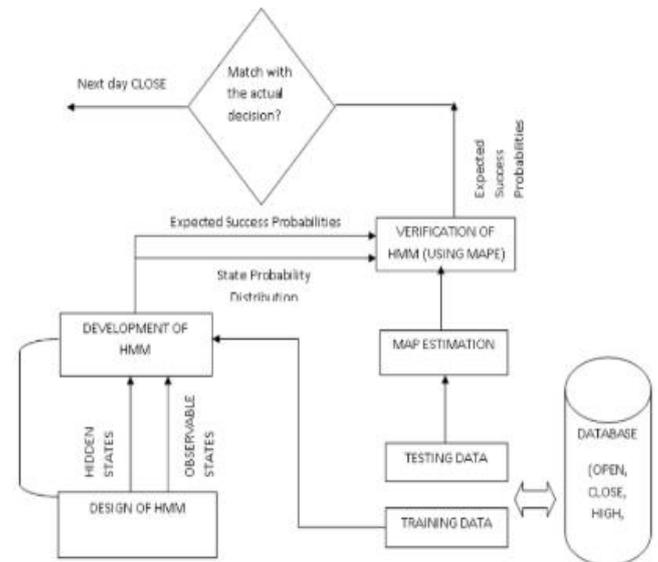


Fig-3: Flow chart of hmm

The algorithm used here in predicting is Viterbi algorithm. The Viterbi algorithm is a dynamic programming algorithm for finding the most likely sequence of hidden states—called the Viterbi path—that results in a sequence of observed events, especially in the context of Markov information sources and hidden Markov models.

The independence assumptions of the HMM tell us how to break up the big search problem (Find $t^* = \text{argmax}_t P(w | t)P(t)$) into smaller sub problems.

The data structure used to store the solution of these sub problems is called a trellis.

4.6 ANALYSIS

Evaluation metrics explain the performance of a model. An important aspects of evaluation metrics is their capability to discriminate among model results. We quantify the performance on the basis of time and accuracy. We can objectively compare the performance of the implementation in various situations. Time indicates the efficiency of the system with regard to memory and processing usage whereas, accuracy implies effective sentiment analysis. For our model we have used Mean Square Error (MAE) and Root Mean Square Error to calculate accuracy of system. Mean absolute error is

average difference between actual and predicted value. It is computed using following formula:

$$MAE = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j|$$

Root mean square error value is standard deviation between actual and predicted value and is computed using following formula:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Predicted_i - Actual_i)^2}{N}}$$

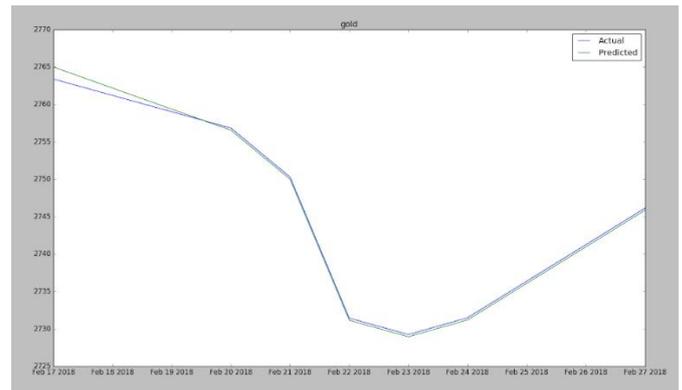


Fig-6: Plot of actual vs predicted values

4.7 RESULTS

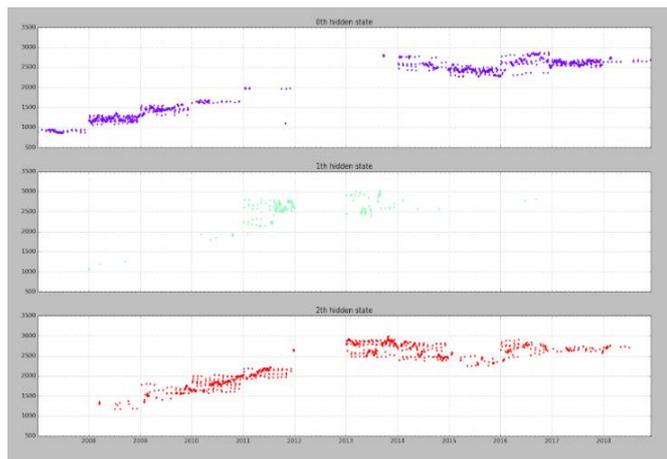


Fig-4: The plot for showing the hidden states

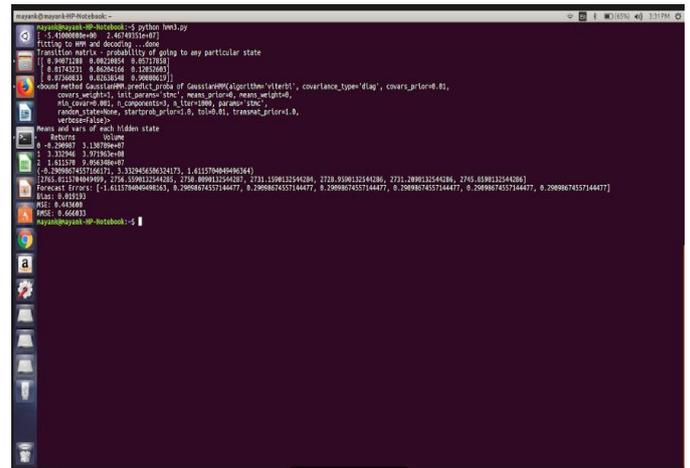


Fig-7: Result as shown in UNIX

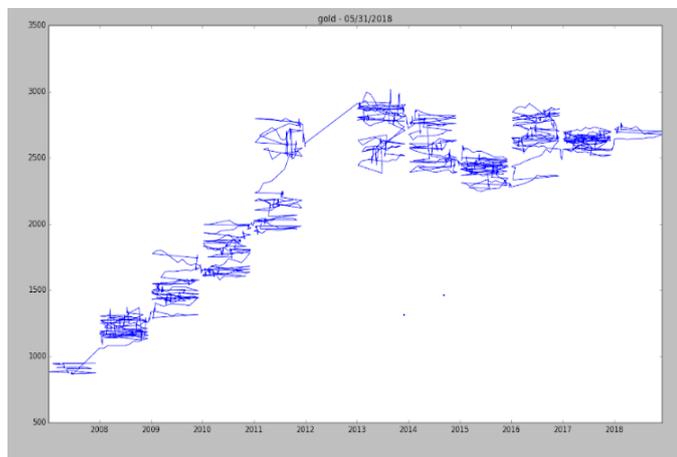


Fig-5: Plot based on the historical data

Results give us transition matrix i.e. the probability of going to any state. It gives us the various parameters. It gives the actual price and the predicted price for the last 7 days

It gives us the calculated bias, MSE and RMSE.

Actual values	Predicted values
2763.4	2765.0115704049499
2756.85	2756.5590132544285
2750.3	2750.0090132544287
2731.45	2731.1590132544284
2729.25	2728.9590132544286
2731.5	2731.2090132544286
2746.15	2745.8590132544286

Fig-8: Actual values vs Predicted values

5. CONCLUSION

Stock's performances are an essential indicator of the strength or weakness of the stock's corporation and economic viability in general.

The proposed algorithm i.e. Virtebri Algorithm for HMM has been implemented using python on UNIX. 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 next day. The master work of this application is to guide the user who is investing in stock market so they get maximum profit.

6. FUTURE WORK

It can be further improved by clubbing two methodologies together as a clustering algorithm to get better results such as HMM with ANN (artificial neural network) as well as we can consider the sentiments of gold customers and fuse it to get better results. Hence this can be improved and further utilised in the area of gold market prediction.

7. REFERENCES

- [1] Aditya Gupta and Bhuwan Dhingra, Stock Market Prediction Using Hidden Markov Models,
- [2] Neelima Budhani, Dr. C.K.Jha and Sandeep K. Budhani, "Prediction of stock Market Using Artificial Neural Network", 978-1-4673-9120-7/141-8..
- [3] Maryam Farshchian and Majid Vafaei Jahan, "Stock Market prediction Using Hidden Markov Model", 978-1-4673-9762-9/15,473-477.
- [4] Poonam Somani, Shreyas Talele, Suraj Sawant, "Stock Market Prediction Using Hidden Markov Model", 978-1-4799-4419-4 /14,89-92.
- [5] Vivek Rajput and Sarika Bobde, "Stock Market Prediction Using Hybrid Approach", 978-1-5090-1666-2/16,82-86.
- [6] Jiahong Li, Hui Bu*, Junjie Wu, "Sentiment-Aware stock market prediction: A deepLearning,978-1-5090-6371-0/17