Bangladesh Share Market Forecasting using Hidden Markov Model

1st Md. Ashekur Rahman
Dept. MSCCSE
Brac University (BracU)
Mohakhali, Dhaka, Bangladesh
md.ashekur.rahman@g.bracu.ac.bd

2nd Asma Ul Hussna
Dept. MSCCSE
Brac University (BracU)
Mohakhali, Dhaka, Bangladesh
asma.ul.hussna@g.bracu.ac.bd

3nd Sanjida Ali Shusmita
Dept. MSCCSE
Brac University (BracU)
Mohakhali, Dhaka, Bangladesh
sanjida.ali.shusmita@g.bracu.ac.bd

4ndSamiha Khan
Dept. MSCCSE
Brac University (BracU)
Mohakhali, Dhaka, Bangladesh
samiha.khan@g.bracu.ac.bd

5nd Iffat Immami Trisha
Dept. MSCCSE
Brac University (BracU)
Mohakhali, Dhaka, Bangladesh
iffat.immami.trisha@g.bracu.ac.bd

6nd Md Sabbir Hossain
Dept. MSCCSE
Brac University (BracU)
Mohakhali, Dhaka, Bangladesh
md.sabbir.hossain1@g.bracu.ac.bd

7nd Annajiat Alim Rasel
Dept. MSCCSE
Brac University (BracU)
Mohakhali, Dhaka, Bangladesh
annajiat@gmail.com

***Abstract*—Stock markets are one of the most complicated systems, and modeling them in terms of dynamical equations is nearly difficult. The fundamental reason for this is because stock prices are influenced by a number of unknown factors such as economic conditions, corporate policy changes, supply and demand among investors, and so on. These variables are continually changing, making stock markets very volatile. Stock price prediction is a typical non-stationary pattern recognition issue in Machine Learning. There has been a lot of study into utilizing Artificial Intelligence and Machine Learning techniques like the Hidden Markov Model to forecast the behavior of stocks based on their previous performance. We used Hmm in our stock price prediction. In our model used the daily opening, closing, high and low indices as continuous observations from underlying hidden. It may be able to predict the price of a stock in the future. It will benefit people, and they will be interested in putting their money in the stock market as a result.**

***Index Terms*—Hidden Markov Model (HMM), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and Mean Absolute Percentage Error (MAPE)**

I. Introduction

Stock or share market is a system where entrepreurs need fund for the business are get those money from small institutions at low costs. People has been benefited by investing their money in stock market. Bangladesh stock market was considered a great year for investors. There was rarely any investor who losses his money. In the beginning of 2011, the Bangladesh stock market fall down. So we can say it is one of the most complex systems and forecasting the price quite unpredictable. There are a number of unknown variables, such as economic conditions, company's policy change, supply and demand between investors, etc. which creates influence on the stock prices. These variables are constantly changing, making financial markets extremely volatile or unpredictable.Stock price prediction is a problem in Machine Learning that involves non-stationary pattern recognition. We are developing a system so that it could forecast future price of a stock and people will be benefited and will show interest in investing their money in stock market. For analyzing the stock markets we are using Hidden Markov Models(HMM) which have a strong probabilistic framework for recognizing patterns in stochastic processes. HMM can work better for the stocks with high volatility .

II. Proposed Model

A. Hidden Markov Models (HMM)

Hidden Markov Models have a high probability of success for pattern recognition. They have been used for analyzing patterns in speech, handwriting, gestures and wide variety of DNA sequences. The transition from one state to another is a Markov Process where the next state depends on the present state. HMM is the model of the observations depend on the states of the system which are 'hidden' to the observer hence the name Hidden Markov Models. In HMM states are always discrete, but the observations can either be discrete or continuous or both. Stock markets can be viewed as a Hidden Markov Process where the investor can only observe the stock prices and the underlying states which are driving the stock prices are unknown.

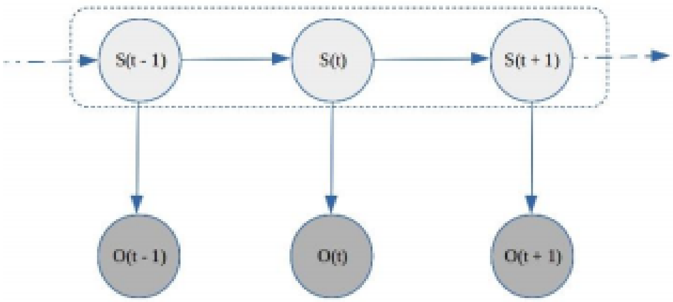

Figure 1. Hidden Markov Process

B. Financial Time Series analysis

Markov chain with a hidden Markov Models have proven to be an effective method for studying non-stationary processes. The stock market is a non-stationary system of constant findings. Consider, Observation Sequence 0_t be a vector of four elements- daily close, open, high and low States and S_t o be one of the assumed states on day t. Observations 0_t are independent and modelled as Multivariate Gaussian distributed takes real values. HMM is a finite state machine so S_t can take only discrete values. We will be using the some notations to define Hidden Markov Models.

Number of observations, T
Latency, K
Number of States, $N(S_t$ = S_1, S_2 $\ldots$. , $S_N)$
Observation Sequence, 0_t
Initial State Probability, P_0
State Transition Matrix, A = [a_ij] where a_ij the state transition probability from i to j
Observation Probabilities,μ_i $\sum_i$ i= 1,2, . . N where μ_i, $\sum_i$ are the mean and covariance matrix for Gaussian distribution for state i The Hidden Markov Model can be represented as λ = (A, μ, $\sum$,P_0)

C. Prediction of Stock Prices

To calculate the log-likelihood of K for predicting the next day's stock price. From previous observations and comparing it with the log-likelihood of all the previous sub-sequences of same size by shifting the window by one day in the direction of past data1. Then we identify a day in the past whose log-likelihood of its K previous observations is the closest to the sub-sequence whose next day's price is to be predicted.

j = $argmin_i$(| P (O_t, $O_t - 1$, $O_t - 2$,.., $O_t - K$) | λ) - P ($O_t - i$, $O_t - i - 1$, $O_t - i - 2$,.., $O_t - K$) | λ |) where i is 1,2,..,T/K

The difference in price between the indicated day and the next day is then calculated. This adjustment is then added to the current day's price to arrive at our projection for the next day.

$O_t + 1$ = O_t + $O_t - j + 1$ - $O_t - j$

To verify that our model does not diverge, we obtain the true observation, add it to our dataset, and fine-tune our model parameters. In other words, we fix the size of our sub-sequence and look for a comparable sub-sequence in previous data. The behavior of the detected sub-sequence is then mapped to the sub-sequence utilized for prediction.

We train a collection of models by varying the number of states (N) from the state space to find a model with the best number of states. We looked at the number of states in G ranging from [2,25]. If we take more sates it will give much accuracy but may cause overfitting. We will calculate the negative log-likelihood of the training data. Negative vakues used for each of the models and chose the model which has the lowest. This tends to prefer a complex model implying that the number of states chosen may tend to a higher value and might result in overfitting. To avoid this issue, a penalty term is added to the negative log probability. We apply varied degrees of limitations on the model depending on the penalty term chosen. We have explored Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) which are known as two different performance measure metrics. To derive the performance measure of the model in AIC, we add the number of model parameters to the negative loglikelihood value, but in BIC, we add the product of the number of model parameters and the logarithm of the number of observation samples utilized.

AIC = - 2 log (P ($O_t rain|$ λ)) - 2p

BIC = - 2 log (P ($O_t rain|$ λ)) - p.log(T)

These are the product of the model for performance measuring. When a false negative result is more misleading than a false positive, AIC is better; when a false positive is as misleading as, or more misleading than, a false negative, BIC is better. In our project, we have used BIC as the model's performance measure to select the number of states in our target model.

III. Implementation

We used in this project is Mean Absolute Percentage Error (MAPE) as performance metric which is define as

$$MAPE = \sum_{i=1}^{m} \frac{|Predicted(i) - True(i)|}{True(i)}$$

Our main objective is to determine the efficiency of HMM in predicting the stock prices.To train the model and get the probability of the data, we used hmmlearn, an open source python package. We have collected dataset from investing.com of six companies which are: Unilever Consumer Care Limited (UNILEVERCL), Grameenphone Ltd (GP), Dutch-Bangla Bank Ltd (DUTCHBANGL), Beximco Pharmaceiticals

Ltd. (BXPHARMA),Berger Paints Bangladesh Ltd. (BERGERPBL), British American Tobacco Bangladesh Company Limited (BATBC) . We used 4 types of features opening price, closing price, high and low for at least the past 1599 working days which more the 4 years of daily stock prices. The most recent 90 observations were set aside for testing, while the remainder of the data were utilized to train the model. For the last 90 days, we projected the prices. Beginning on the 100th day and working backwards observation to improve the model's ability to forecast the 89th and so forth. The number of training samples will increase by one if needed, when each iteration we return the model. We used 50 latency and 10000 iters.

Company Name Code	Start Date	End Date	No. of Days
UNILEVERCL	05-Jun-2014	16-May-2021	1613
GP	18-Nov-2009	16-May-2021	2707
DUTCHBANGL	26-Nov-2007	16-May-2021	3189
BXPHARMA	11-Aug-2014	16-May-2021	1598
BERGERPBL	26-Nov-2007	16-May-2021	3097
BATBC	03-Jul-2014	16-May-2021	1599

Table I
DATA SUMMARY

To observe sequence of data we have used Forward algorithm. For best hidden state sequence Viterbi algorithm can be used. For optimum model parameter it is solved by Baum-Welch algorithm. To compare the findings, we computed the MAPE and showed the predicted and actual prices. We used minimum likeyhood vector. We then choose the model with the lowest BIC value, which is a function of the number of states, to optimize our model.

IV. RESULTS

In this part, we'll talk about what the end outcome will be. We'll also put our project to the test and compare the results to the final outcome. Before we can start working on the project, we need to gather data. We must also compute the value correctly. Finally, we must put all of these characteristics to the test.

Code	Close	Open	High	Low
UNILEVERCL	0.0254658	0.0279011	0.0271942	0.0241991
GP	0.0138668	0.0179751	0.0137767	0.0136781
DUTCHBANGL	0.013355	0.0156578	0.012921	0.0105929
BXPHARMA	0.0416561	0.0401255	0.0297477	0.0401381
BERGERPBL	0.0254288	0.031876	0.0229392	0.0218087
BATBC	0.0270006	0.0340903	0.0297567	0.0250588

Table II
DATA SUMMARY

V. CONCLUSION

The Hidden Markov Model is a technique for forecasting future data. We utilize it to forecast the stock price of Bangladesh's stock market. After implementing the model, we discovered that the accuracy level is high except Baximco. It

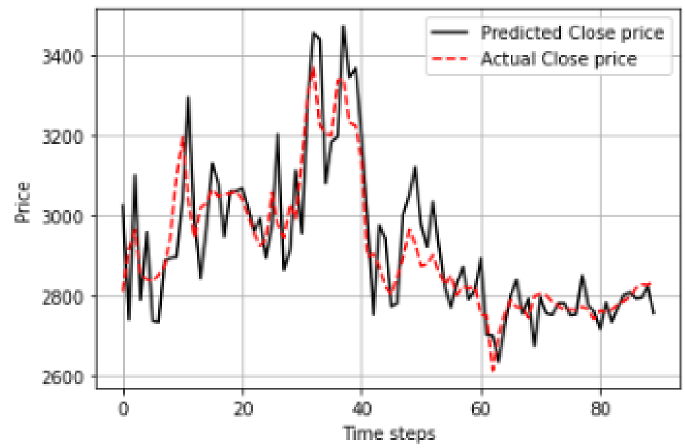

Figure 2. Unilever Consumer Care Limited, Close, State = 21

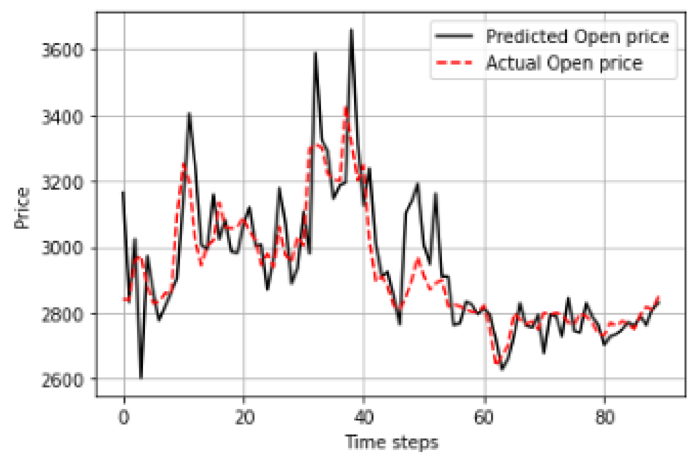

Figure 3. Unilever Consumer Care Limited, Open, State = 21

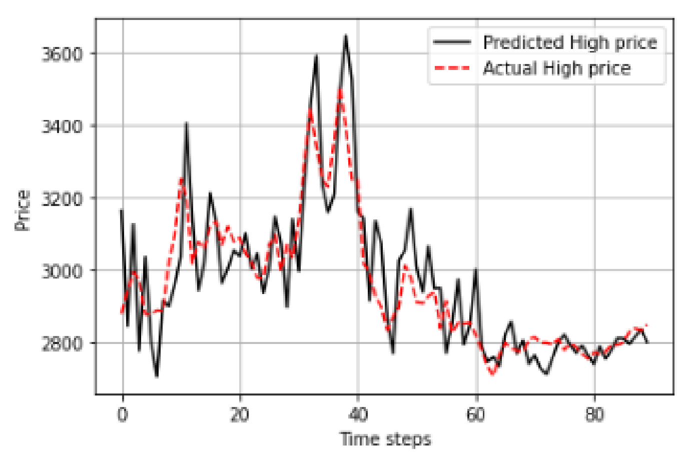

Figure 4. Unilever Consumer Care Limited, High, State = 21

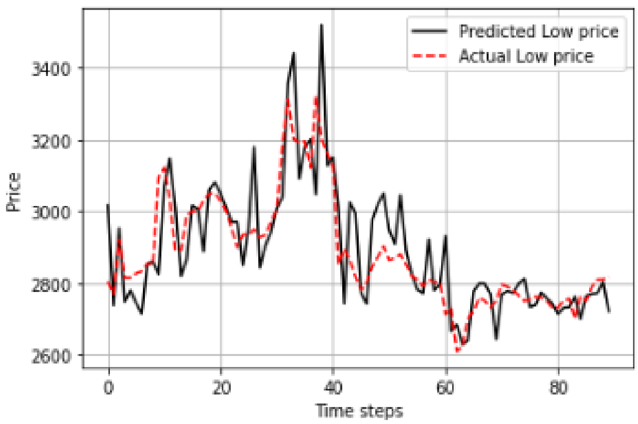

Figure 5. Unilever Consumer Care Limited, Low, State = 21

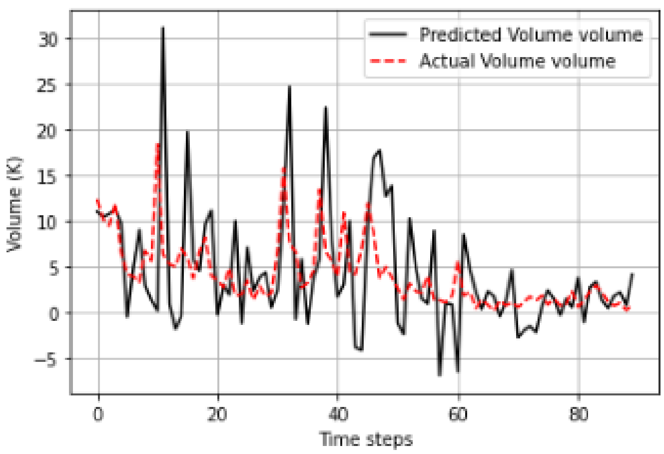

Figure 6. Unilever Consumer Care Limited, Volume, State = 21

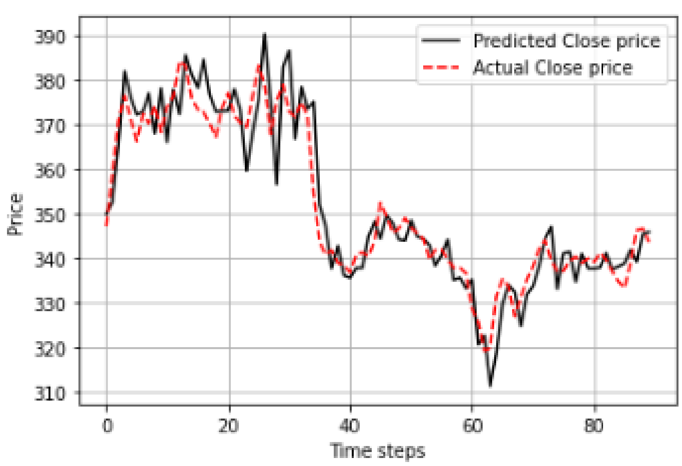

Figure 7. Grameenphone Ltd. Close,State = 24

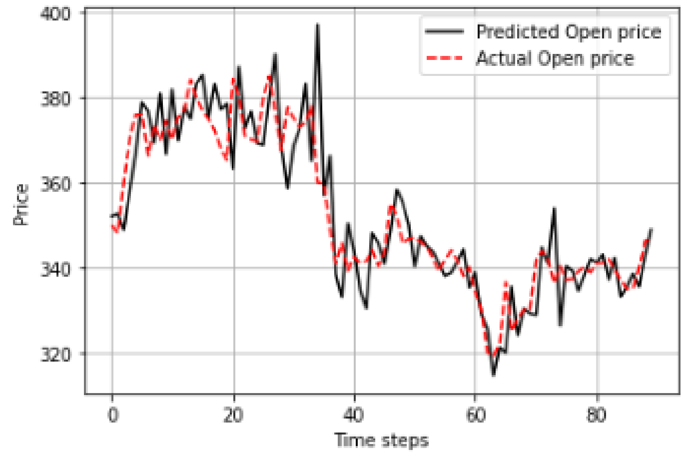

Figure 8. Grameenphone Ltd., Open, State = 24

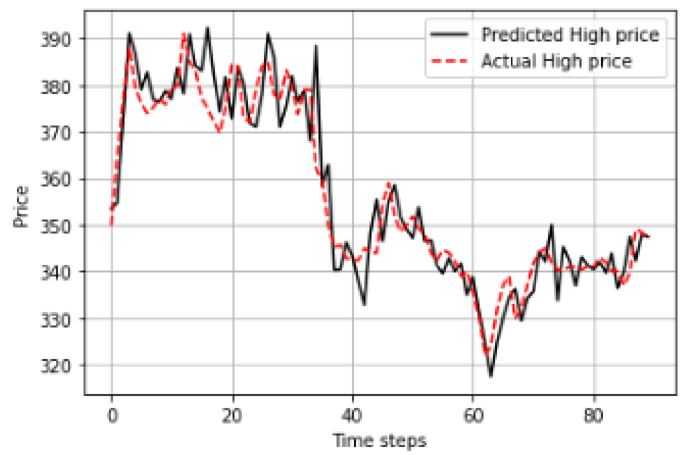

Figure 9. Grameenphone Ltd., High, State = 24

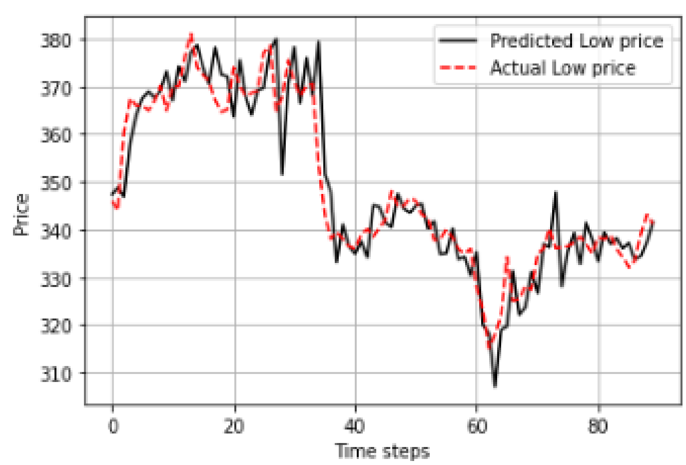

Figure 10. Grameenphone Ltd., Low, State = 24

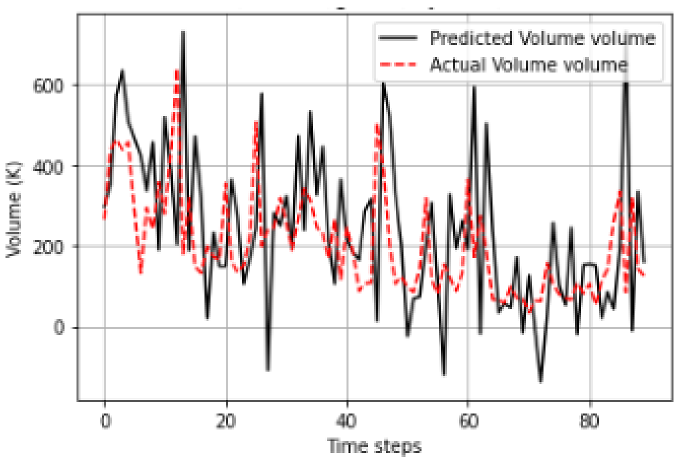

Figure 11. Grameenphone Ltd., Volume, State = 24

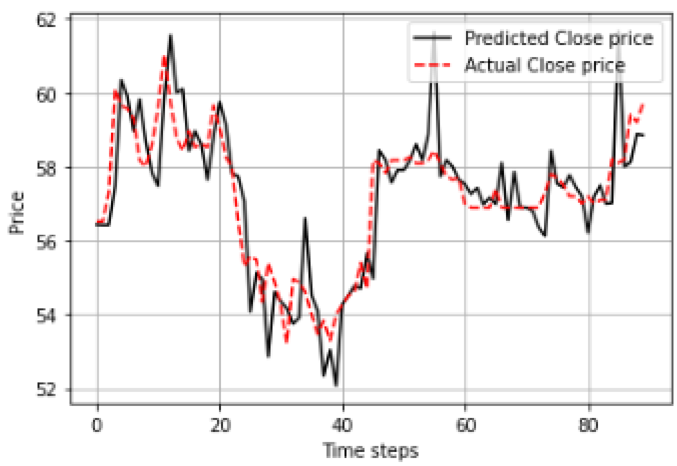

Figure 12. Dutch-Bangla Bank Ltd., Close,State = 24

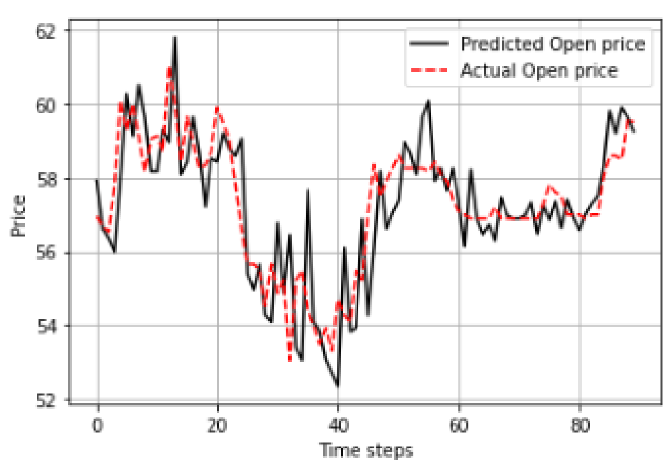

Figure 13. Dutch-Bangla Bank Ltd., Open, State = 24

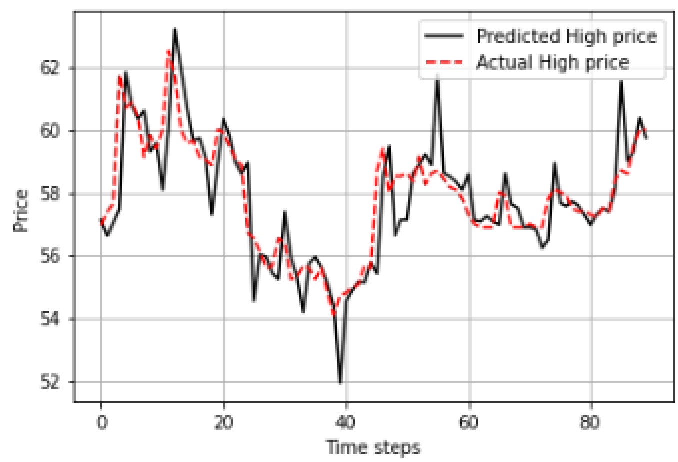

Figure 14. Dutch-Bangla Bank Ltd., High, State = 24

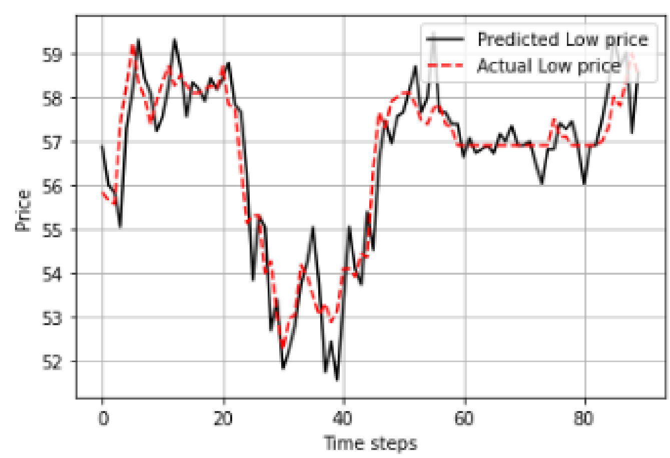

Figure 15. Dutch-Bangla Bank Ltd., Low, State = 24

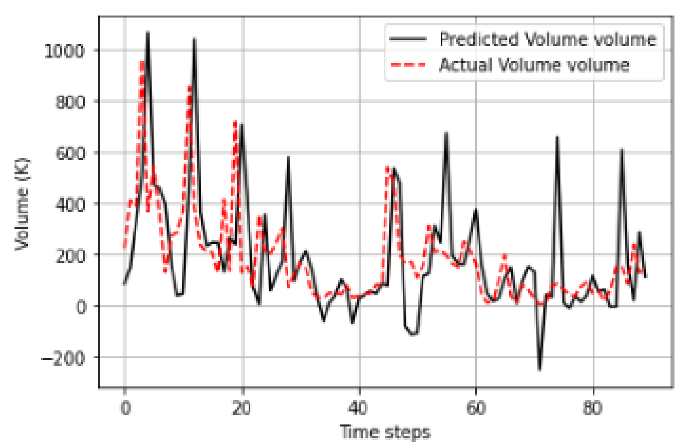

Figure 16. Dutch-Bangla Bank Ltd, Volume, State = 24

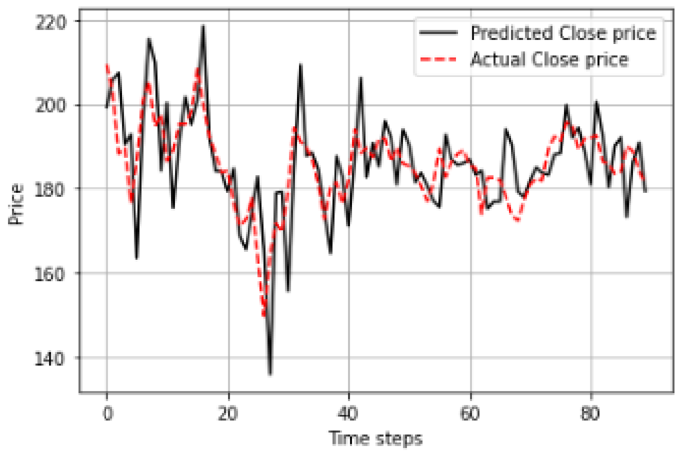

Figure 17. Beximco Pharmaceiticals Ltd., Close,State = 20

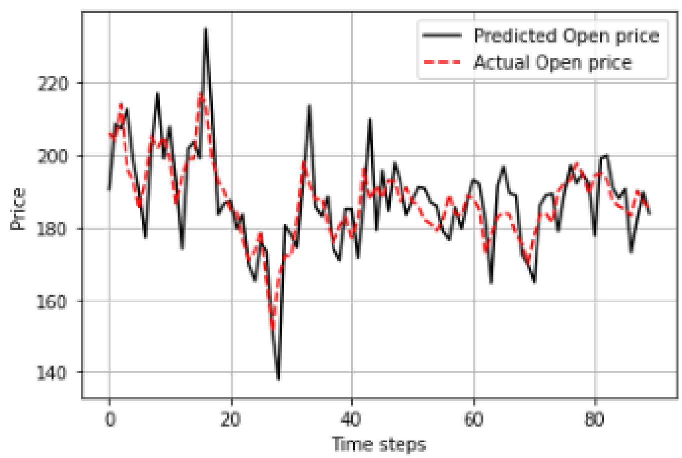

Figure 18. Beximco Pharmaceiticals Ltd.., Open, State = 20

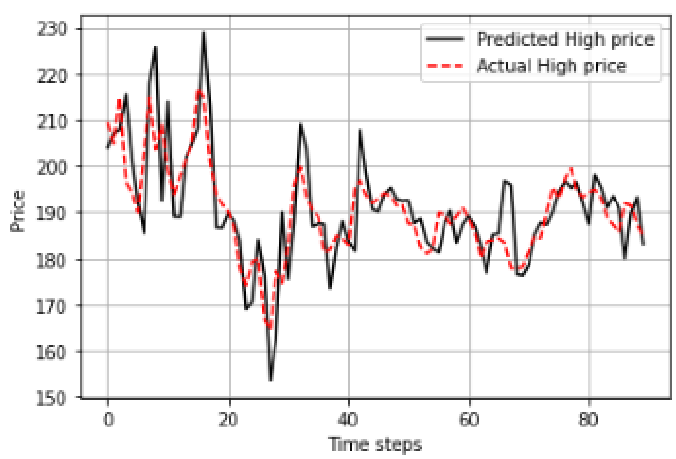

Figure 19. Beximco Pharmaceiticals Ltd.., High, State = 20

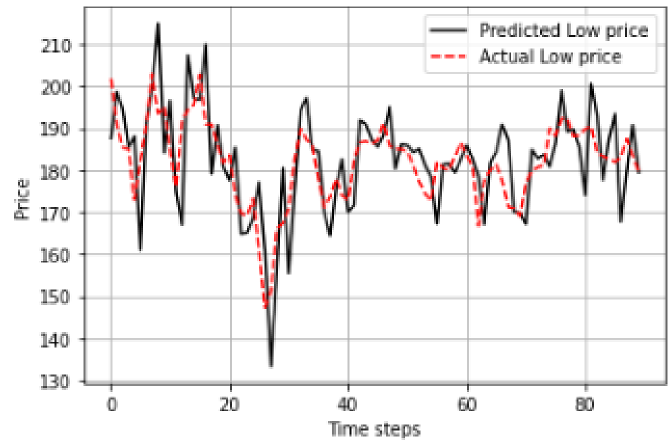

Figure 20. Beximco Pharmaceiticals Ltd.., Low, State = 20

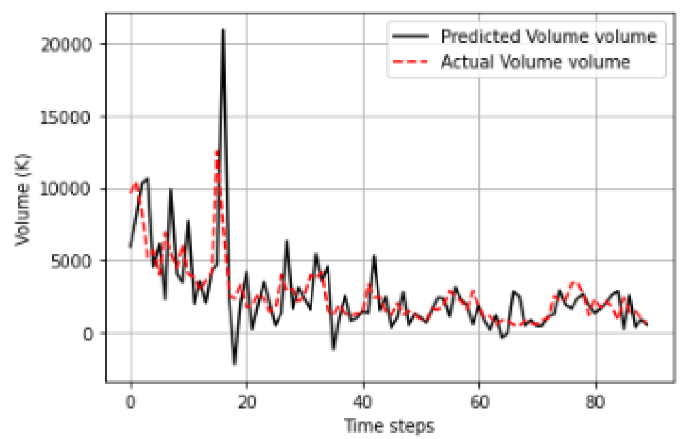

Figure 21. Beximco Pharmaceiticals Ltd. Volume, State = 20

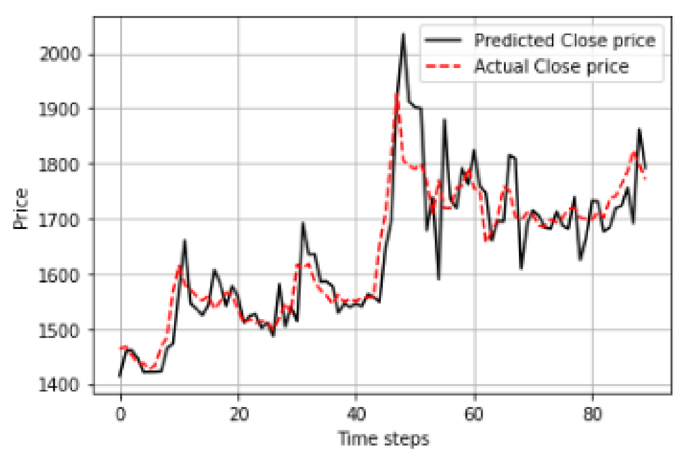

Figure 22. Berger Paints Bangladesh Ltd., Close,State = 24

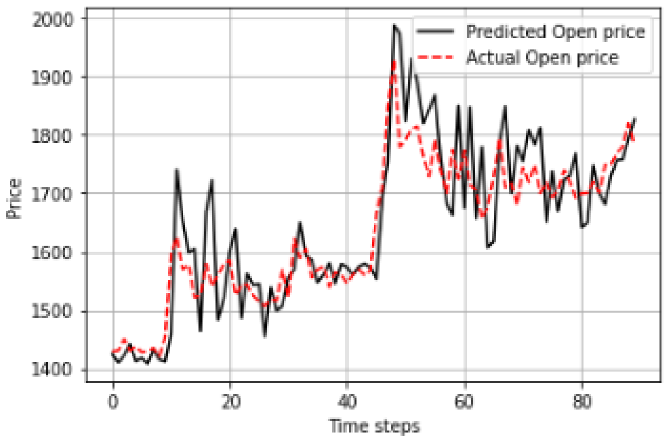

Figure 23. Berger Paints Bangladesh Ltd, Open, State = 24

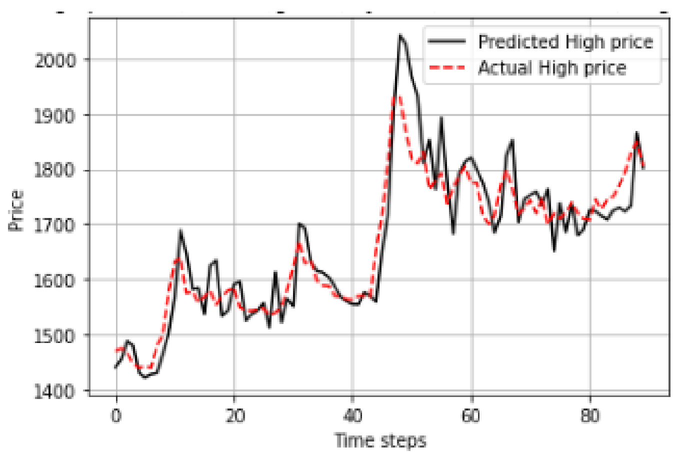

Figure 24. Berger Paints Bangladesh Ltd, High, State = 24

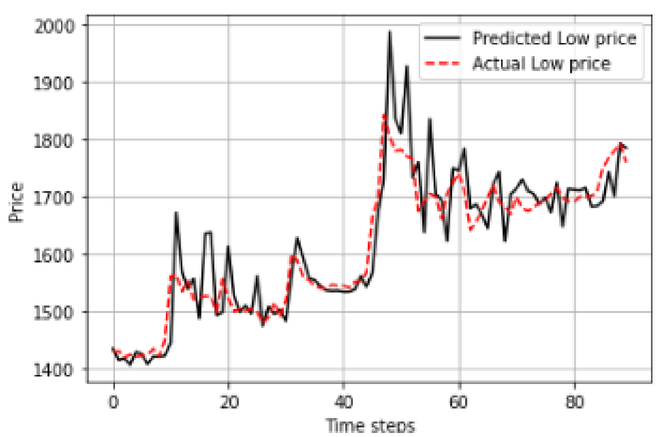

Figure 25. Berger Paints Bangladesh Ltd., Low, State = 24

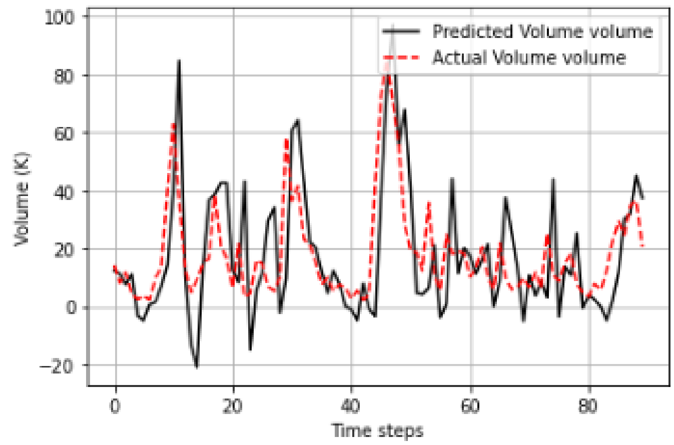

Figure 26. Berger Paints Bangladesh Ltd. Volume, State = 24

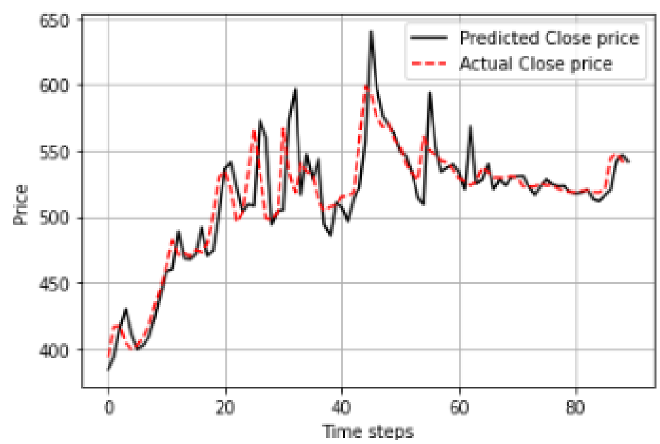

Figure 27. BATBC, Close,State = 23

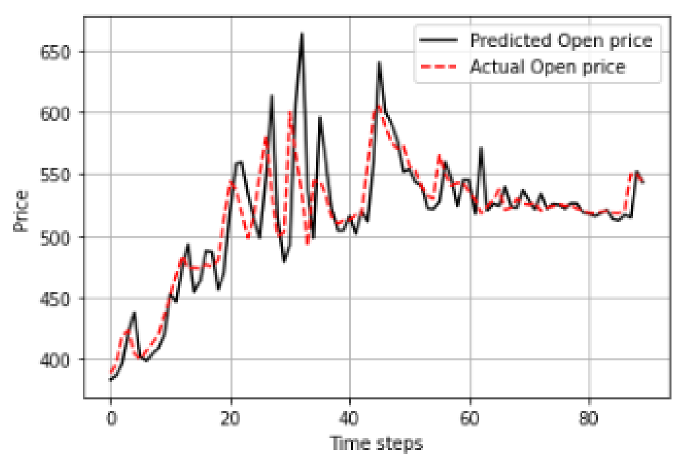

Figure 28. BATBC, Open, State = 23

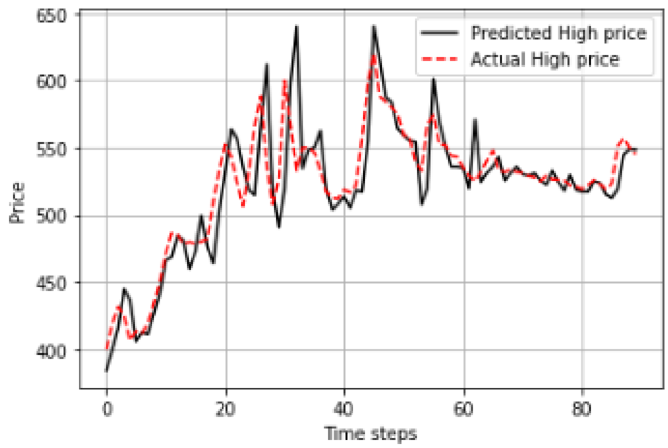

Figure 29. BATBC, High, State = 23

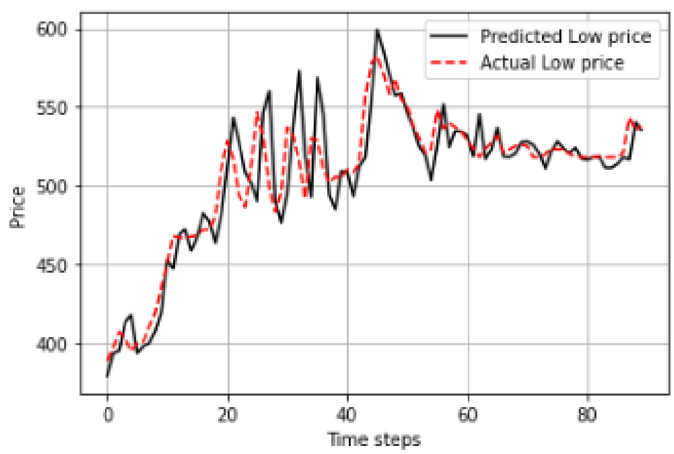

Figure 30. BATBC, Low, State = 23

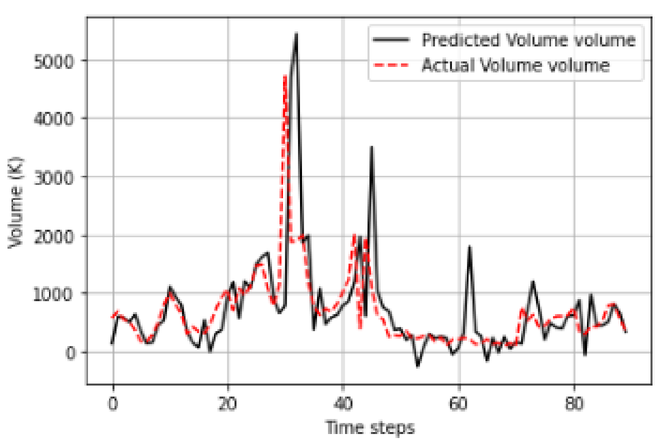

Figure 31. BATBC Volume, State = 23

is due to the Bangladesh's current situation. The stock price in Bangladesh's stock market is determined by a variety of factors.

It's referred to as a factor. The stock price rises and falls as these circumstances change. If we don't, we're doomed.

If those considerations are taken into account, the accuracy level may be higher.

A. *Limitations & future work*

We have some limitation in our project. We can't provide much company for prediction. Our accuracy can be more than resulted value because if we can take more states by handling over-fitting. But it gives good result. HMM can be used for various sector one of our project member has already implemented this approach in his company to calculate:

References

1 Ayush Jain,Prateek Mehta, andRavi Teja Darbha, "Stock Forecasting using Hidden Markov Models." Foundations and trends in information retrieval 2.1-2 (2008): 1-135.J.

2 MD. Almas Hossain,MD. Ariful islam shaikat, "STOCK PREDICTION OF BANGLADESH SHARE MARKET " Daffodil Internation University, Bangladesh (2019):.

3 Han, Jiawei, Micheline Kamber, and Jian Pei. Data mining, southeast asia edition: Concepts and techniques. Morgan kaufmann, 2006.

4 Thammano A (1999), Neuro-fuzzy Model for Stock Market Prediction, Proceedings of the Artificial Neural Networks in Engineering Conference, ASME Press, New York, pp. 587-591.

5 j Huang X, Ariki Y, Jack M (1990), Hidden Markov Models for speech recognition. Edinburgh University Press

6 Md. Rafiul Hassan, Baikunth Nath (2005), Stock Market Forecasting Using Hidden Markov Model: A New Approach, 5th International Conference on Intelligent Systems Design and Applications (ISDA'05).

7 Nguyet Nguyen, Stock Price Prediction using Hidden Markov Model, Youngstown State University, Ohio, USA

8 Hassan, Md. R.; Nath, B.; Stock Market Forecasting Using Hidden Markov Models: A New approach. Proceeding of IEEE fifth International Conference on Intelligent Systems Design and Applications 2005.