Contemporary Mathematics
http://ojs.wiserpub.com/index.php/CM/

Research Article

Trend Analysis of BSE Stock Prices Using Hidden Markov Models and Viterbi Algorithm

S. Aarthi, S. Indrakala*

Department of Mathematics, Kunthavai Naacchiyaar Government Arts College for Women (Autonomous)Affiliated to Bharathidasan University, Thanjavur, 613007, Tamil Nadu, India
E-mail: s.indirakala@yahoo.com

Received: 19 April 2024; **Revised:** 16 June 2024; **Accepted:** 18 June 2024

Abstract: Many forecasting techniques have been put forth and used in recent years to predict stock market trends. Recently, many researchers developed models based on Artificial Neural Network (ANN), Support Vector Machine (SVM), Fuzzy Logic (FL) and Moving Average (MA). This paper presents the trend analysis of the stock market prediction using the Hidden Markov Model and Viterbi algorithm with a 1-day, 2-day, 3-day, 4-day and 5-day variation in the close value for the specified time frame. In this work, we developed a BSE price forecasting model based on Hidden Markov Model due to its proven fittingness for modeling vigorous systems and pattern classification. We apply the HMM methodology to forecast the BSE closing price from Jan 2021 to Dec 2021 using available past datasets from Investopedia. The trend percentage of stock prices, which is computed for every observed sequence and hidden sequence, is provided by the probability values π. In situations of uncertainty, decision makers can use the proportion of probability values derived from the steady state probability distribution as a guide when making judgments.

***Keywords*:** stock trend analysis, HMM, viterbi algorithm, TPM, EPM, optimal state sequence

MSC: 62M05, 47N30, 60G25, 60J22, 93A30

1. Introduction

With more companies going public in recent years, stocks have emerged as a popular topic in the financial industry. A growing number of financial analysts and investors are also interested with stock price prediction because, on the one hand, the trend of stock price will influence the trend of various economic behaviors to some degree. However, with more people investing in the stock market every year, the only way to swiftly identify market trends and increase investment returns is to precisely analyze the future trajectory of stock prices. Financial research focuses on stock price prediction, which is typically seen as a difficult undertaking due to the extreme volatility of financial markets.

For the AI community, one of the most challenging problems has been stock price forecasting. Forecasting research has usually been surpassed by traditional AI research, which is mostly concerned with developing intelligent solutions intended to mimic human intelligence. However, due to its nonstationary, cyclical, and stochastic nature, stock price forecasting is still very limited. The rate at which prices fluctuate in such a series is influenced by a number of variables, including as equity, interest rates, securities, warrants, options, and mergers and acquisitions of noteworthy

DOI: https://doi.org/10.37256/cm.6120254788

financial institutions. Ordinary investors cannot consistently earn in such a market. Because of this, regular investors would be very interested in and in demand for an intelligent stock market predicting program.In an efficient market, stock prices would be determined primarily by fundamentals, which, at the basic level, refer to a combination of two things:

• An earnings base, such as earnings per share (EPS).

• A valuation multiple, such as a P/E ratio.

It is common practice in the corporate world, particularly in the stock market, to hire a competent advisor or someone to monitor stock prices and trends. This is done in order to investigate the chaotic system that contains a lot of data and different elements that affect the stock price. Many researchers gave presentations on their research on stock market prediction, which is confined to the index status alone. However, because the anticipated outcomes depend on several variables or predictors, this paper forecast and predict the closing prices more precisely, which is more complex. Company stocks typically follow the market, as well as the peers in their industry or sector. Some well-known investment businesses contend that most of a stock's movement is determined by the mix of sector and market movements as a whole rather than by the performance of any one company (Studies have indicated that 90% of it can be attributed to economic and market forces). For instance, "guilt by association" lowers demand for the entire sector when one retail stock suddenly has a poor outlook, which typically hits other retail companies.

The selection of target stocks and the prediction of stock prices in the traditional quantitative investment field are primarily based on the outcomes of long-term stock market experience [1]. Empirical stock analysis methods are difficult to spread and promote and frequently have poor antirisk and long-term prediction abilities [2].

Furthermore, the conventional methods' analysis speed was frequently slow. Consequently, statistical and financial approaches to stock analysis emerged, marking the start of mathematical stock modeling. These approaches include the autoregressive model [3], the stochastic volatility model [4], and the Markov model [5]. Their predictive and analytical capabilities surpass those of empirical methods. Furthermore, because mathematical modeling is used, these models can only be applied to the present large-scale data scenarios and are best suited for computer analyses, which are based on sparse input data.

The stock price is an observable time series, meaning that its determinants are unknown variables. This characteristic is congruent with the hidden Markov model (HMM), which numerous academics have used to predict stock prices. A statistical model called HMM has been applied to image processing, pattern identification, DNA sequence analysis, and automatic speech recognition [6]. This paper's primary contribution is the construction of a HMM based stock price prediction model. The rest of this paper is designed as follows. Proposed Scheme is reviewed in Section 2. The HMM and the Viterbi Algorithm-based stock price prediction model is introduced in Section 3. Section 4 presents the outcomes of the experiment. Section 5 concludes the paper.

2. Proposed scheme

A machine with a finite number of states is called a Hidden Markov Model (HMM). It offers a probabilistic framework for modeling a multivariate time series of observations. The use of hidden Markov models as a voice recognition tool dates back to the early 1970s. Due to its solid theoretical foundation and robust mathematical structure, this statistically based model has gained popularity over the past several years in a variety of fields. It is evident that HMM is an incredibly useful tool with a wide range of uses. The benefits of HMM can be summed up as follows:

• It can handle fresh data robustly.

• It has a strong statistical base.

• Efficient in terms of computation to create and assess (since pre-existing training algorithms exist).

• It has an effective ability to predict similar patterns.

The fundamentals of HMM and how it might be applied to signal prediction are covered in the Rabiner tutorial. The HMM, in contrast to the Markov chain, will select a specific course of action based on the observation probability as well. The ability of the HMM to select the optimal overall strategy sequence given an observation sequence is crucial to its performance. More representation capacity is provided by adding density functions to the HMM's states than by

predefined methods linked to the states. New patterns can be discovered by the unsupervised learning method of the general HMM approach framework. It is not necessary to impose a "template" for it to learn because it can tolerate input sequences of varying lengths. The HMM is briefly explained in the following session.

3. Hidden markov model

An HMM comprises of a five-tuple: $(\boldsymbol{S}, \boldsymbol{K}, \boldsymbol{\Pi}, \boldsymbol{A}, \boldsymbol{B})$.
• $S = \{1, \ldots, N\}$ is the set of states. s_t denotes the state at time t.
• $K = \{k_1, \ldots, k_M\}$ is the output alphabet. M is the number of observation ranges.
• $\Pi = \{\pi_i, i \in S\}$ is the initial state distribution and π_i is defined as

$$\pi_i = P(s_1 = i)$$

• Transition probability distribution $A = \{a_{ij}\}, i, j \in S$.

$$a_{ij} = P(s_{t+1} \mid s_t), \quad 1 \leq i, j \leq N$$

• Emission probability distribution $B = b_j(o_t)$.
The probabilistic function for each state j is:

$$b_j(o_t) = P(o_t \mid s_t = j)$$

We can determine the probability of the observation sequence and the likely underlying state sequences by modeling a problem as an HMM and assuming that the HMM produced a certain set of data. In order to create a more accurate model, we can also train the model's parameters using the observed data. Next, make use of the learned model to forecast unknown data.

Given an observation sequence $O = (o1, \ldots, oT)$ and an $HMM\mu = (\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\Pi})$, find the probability of the sequence $P(O \mid \mu)$. This process is known as decoding. Here the observations are independent of each other the time t, the probability of a state sequence $S = (s1, \ldots, sT)$ generating the observation sequence can be calculated as:

$$P(O \mid S, \mu) = \prod_{t=1}^{T} P(o_t \mid s_t, s_{t+1}, \mu) \tag{1}$$

$$= b_{s_1}(o_1) \cdots b_{s_1 s_2}(o_2) \cdots b_{s_{T-1} s_T}(o_T) \tag{2}$$

and the state transition probability,

$$P(S \mid \mu) = \pi_{s_1} \cdot a_{s_1 s_2} \cdot a_{s_2 s_3} \cdots a_{s_{T-1} s_T}$$

The joint probability of O and S :

$$P(O,\ S \mid \mu) = P(O \mid S,\ \mu)P(S \mid \mu) \tag{3}$$

Therefore,

$$P(O \mid \mu) = \sum_{S} P(O \mid S,\ \mu)P(S \mid \mu) \tag{4}$$

$$= \sum_{s_1 \ldots\ldots s_{T+1}} \pi_{s_1} \prod_{t=1}^{T} a_{s_t s_{t+1}} b_{s_t s_{t+1} o_t} \tag{5}$$

By adding up the observation probabilities for every potential state sequence, the calculation is rather simple. The computation increases exponentially with the length of T in the sequence. It requires $(2T-1)$. N^{T+1} multiplications and $N^T - 1$ additions.

4. Viterbi algorithm

According to [7] the Viterbi algorithm aims to find the optimal estimate for the hidden state sequence within HMM, conditional on a series of system measurements. At each stage, the Viterbi algorithm finds the optimal value for the state in the order, and continues the analysis to the next stage in the inductive way. To find the optimal order in the hidden state

$Q = (q_1,\ q_2,\ \ldots\ q_n)$ in the realization of HMM, conditional on the measurement sequence system $O = (o_1,\ o_2,\ \ldots\ o_n)$, the following variables are defined:

$$V(j) = \max_{Q} P(Q = j \mid \lambda) \tag{6}$$

Where, Q denotes hidden stata sequence, O denotes observed sequence and $V(j)$ is the optimal value for HMM at the time n, considering the first state of S_i as the condition. The value of $V_n(j)$ is calculated as follows:

$$V_n(j) = \max_{i=1} V_{n-1}(i) P_{ij} b_j(o_n) \tag{7}$$

Then to obtain the best value of P calculated using following formula

$$P = \max_{i=1}(i) \tag{8}$$

The three factors are multiplied in equation (7) to extend the previous path by calculating the Viterbi probability at time n are follows:

- $V_{n-1}(i)$ is the probability of the previous Viterbi path from the previous time step.
- P_{ij} is the transition probability from the previous state to the current state.
- $b_j(o_n)$ is the observation state against the observation symbol given the current state j.

5. Mathematical results and discussion

In this section, the data has been taken from Investopedia.com and the analysis focused on BSE daily close value data from January 2021 to December 2021. We used two observation symbols: "I" for increasing states and "D" for decreasing states and it is observed that the symbol is "I" if the differences in close values are larger than 0 and that the sign is "D" if the differences in close values are less than 0. The symbols $S1,\ S2,\ S3$ and $S4$ stands for the four hypothesized concealed states, which are low, moderate low, moderate high and high respectively. The states cannot be observed immediately. The stock advertisement's conditions are thought to be concealed. Given an arrangement of perception we are able discover the covered up state grouping that created those perceptions. Figure 1 gives the daily closing value of the BSE stock price values.

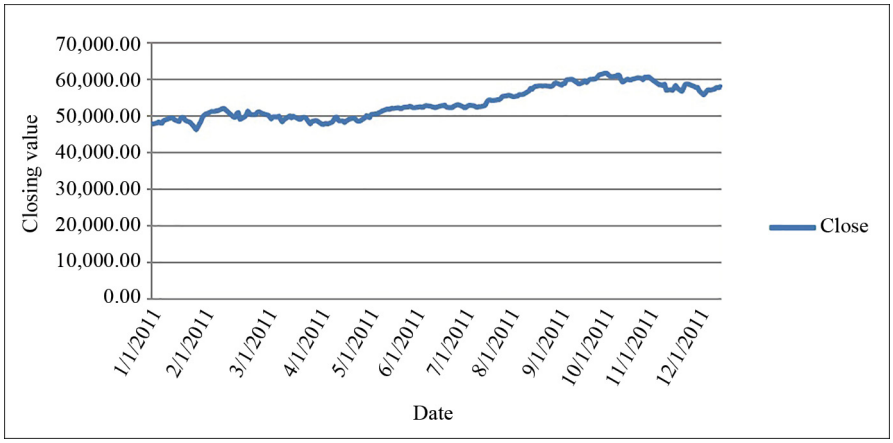

Figure 1. Daily closing value of BSE

Table 1. Interval values

Energy states	
$S1 = -15,000$ to -500	Low (L)
$S2 = -501$ to 250	Moderate low (ML)
$S3 = 251$ to 500	Moderate high (MH)
$S4 = 501$ to 15,000	High (H)

The various probability values of TPM, EPM and π for difference in one day, two days, three day, four days and five days are calculated and given as below in Table 1-11:

Initial State Probability for difference in one day close value:

$$\pi_1 = \begin{bmatrix} 0.16194332 & 0.582995951 & 0.133603239 & 0.12145749 \end{bmatrix}$$

Table 2. TPM for difference in 1 day close value

	L	ML	MH	H
L	0.225	0.65	0.025	0.1
ML	0.118881119	0.65034965	0.125874126	0.104895
MH	0.151515152	0.454545455	0.333333333	0.060606
H	0.3	0.3	0.1	0.3

Table 3. EPM for diff in 2 day

	I	D
L	0	1
ML	0.354166667	0.645833333
MH	1	0
H	1	0

Initial State Probability for difference in two day close value:

$$\pi_2 = \begin{bmatrix} 0.199186992 & 0.528455 & 0.097561 & 0.174797 \end{bmatrix}$$

Table 4. TPM for difference in 2 day close value

	L	ML	MH	H
L	0.102041	0.510204	0.102041	0.285714
ML	0.184615	0.515385	0.130769	0.169231
MH	0.217391	0.608696	0.043478	0.130435
H	0.348837	0.534884	0.023256	0.093023

Table 5. EPM for diff in 2 day

	I	D
L	0	1
ML	0.354166667	0.645833333
L	0	1
ML	0.369231	0.630769
MH	1	0
H	1	0

Initial State Probability for difference in three days close value:

$$\pi_3 = \begin{bmatrix} 0.310204082 & 0.302041 & 0.102041 & 0.285714 \end{bmatrix}$$

Table 6. TPM for difference in 3 day close value

	L	ML	MH	H
L	0.157895	0.197368	0.144737	0.5
ML	0.178082	0.452055	0.082192	0.287671
MH	0.48	0.24	0.12	0.16
H	0.557143	0.285714	0.057143	0.1

Table 7. EPM for diff in 3 day

	I	D
L	0	1
ML	0.337838	0.662162
MH	1	0
H	1	0

Initial State Probability for difference in four days close value:

$$\pi_4 = \begin{bmatrix} 0.352459016 & 0.217213 & 0.045082 & 0.385246 \end{bmatrix}$$

Table 8. TPM for difference in 4 day close value

	L	ML	MH	H
L	0.104651	0.127907	0.05814	0.709302
ML	0.346154	0.288462	0.057692	0.307692
MH	0.454545	0.090909	0	0.454545
H	0.574468	0.276596	0.031915	0.117021

Table 9. EPM for diff in 4 day

	I	D
L	0	1
ML	0.264151	0.735849
MH	1	0
H	1	0

Initial State Probability for difference in five days close value:

$$\pi_5 = \begin{bmatrix} 0.423868313 & 0.090535 & 0.037037 & 0.44856 \end{bmatrix}$$

Table 10. TPM for difference in 5 day close value

	L	ML	MH	H
L	0.184466	0.038835	0.009709	0.76699
ML	0.454545	0.090909	0.136364	0.318182
MH	0.444444	0.111111	0.111111	0.333333
H	0.648148	0.138889	0.037037	0.175926

Table 11. EPM for diff in 5 day

	I	D
L	0	1
ML	0.363636	0.636364
MH	1	0
H	1	0

Here N = No. of Hidden States = 4 = $\{L, ML, MH, H\}$;

T = No. of Observations = 2 = $\{I, D\}$.

So we have $m = N^T = 16$ combination of sequences. Using TPM, EPM, initial probability and joint probability formula as stated in (1), (2) we calculate the probabilities for 16 combination of sequences and find the maximum probability which gives the likelihood hidden state sequence as D, D, I, I for the difference in one day, two days, three days, four days and five days closing value using HMM.

Probabilities for the observation on Hidden State using Viterbi Algorithm in Figure 2 are given below:

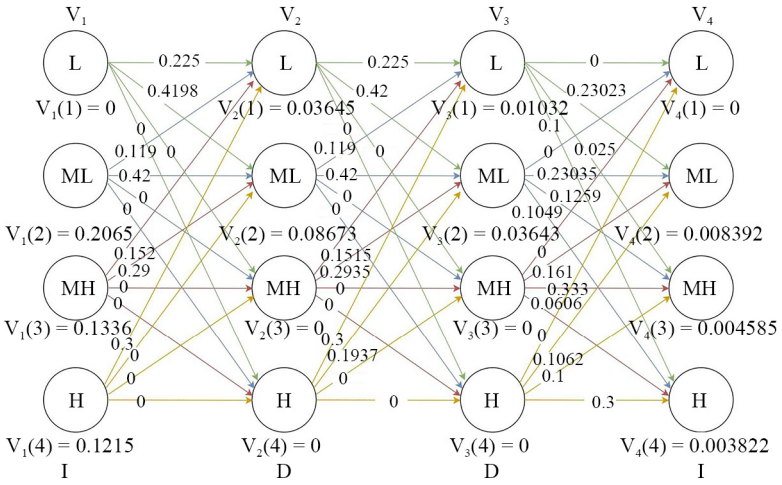

Figure 2. Probabilities for the observation for the difference in one day close value using the Viterbi Algorithm

From the above figure, in the first column (i.e., in the Increasing state) $V_1(2)$ has the maximum value, in 2nd column (i.e., in the Decreasing state) $V_2(2)$ has the maximum value, in 3rd column $V_3(2)$ has the maximum value and in the 4th column $V_4(2)$ has the maximum value. Therefore the maximum values in the Viterbi path are

$$V_1(2) = 0.2065,\ V_2(2) = 0.08673,\ V_3(2) = 0.03643,\ V_4(2) = 0.008392.$$

So the best sequence for the difference in one day close value in the order of the first day is , the second day is *ML*, the third day is *ML* and the fourth day is *ML*.

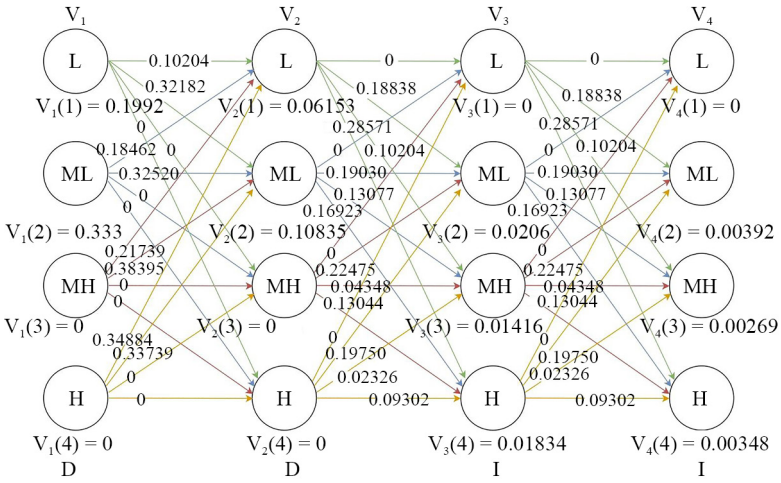

Figure 3. Probabilities for the observation for the difference in two days close value using the Viterbi Algorithm

From the Figure 3, in the first column (i.e., in the Increasing state) $V_1(2)$ has the maximum value, in 2nd column (i.e., in the Decreasing state) $V_2(2)$ has the maximum value, in 3rd column $V_3(2)$ has the maximum value and in the 4th column $V_4(2)$ has the maximum value. Therefore, the maximum value in the Viterbi path are

$$V_1(2) = 0.333,\ V_2(2) = 0.10835,\ V_3(2) = 0.0206,\ V_4(2) = 0.00392.$$

So the best sequence for the difference in two days close value in the order of the first day is *ML*, the second day is *ML*, the third day is *ML* and the fourth day is *ML*.

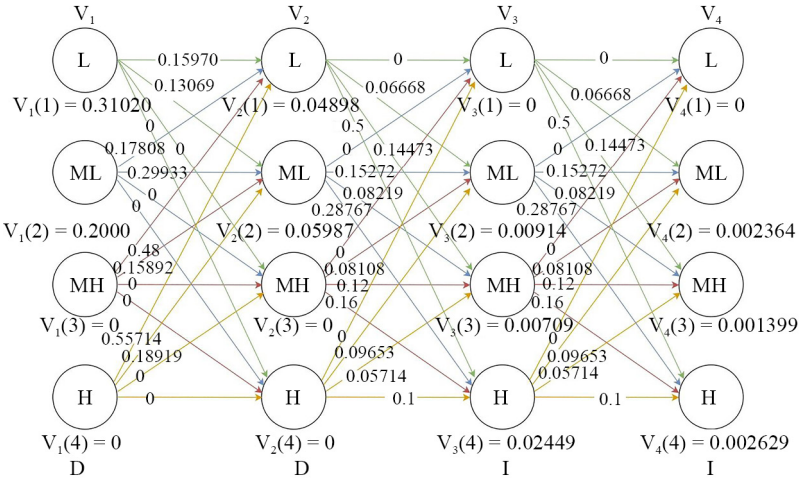

Figure 4. Probabilities for the observation for the difference in three days close value using the Viterbi Algorithm

From the Figure 4, in the first column (i.e., in the Increasing state) $V_1(1)$ has the maximum value, in 2nd column (i.e., in the Decreasing state) $V_2(2)$ has the maximum value, in 3rd column $V_3(4)$ has the maximum value and in the 4th column $V_4(4)$ has the maximum value. Therefore, the maximum value in the Viterbi path are

$$V_1(1) = 0.31020,\ V_2(2) = 0.05987,\ V_3(4) = 0.02449,\ V_4(4) = 0.002629.$$

So the best sequence for the difference in three days close value in the order of the first day is *L*, the second day is *ML*, the third day is *H* and the fourth day is *H*.

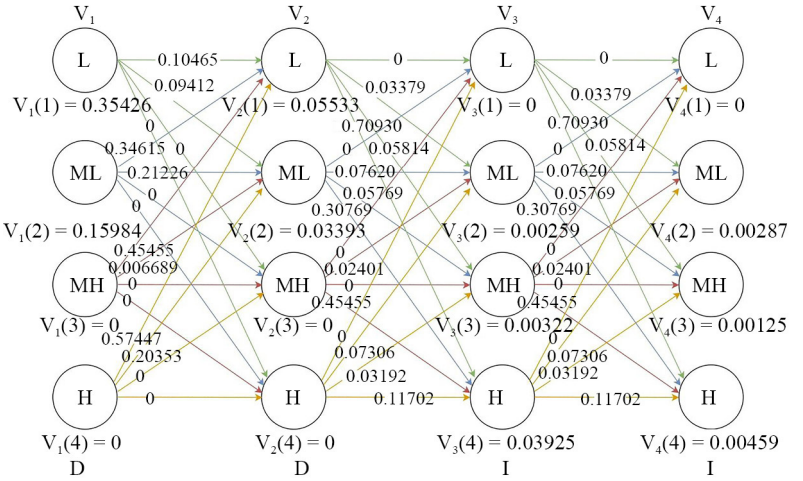

Figure 5. Probabilities for the observation for the difference in four days close value using the Viterbi Algorithm

From the Figure 5, in the first columm (i.e., in the Increasing state) $V_1(1)$ has the maximum value, in 2nd column (i.e., in the Decreasing state) V_2 (1) has the maximum value, in 3rd column V_3 (4) has the maximum value and in the 4th column $V_4(4)$ has the maximum value. Therefore, the maximum value in the Viterbi path are

$$V_1(1) = 0.35246,\ V_2(1) = 0.05533,\ V_3(4) = 0.03925,\ V_4(4) = 0.00459.$$

So the best sequence for the difference in four days close value in the order of the first day is *I*., the second day is *L*, the third day is *H* and the fourth day is *H*.

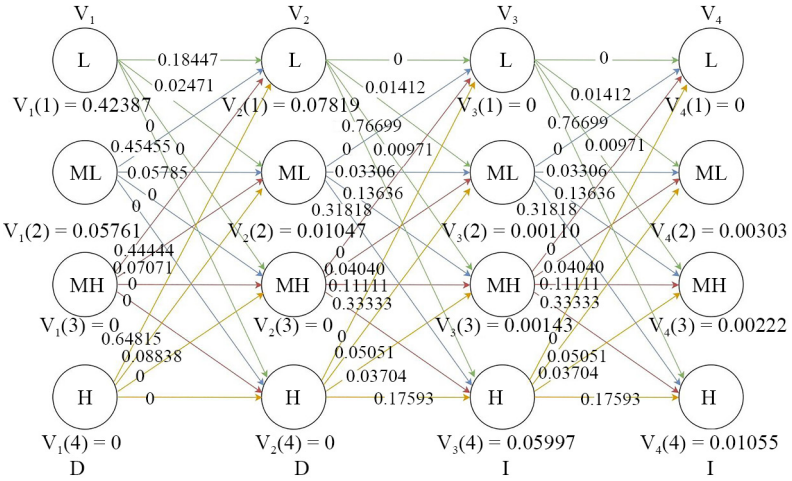

Figure 6. Probabilities for the observation for the difference in five days close value using the Viterbi Algorithm

From the Figure 6, in the first column (i.e., in the Increasing state) $V_1(1)$ has the maximum value, in 2nd column (i.e., in the Decreasing state) V_2 (1) has the maximum value, in 3rd column $V_3(4)$ has the maximum value and in the 4d. column $V_4(4)$ has the maximum value. Therefore, the maximum value in the Viterbi path are

$$V_1(1) = 0.42387,\ V_2(1) = 0.07819,\ V_3(4) = 0.05997,\ V_4(4) = 0.01055$$

So the best sequence for the difference in five days close value in the order of the first day is *L*, the second day is *L*, the third day is *I* and the fourth day is *II*.

The optimum sequence of states obtained from the all five day's differences using IIMM and Viterbi algorithm with TPM and EPM is given below:

1.	D S2	→	D S2	→	I S2	→	I S2
2.	D S2	→	D S2	→	I S2	→	I S2
3.	D S1	→	D S2	→	I S4	→	I S4
4.	D S1	→	D S1	→	I S4	→	I S4
5.	D S1	→	D S1	→	I S4	→	I S4

However, using HMM we can easily predict whether the next day will be increasing (I) or decreasing (D). By using the Viterbi algorithm, we can provide the best route for the state in the order of the first day is L (low), the second day is L (low), third day is H (high) and fourth day is H (high). Hence, the five day difference of TPM and EPM has the shortest path. So the best optimum sequence is found from five day difference in close value.

6. Conclusion

In order to account for hidden states that may have an impact on precise forecasting, the HMM prediction method generates starting state, transition probability, and emission probability. Here, the stock market's four states were quickly identified by the Hidden Markov model, which was also utilized to forecast future values. The higher performance of the specific sequence has the highest value in the Optimum State Sequences. Predicting the ideal sequence is more accurate with the suggested model. To make it simple to determine if the sequence's level is increasing or decreasing for the next day, hidden states and sequences have been created. Additionally, it was determined whether the amount of increase was moderate high or high, and whether the level of decrease was moderate low or low. Investors with both long-term and short-term goals will find great value in this strategy.

Conflict of interest

The authors declare no competing financial interest.

References

[1] Tang C, Zhu W, Yu X. Deep hierarchical strategy model for multi-source driven quantitative investment. *IEEE Access*. 2019; 7: 79331-79336. Available from: https://doi.org/10.1109/ACCESS.2019.2923267.

[2] Tiwari AK, Dar AB, Bhanja N, Gupta R. A historical analysis of the US stock price index using empirical mode decomposition over 1791-2015. *Economics-The Open Access Open-Assessment E-Journal*. 2016; 10(9): 1-15.

[3] Khan MK, Teng J, Khan MI, Khan MF. Stock market reaction to macroeconomic variables: An assessment with dynamic autoregressive distributed lag simulations. *International Journal of Finance and Economics*. 2021; 28(3): 2436-2448.

[4] Alghalith M. Estimating the stock portfolio volatility and the volatility of volatility: A new simple method. *Econometric Reviews*. 2016; 35(2): 257-262.

[5] Huang JC, Huang WT, Chu PT, Lee WY. Applying a markov chain for the stock pricing of a novel forecasting model. *Communications in Statistics-Theory and Methods*. 2017; 46(9): 4388-4402.

[6] Palaz D, Magimai-Doss M, Collobert R. End-to-end acoustic modeling using convolutional neural networks for HMM-based automatic speech recognition. *Speech Communication*. 2019; 108: 15-32. Available from: https://doi.org/10.1016/j.specom.2019.01.004.

[7] Su Z, Yi B. Research on HMM-based efficient stock price prediction. *Mobile Information Systems*. 2022; 2022(10): 1-8.