Visualizing and Forecasting Stocks Using Machine Learning

Shanthi Swaroopa H P [1], Mohammed Fiquar [2], Laxmi V[3]

[1]Dept. ISE, BNMIT
[2]Dept. ISE, BNMIT
[3]Asst. Professor, Dept. ISE, BNMIT

---***---

Abstract - *One of the most complicated systems that is nearly hard to represent using dynamical equations is the stock market. The primary reason is that a number of ambiguous factors, such as the state of the economy, a company's policies, and investor demand and supply, and so on, affect stock prices. There has been a lot of research in predicting the behavior of stocks based on their historical performance using Artificial Intelligence and Machine Learning techniques like- Artificial Neural Networks, Fuzzy logic and Support Vector Regression. These models have predicted the prices with less accuracy. Hidden Markov Models are one technique for stock market analysis that is less popular than the approaches previously described. This study looks for a hidden link between matching stock prices and the Hidden Markov Model, unlike the stock forecasts that are previously accessible. Therefore, in this study, the concentration is on Hidden Markov Models and evaluate their effectiveness in comparison to Support Vector Regression Models. The experimental findings show that this technique may deliver surprisingly accurate results, especially for short-term forecasts, where the reached accuracy is > 80%.*

Key Words: Hidden Markov Model

1. INTRODUCTION

For identifying patterns in random processes, Hidden Markov Models provide a robust probabilistic foundation. They have been and are currently widely used for identifying patterns in speech, handwriting, and gestures. A wide range of DNA sequences were successfully analysed by HMMs in the future. HMMs' core tenet is that the likelihood of an observation depends on the system states that are 'hidden' from the observer. The name Hidden Markov Models comes from the fact that the change from one state to another is a Markov Process, meaning that the next state depends solely on the current one. States whereas the observations may be either discrete, continuous, or both, in HMMs are always discrete. The stock market may be thought of as a Hidden Markov Process, where the investor can only see the stock values and does not know what underlying states are influencing those prices. For the purpose of creating a prototype, considering the close price for the TATA MOTORS stock. In this project, its assumed that the distribution of the observations is multivariate Gaussian.

2. RELATED WORK (LITERATURE SURVEY)

HMMs have been used in the past to make stock market forecasts. By looking for a similar trend in the historical data, Hassan and Nath [1] employed fixed state HMMs to predict several airline stocks. In order to measure the performance of the model with regard to the number of hidden states, Nguyet Nguyen [2] expanded on the work of Hassan and Nath by using the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). To increase the possibility of observing all likely sequences, Aditya and Bhuwan [3] employed a fixed state HMM-based MAP estimator.

A HMM-Fuzzy logic model and an artificial neural network were used to compare their findings. Utilizing Support Vector Regression (SVR), which was developed by Cao and Francis [7], is an additional strategy. In order to compare the performance of the HMM model to that of Nguyet Nguyen's SVR model, an HMM model is developed comparable to his.

3. PROPOSED SYSTEM ARCHITECTURE

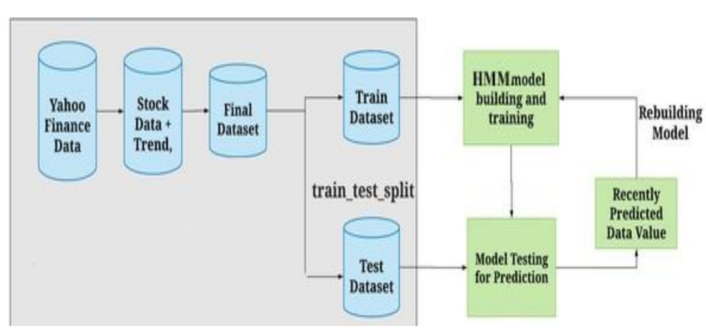

Fig.1 System Architecture

3.1 A media entity that is a component of the Yahoo! network is Yahoo! Finance. In addition to stock quotations, news articles, financial reports, and unique material, it offers financial news, data, and opinion. Additionally, it provides a few online tools for managing personal finances. It publishes unique material written by its own writers in addition to partner content from other websites. On a list of the biggest news and media websites, it is placed 21st by SimilarWeb. Table 1 shows the dataset obtained from yfinance.

Month	Ordinary Shares						'A' Ordinary Shares					
	BSE			NSE			BSE			NSE		
	High (₹)	Low (₹)	No. of Shares	High (₹)	Low (₹)	No. of Shares	High (₹)	Low (₹)	No. of Shares	High (₹)	Low (₹)	No. of Shares
Apr-16	419.30	370.50	9162156	419.05	370.50	124230334	306.55	272.05	1974851	306.50	271.25	34228686
May-16	458.20	380.10	18347837	459.65	379.85	310845164	311.65	260.10	5196859	314.50	259.90	56057627
Jun-16	488.00	440.35	23047496	487.95	440.40	242209799	323.85	287.40	12546218	324.70	287.55	74359259
Jul-16	510.10	455.25	15269479	510.10	455.15	130370299	330.00	291.25	7251939	330.65	290.95	40167587
Aug-16	537.45	479.15	19255368	537.70	478.40	158425947	343.10	313.65	6330649	343.55	313.40	39381756
Sep-16	589.35	526.20	19454513	588.70	526.00	164094093	373.35	334.70	2769462	373.50	335.10	37243000
Oct-16	565.70	521.85	9415442	565.70	522.00	108142885	365.45	338.75	2827255	366.15	338.80	30945237
Nov-16	540.20	453.05	13558398	540.20	452.35	169878493	348.45	285.55	2570880	348.55	285.30	50489432
Dec-16	472.80	433.05	10279248	473.15	432.90	181549238	304.10	287.05	1296757	304.60	287.75	29997642
Jan-17	548.00	481.25	10130929	548.90	481.10	110527645	341.75	307.95	2154717	341.85	308.40	26577374
Feb-17	541.70	436.55	14480959	542.25	436.45	163811534	343.30	268.15	3132935	344.75	268.30	52622569
Mar-17	480.40	449.10	12098007	480.95	449.45	134165992	292.40	274.60	8732943	292.85	274.55	54498816

Table 1. Dataset

3.2 Analysis of Financial Time Series Using Hidden Markov Models

A generative probabilistic model called the Hidden Markov Model (HMM) assumes that the system is transitioning between a predetermined number of states. A matrix of state transition probabilities may be used to define the state transition since it is a Markov Process. Although the state sequence cannot be seen in its entirety, some of the state dynamics may be seen.

The observations in stock markets are continuous in nature and they are part of non-stationary systems. Assume that St represent's the state on day t and that Ot is a vector with four elements: the daily close, open, high, and low. One of the presumptive states is the state of St. A typical Hidden Markov Process is seen in Figure 2.

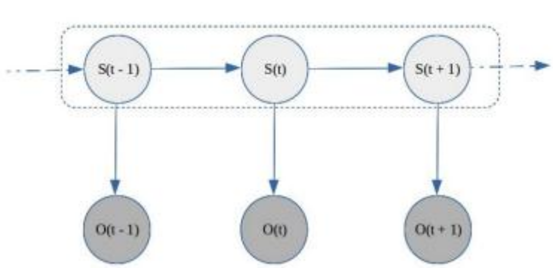

Fig.2 Hidden Markov Model

Observations may be modelled as Multivariate Gaussian distributed since the vector "Ot" accepts real values. Although the components of an observation may be connected, the observations are presumed to be independent. Since HMM is a finite state machine, the state St can only accept discrete values. Let's define a few terms used to describe hidden markov models presently.

Number of observations, T Latency,

K Number of States, N ($St = s1, s2, \ldots, sN$) Observation Sequence,

O Initial State Probability,

$P0$ State Transition Matrix, $A = [aij]$ where aij is the state transition probability from si to .

Observation Probabilities, Σi $i = 1,2, . . N$, where , Σi are the mean and covariance matrix for Gaussian distribution for state i.

The Hidden Markov Model can be represented as $\lambda = (A, \mu, \Sigma, P0)$ *eq-1*

The following inquiry must now be addressed: How the Hidden Markov Model be employed? This is resolved. by answering the three questions that follow.

- How probable is it to witness the provided sequence of data, given the model?
- What is the ideal hidden state sequence given the model and observations?
- What are the ideal model parameters, given the observations?

Utilizing the Forward method to solve the first issue. The Viterbi algorithm can resolve the second issue. Baum-Welch method is used to resolve the third issue.

3.3 Prediction of Stock Prices

The key to forecasting the stock price for the following day is to compute the log-likelihood of K prior observations and compare it to the log-likelihood of all prior subsequences of the same size by moving the window by one day in the direction of the prior data. The next step is to choose a day in the past whose log-likelihood of its K prior observations is most similar to the subsequence whose price is to be forecasted for the following day.

j = argmini (|*P*(*Ot* ,*Ot*−1,*Ot*−2, … ,*Ot*−*K* | *λ*) − *P*(*Ot*−*i* ,*Ot*−*i*−1,*Ot*−*i*−2, … ,*Ot*−*i*−*K* | *λ*)|) *eq-2*

wh ere, *i* = 1,2, … , *T* / K.

The difference in price between the indicated day and the following day is then calculated. The price for the current day is then adjusted for this change to obtain the forecast for the following day.

Ot+1 = *Ot* + (*Ot*−*j*+1 − *Ot*−*j*) *eq-3*

Then, once the correct observation is obtained it will include it into the dataset and make necessary adjustments to the model's parameters to prevent divergence. Simply put, fixing of the size of the subsequence and look for a different subsequence from earlier data that has a like pattern. The behaviour of the detected subsequence is then mapped to the subsequence being predicted.

Training a collection of models by altering the number of states (N) from the state space G in order to choose the model with the ideal number of states. Consideration of the values of the number of states in G between [2, 25]. Afterward, determined the negative log-likelihood of the training data used for each model, and selected the model with the lowest value.

However, this method favors complicated models, which suggests that the number of states picked may likely to be greater and may lead to overfitting. To get around this problem, Altration the negative log likelihood with a penalty term. Depending on the chosen penalty period, impose various levels of limits on the model and have looked into two different performance assessment metrics: the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). For a BIC model, the number of model parameters is multiplied by the logarithm of the number of observation samples used, but for an AIC model, the number of model parameters is added to the negative log-likelihood value.

□□□ = −2 log ((□□□□□□ | □)) − 2□ □□□ = −2 log ((□□□□□□ | □)) − □. log(□) □ℎ□□□, □ = □ 2 + 2□ − 1 *eq-4.*

4. IMPLEMENTATION

Before concluding that the Hidden Markov Model employed may be effectively utilised in the stock market, perform empirical assessments on the support vector regression analysis prediction method and the hidden markov model prediction method, respectively.

The performance metric that is used in this project is Mean Absolute Percentage Error (MAPE) which is defined as □□□□ = 1 □ ∑ |□□□□□□□(□) − □□□□(□)| □□□□(□) □ □=1

Identifying the effectiveness of HMMs in stock price prediction was the key goal. To train the model and determine the probability of the observations, utilized the free source Python module hmmlearn. TATA Motors Inc., Reliance Inc., and YES BANK Corp. is chosen to be the stocks. The closing price is taken into account together with the market's attributes during the last 2520 working days (about 10 years). The most recent 100 observations were set aside for testing, while the remaining observations were utilised to train the model. Starting with the 100th day and utilising its actual observation to fine-tune the model for forecasting the 99th day and so on, projected the prices for the previous 100 days. As a result, it will add one more training sample each time then fine-tune the model. Putting of the fixed model into practise by limiting the number of states to four. The stock price predictions for TATA MOTORS, Reliance, and YES Bank utilising HMM with four states are shown in Figures (3-5). To compare the outcomes, compute the MAPE and showed the forecasts and the actual prices. After that, improved the model by choosing the one with the lowest BIC value, which depends on the quantity of states.

*Axis Key - **(**X axis = Date, Y axis = Price)*

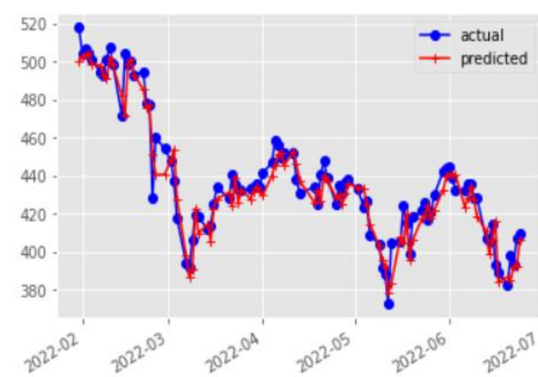

Fig. 3 TATA Motors

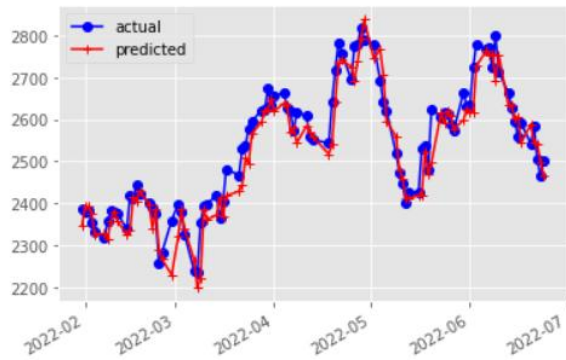

Fig. 4 RELIANCE

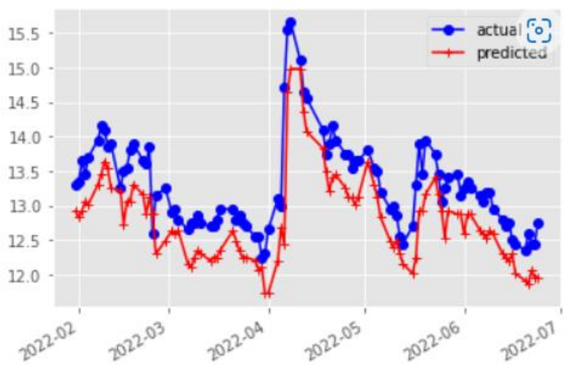

Fig. 5 YES BANK

Table of Comparison

MAPE values for TATA MOTORS Model Closing price.

Model	Closing
HMM	0.0053
SVR	0.0052

Table. 2

MAPE values for RELIENCE Model Closing price.

Model	Closing
HMM	0.0059
SVR	0.0075

Table. 3

MAPE values for YES BANK Model Closing price.

Model	Closing
HMM	0.0124
SVR	0.0098

Table. 4

Support vector regression (SVR), a non-probabilistic model, was also utilised to model the stocks. Take a look at the data set (x1, y1), (x2, y2),..., (xT, yT). In SVR, yi is a response variable as opposed to Support Vector Machines for classification, where yi is a label. Searched for a function that uses the prices of the current day to forecast the close prices of the following day using SVR. To forecast the four prices, created four models. The testing and training procedures are the same as for the HMM. Retrained the SVR to forecast for subsequent days by starting predictions with the 100th day and using its real observation. Trained the SVR using historical data up to the previous day in order to forecast the pricing for the next day.

In order to obtain the forecasts for the following day, the current day's pricing is passed. To compare the outcomes from SVR and HMM, computed the MAPE and plotted the predictions on the identical HMM plots. Figures 3-5 display the actual and forecast closing prices for the Stacks based on HMM. The MAPE values for all three equities are shown in Tables 2-4 using the HMM and SVR models.

5. RESULTS

In both the HMM and the SVR implementations, the projected values are shown for Close price closely track the patterns displayed by its corresponding real values, and the MAPE values were also found to be comparable. The SVR model's predictions were not confirmed by jarring changes in the stock price. However, it was discovered that the HMM-based model was extremely susceptible to changes in stock price. This conclusion is also in line with the scatter plot of the prediction error for each day for both models, where the distribution of the error points was seen to be more dispersed around zero for the HMM and concentrated around zero for SVR. Figure 6 displays the HMM prediction for the next four days for the stock TATA MOTORS.

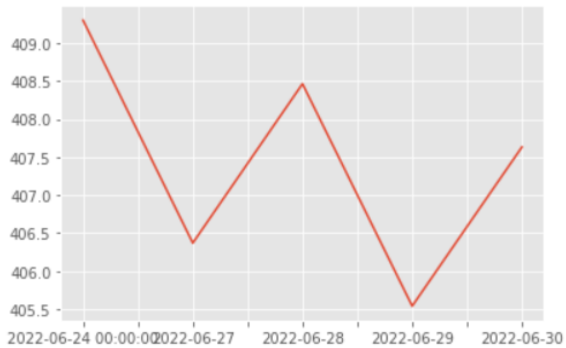

Fig. 6 Prediction plot for Tata Motors

6. CONCLUSION & FUTURE ENHANCEMENT

Hidden Markov Model seeks to minimise the impact of factor selection, combination, and transformation, in contrast to the support vector regression analysis prediction method and other forecasting techniques. To minimise unneeded mistakes, this model makes a wise decision by employing the implicit prediction and letting the stock price-influencing components be included in the hidden state model. Therefore, the stock price may be predicted using the Hidden Markov Model. Modifying the original Hidden Markov Model approach to account for the single forecast uncertainty and unpredictability. An average weighted factor is performed to complete the development of the forecast accuracy, and lower the investment risk by taking into account the impact factor prediction data. It is true that the stock market forecast approach has significant practical implications.

Even though, in general, the number of states in Hidden Markov Models would have a substantial impact on the observations, when tried to determine the best model utilising the optimum states using BIC, it did not make a significant difference. When the prices for the following day are forecast, both HMM and SVR have comparable accuracy. SVR provides more consistent forecasts, whereas HMM captures the volatility of stock values. HMM may thus be more effective for equities with significant volatility and SVR may be more effective for stocks with more stability

In conclusion, modelling and prediction of Hidden Markov Models for multivariate financial time series offers a set of useful frameworks for probabilistic analysis and is a key tool in the field of pattern recognition. The appropriate observed vector is chosen, the quantity of hidden states, the emission probability density functions, and determine the goal time series change pattern based on the Hidden Markov Model. In order to produce a probability and support the prediction findings with statistics, selection of pattern matching sequences by looking at the historical sequence.

In future, it is possible to add different parameters to the model for better prediction of result and add new features like sentimental analysis to the web service to make this system much more robust and sustainable.

REFERENCES

[1] B. L. Pooja, Suvarna Kanakaraddi ,Meenaxi .M. Raikar. (2020) “Sentiment Based Stock Market Prediction”. School of Computer Science and Engineering KLE Technological University Hubli, India.

[2] Mohammad Mahabubul Hasan, Pritom Roy, Sabbir Sarkar and Mohammad Monirujjaman Khan (2021) "Stock Market PredictionWeb Service Using Deep Learning by LSTM " Department of Electrical & Computer Engineering North South University Dhaka-1229, Bangladesh.

[3] Sidra Mehtab & Jaydip Sen (2020) "A Time Series Analysis-Based Stock Price Prediction Framework Using Machine Learning and Deep Learning Models "School of Computing and Analytics, NSHM Knowledge Campus Kolkata – 700053, India.

[4] Dr. Devpriya Soni1, Sparsh Agarwal, Tushar Agarwal, Pooshan Arora, Kopal Gupta (2021) "Optimised Prediction Model For Stock Market Trend Analysis " Department of CSE, Jaypee Institute of Information Technology, Noida, Uttar Pradesh, India.

[5] Tao Xing, Yuan Sun, Qian Wang, Guo Yu (2021) "The Analysis and Prediction of Stock Price"International School of Software Wuhan University Wuhan, China.

[6] Faisal momin, Sunny patel (2020) "Stock Market Prediction System Using Machine Learning Approach" department of computer engineering, sandip foundation, nashik, India.

[7] M Therasa S Sarveshwar M Madhusudhaanan V Naveen Kumar (2020) "Survey On Stock Market Forecast Using Machine Learning And Text Mining With Cloud’s CaaS" Department of Computer Science and Engineering, Panimalar Institute of Technology, Chennai-600123.