International Journal of Innovation Studies

COMPARATIVE ANALYSIS OF HIDDEN MARKOV MODELS AND ARIMA FOR STOCK PRICE PREDICTION: A PERFORMANCE EVALUATION

Poornima M
Department of Mathematics, D.K.M. College for Women (Autonomous), Vellore – 1, Affiliated to Thiruvalluvar University, India, poornimadkmcw@gmail.com

N. Nithyapriya
Department of Mathematics, D.K.M. College for Women (Autonomous), Vellore – 1, Affiliated to Thiruvalluvar University, India, nithyapriyamath@gmail.com

Abstract—Stock market prediction involves using analytical techniques to predict stock prices or indices based on records. This makes accuracy essential to the depositors, dealers, and monetary organizations operating near marks regarding knowledgeable buying and selling, along with effective management of the portfolios. Strong influence would be exercised in effective models to predict changing investment strategies in volatile markets to enhance risk management. Reliability in forecasts would give an advantage over the competitors as it helps stakeholders know the market's movement and optimize returns, thereby minimizing losses. This paper compares the performances of two popular models used to predict stock price: Hidden Markov Models (HMM) and Auto-Regressive Integrated Moving Average (ARIMA). Our three key metrics under consideration in this analysis are MAPE, SMAPE, and Coefficient of Determination, which is R^2. The outcomes indicate that HMM dramatically outperforms ARIMA under all evaluations. For instance, HMM attains a value of 1.51 MAPE, 1.51 SMAPE, and an R^2 value of 0.8882. This depicts great precision with a very good fit for the model. On the contrary, ARIMA shows a much greater MAPE value of 5.91 and an SMAPE of 5.71 and holds an even terrible R^2 at -0.3709 thus indicating that it is quite a bad fit. The analysis also reveals a good performance in time series for HMM concerning financial domains that include stock price predictions where they require accurate forecasts and reliable information to determine what decisions to be taken. An improved concept and robust framework on the implementation of HMM into financial forecasting research has been given for accuracy that surpasses traditional models ARIMA.

Keywords—Prediction, ARIMA, HMM, LSTM, MAE, RMSE, R-squared.

Introduction

The stock market has become an integral part of the universal family in recent years. Somewhat variations fashionable this marketplace significantly impact individual finances, commercial decision-making, and a country's overall economic stability. Despite its potential for high returns, investing in the typical market carries inherent risks payable towards the aforementioned changeable nature. Predicting typical marketplace trends remains a long-standing challenge that has been thoroughly explored through various machine-learning methodologies. Accurate prediction of future stock prices may result in enormous financial

benefits. Hence, developing intelligent models for stock market prediction is highly valuable and broadly relevant.
Modern approaches such as Neural Networks, Support Vector Machines, Hidden Markov Models, and ARIMA provide efficient techniques for the standard marketplace tendencies' forecasting. This paper briefly reviews the prominent methods used in the standard marketplace prediction.

LITERATURE REVIEW

A technique was applied close to the prediction of the next period's exiting value based on the HMM. Stocks from four different companies, which are Apple Inc., TATA Steel, Dell Inc., and IBM Corporation, were considered. The features used were Low, High, Close, and Open prices. Each stock was considered as a separate case for modeling. The classical was efficient for more than seven months and tested based on MAPE values [1].
Neural networks have numerous applications, such as predicting financial markets. This uses an Artificial Neural Network with past market data taken as input to study share price prediction. A feed-forward architecture, trained on one year of data, was able to predict market trends, though not actual values [2].
Stock market prediction is a challenging task in time series forecasting. They propose a nonlinear combination model using SVM regression, combining linear regression for extracting linear features and Neural Networks for nonlinear features. Applied to the Shanghai Stock Exchange index, the model outperforms others, demonstrating its effectiveness in financial forecasting [3]. It presents a likelihood system for predicting mid-term trends in the Taiwan Stock Exchange Weighted Stock Index (TSEWSI). The system enhances prediction by using ARIMA(1,2,1) features with a recurrent neural network trained on the second difference data. Trained on four years of weekly data, it well predicts market trends six weeks in advance [4].
This backpropagation and regression analysis work employed error in predicting untranslated and translated Nigerian Stock Market Prices (NSMP) over 720 days. The network topology employed was a 5-j-1 network topology that normalized the inputs to zero mean unit variance. Results showed that translated NSMP predictions were better than untranslated ones in terms of stability and mean relative percentage error. The translated NSMP method was more accurate and reliable [5]. Stock market volatility is difficult to predict because it is dynamic and complex. It introduces a service-oriented multi-kernel learning (MKL) structure for analyzing typical unpredictability, using multiple data sources such as ancient values, interchange volumes, and news articles. The framework includes data preparation, source-specific modeling, and cross-source correlation analysis using sub-kernels with weight adjustments. Experiments on HKEx 2001 data show that MKL outperforms single-kernel methods, achieving higher accuracy and improved volatility prediction by integrating diverse information sources [6].
Neural networks are widely used for stock market prediction, though no standard method exists for optimal model selection. It compares a feedforward MLP and an Elman recurrent network for predicting stock values based on historical data. Results indicate that MLP excels in forecasting price fluctuations, while the linear regression and Elman network better predict change directions, as standard evaluation metrics show [7]. This three-layered feedforward BP neural network uses for prediction to address the complexity of stock price systems. It describes

the network topology, preprocessing of sample data, and selection of parameters, and employs the Levenberg-Marquardt algorithm to avoid local extrema and improve convergence. Results from simulations using Shanghai Stock Exchange data for short-term stock prediction are found in [8].

This models uses of the S&P 500 Index through neural networks, namely multilayer perceptrons and probabilistic neural networks, in attempts to forecast market trends. The trading advice of the networks outperformed the index through the probabilistic neural network marginally outperforming the multilayer perceptron [9]. This Hidden Markov Models (HMM) discusses the utilization to predict stock prices across interrelated markets, with a specific focus on stocks in the airline industry. Once trained on data from the past, the HMM detects behavioral patterns and interpolates neighboring values for the generation of forecasts. Results indicate that HMM is a promising tool for stock market prediction, and it provides a new approach in this research domain [10].

The daily and weekly prediction of S&P 500 returns using historical price data with auto-regressive (AR) and neural network (NN) models. It introduces a novel three-model system, where each model predicts one of three market trends (bear, choppy, bull), guided by a trend classification algorithm and decision rules. Results show that nonlinear models outperform linear ones, and the three-model NN system achieved double the return of a buy-and-hold strategy for weekly predictions [11].

Methodologies

Description of Data

For this stock price forecasting analysis, the HDFC Bank stock price dataset from July 20, 2023, to July 19, 2024, was downloaded from Yahoo Finance. The data is managed in a Pandas DataFrame for efficient management, easy manipulation, and smooth analysis. This dataset was accessed through publicly available sources, including the Yahoo Finance repository [12].

ARIMA

It examines the top 4 NSE Nifty Midcap50 companies using historical stock data and the ARIMA model with AICBIC to predict market trends and aid investment decisions [13]. The model discusses the key components of time series data and implements the ARIMA model to predict future stock values using NIFTY daily data of the Nifty50 index [14]. It is an ARIMA-based stock price predictive model that has been applied and displayed its potential as a very excellent short term predictor compared to the other methods with published data from the NYSE and NSE [15].

HMM

This Hidden Markov Model (HMM) uses to analyze stock market trends to determine the sequences of hidden states under optimal conditions through a one-day difference in close values and probability distributions at steady-state levels for predictors of future trends [16]. The daily stock prices of Apple, Google, and Facebook have been predicted in this paper using the Hidden Markov Model, by choosing an optimal number of states using AIC and BIC. HMM outperformed the naïve method in forecasting and gave better returns for active traders [17].

EXPERIMENTAL PARAMETER SETTINGS

This one existed to establish the capability of the RNN prediction perfect in comparison against three other deep learning models: AUTOENCODER, CNN-LSTM, and LSTM.

Dataset Collection

For the analysis of this paper, for stock price forecasting, the data set of HDFC Bank was taken from the Yahoo Finance website between July 20, 2023, and July 19, 2024. The data is set up in a Pandas DataFrame that enables this manageable, easily manipulated, and analyzed data set.

Data Pre-processing

High-quality pipeline for preprocessing HDFC Bank's stock price data. Initially, the dataset is loaded from a CSV file. The 'Date' column is converted to DateTime format for easier time-series analysis. The 'Close' prices are extracted as the final trading price of the stock for every day and reshaped for modeling. These 'Close' prices are normalized using z-score normalization to make it uniform for enhancing the performance of the model; subtraction by the mean and division by standard deviation scales the data, making it independent of original units and reducing the influence of outliers. In addition, processed dates are saved for visualization purposes at later stages to ensure that the analysis is on temporal trends. This preprocessing leads the way to the Hidden Markov Model (HMM) and further predictive analysis.

Data Normalization (Z-Score Scaling)

Close prices scaled =(close prices-μ)/σ

- μ: Mean of the closing prices.
- σ: Standard deviation of the closing prices.
- This formula standardizes the data to have a mean of 0 and a standard deviation of 1.

ARIMA

The code uses the ARIMA (AutoRegressive Integrated Moving Average) model to analyze and predict the closing stock prices of HDFC Bank. After loading the dataset and converting the 'Date' column into datetime format, the 'Close' prices are extracted for modeling. Specified (p, d, q) parameters, in this case (5, 1, 0), are applied to configure the ARIMA model so that it captures the trends and patterns of a time series. The model is then fitted to available historical data with generated predictions over the same period. Metrics of performance include Mean Absolute Percentage Error (MAPE), Symmetric Mean Absolute Percentage Error (SMAPE), and R-squared (R^2). This means Figure 1 is the plot of the simulated closing prices forecasted against actual closing prices that reflects the predictive performance of the ARIMA model. ARIMA stands for Auto-Regressive Integrated Moving Average. It consists of three parameters: p, d, and q:

- p: Number of Auto-Regressive (AR) model lag observations.
- d: Degree of differencing to make the data stationary.
- q: Size of the Moving Average (MA) window.

ARIMA Components:

1. **Auto-Regressive (AR)**:

$$X_t=\phi_1 X_{t-1}+\phi_2 X_{t-2}+\cdots+\phi_p X_{t-p}+\epsilon_t$$

ϕ_i: Coefficients of lag terms.

ϵ_t: White noise (error).

2. **Differencing (Integrated)**: To make the series stationary:

$$X_t'=X_t-X_{t-1}$$

For d-order differencing:

$$X_t^{(d)} = X_t^{(d-1)} - X_{t-1}^{(d-1)}$$

3. **Moving Average (MA):**

$$X_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots\ldots\ldots + \theta_q \epsilon_{t-q}$$

θ_i: Coefficients of error terms.

The combination of these gives the ARIMA model:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + X_t + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots\ldots\ldots + \theta_q \epsilon_{t-q}$$

HMM

The method used a time-series analysis using a Hidden Markov Model (HMM) to forecast HDFC Bank's closing stock prices. The dataset is preprocessed by converting dates into a DateTime format and isolating the 'Close' price column, which is normalized to improve the model's stability and performance. An HMM with three hidden states is defined and trained on the normalized data to capture the underlying patterns in stock price movements. The model predicts the hidden states, which are used to estimate the closing prices. These predictions are then inverse-transformed to their original scale for accuracy comparison. Performance metrics such as Mean Absolute Percentage Error (MAPE), Symmetric Mean Absolute Percentage Error (SMAPE), and R-squared (R^2) are computed to evaluate the model. Finally, figure 2 shows the predicted prices are plotted against the actual prices, providing a clear visualization of the model's forecasting accuracy.

Inverse Scaling (De-normalization)

Predicted prices rescaled=predicted prices·σ+μ

Rescales the predicted values back to the original scale using the previously computed mean (μ) and standard deviation (σ).

Gaussian Hidden Markov Model

The HMM is a probabilistic model where:

$$P(O_t | S_t = i) = \mathcal{N}(\mu_i, \Sigma_i)$$

Observations Ot are modeled using Gaussian distributions

S_t Hidden state at time t

N(μ_i,Σ_i): Gaussian distribution with mean μi and covariance Σi for each state.

Parameters learned during model fitting:

Transition probabilities:

$$P(S_{t+1} = j | S_t = i)$$

Means and variances of Gaussian emissions.

Gaussian Emission Prediction

Hidden states are predicted using:

Hidden states=argmaxj P(Ot| St=j)

For each time step t, the model assigns the state j with the highest likelihood based on the observed data.

Predicted Prices Based on Hidden States

The predicted price for each observation is the mean of the Gaussian distribution associated with its hidden state:

Predicted prices[t]=μ. Hidden state[t]

Results and Discussion

We can infer the following conclusions based on the insights from Table 1 and Figures 1-3.

The results of the Hidden Markov Model (HMM) and ARIMA (AutoRegressive Integrated Moving Average) models for predicting HDFC Bank's daily closing stock prices are summarized in the table below. The performance metrics used for evaluation are Mean Absolute Percentage Error (MAPE), Symmetric Mean Absolute Percentage Error (SMAPE), and R-squared (R^2). These metrics provide a comprehensive assessment of the models' prediction accuracy and goodness of fit.

HMM. The MAPE for the HMM model is 1.51%, which indicates that, on average, the model's predictions deviate from the actual stock prices by 1.51%. This is a relatively low error, demonstrating that the Hidden Markov Model is highly accurate in predicting the daily closing prices of HDFC Bank's stock.

ARIMA. In comparison, the ARIMA model shows a higher MAPE of 5.91%, suggesting that the ARIMA model's predictions are, on average, 5.91% off from the actual closing prices. This higher error reflects the reduced accuracy of ARIMA in capturing the underlying patterns in the stock price data for this particular dataset.

Symmetric Mean Absolute Percentage Error (SMAPE)

HMM: The SMAPE for the HMM model is 1.51%, which aligns with the MAPE result, as both metrics are quite close. The SMAPE value confirms that the model's error is relatively small and evenly distributed across all prediction points, ensuring a symmetric evaluation of errors.

ARIMA: The ARIMA model’s SMAPE is 5.71%, again demonstrating a larger prediction error compared to HMM. The higher SMAPE value indicates that ARIMA has relatively higher deviations when predicting stock prices, making it less reliable for this task.

R-squared (R^2):

HMM: The R^2 value of 0.8882 for the HMM model indicates that approximately 88.82% of the variance in the observed stock prices can be explained by the model. This high R^2 value suggests that the HMM is effective in capturing the underlying patterns and dynamics of the stock market, making it a strong predictor of future prices.

ARIMA: On the other hand, ARIMA yields an R^2 value of -0.3709, which is negative. A negative R^2 value indicates that the model performs worse than a simple mean-based model, meaning that ARIMA does not adequately fit the data and its predictions are unreliable.

HMM’s Superiority: The Hidden Markov Model significantly outperforms ARIMA in predicting stock prices, as evidenced by its lower MAPE, SMAPE, and higher R^2 values. The model's ability to identify underlying market states and transitions in the data contributes to its strong performance.

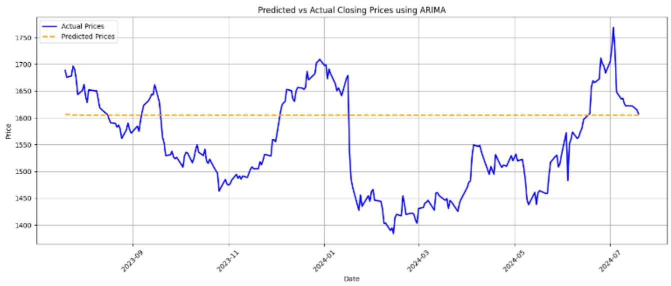
ARIMA's Limitations: While ARIMA is a commonly used time series forecasting model, it struggles to effectively capture the complexities of the stock market dynamics in this case. The negative R^2 value indicates poor model fit, and the higher error metrics suggest that ARIMA fails to accurately predict HDFC Bank stock's closing prices.

Market Modeling with HMM: The superior performance of the HMM model can be attributed to its ability to model stock price fluctuations as a sequence of hidden states, with the observed prices reflecting transitions between these states. This modeling approach allows HMM to better capture the inherent volatility and patterns in the stock data compared to the linear assumptions made by ARIMA.

Overall, the HMM is the more reliable and accurate model for stock price prediction in this study, providing valuable insights for financial market forecasting.

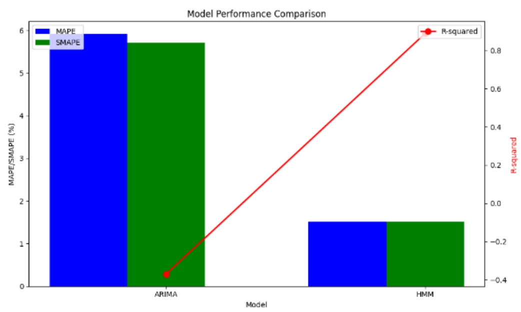
TABLE I. COMPARATIVE PERFORMANCE ANALYSIS OF TWO MODELS ON THE HDFCBANK DATASET

MODEL	MAPE	SMAPE	R-squared
ARIMA	5.91	5.71	-0.3709
HMM	**1.51**	**1.51**	**0.8882**



Historical vs. Predicted Stock Prices on Test Data using ARIMA

Historical vs. Predicted Stock Prices on Test Data using HMM.



Model Performance Comparison: MAPE, SMAPE, and R-squared

CONCLUSION

This study demonstrates the application of AutoRegressive Integrated Moving Average (ARIMA) and Hidden Markov Models (HMM) models for forecasting HDFC Bank's daily stock prices. The results indicate that the HMM significantly outperforms ARIMA in terms of prediction accuracy and model fit. With a MAPE of 1.51%, SMAPE of 1.51%, and an R^2 of

0.8882, HMM offers a robust framework for capturing the underlying patterns and transitions in stock price data. In contrast, ARIMA, with a MAPE of 5.91%, SMAPE of 5.71%, and an R^2 of -0.3709, performs poorly, indicating its limitations in modeling complex stock market behaviors. These findings underscore the effectiveness of Hidden Markov Models in predicting stock prices, highlighting their potential for financial market analysis.

For future work, could explore incorporating additional features such as trading volume and technical indicators to enhance prediction accuracy. Hybrid models combining HMM with deep learning techniques like LSTM or GRU may provide better performance. Real-time predictions and backtesting strategies could be developed for live trading environments. Further exploration of market regimes and volatility models could improve forecasting under varying market conditions. Lastly, integrating stock price predictions into portfolio optimization and risk management frameworks would offer a more comprehensive approach to financial decision-making.

References

Gupta, A., & Dhingra, B. (2012, March). Stock market prediction using hidden markov models. In 2012 Students Conference on Engineering and Systems (pp. 1-4). IEEE.

Abhishek, K., Khairwa, A., Pratap, T., & Prakash, S. (2012, July). A stock market prediction model using Artificial Neural Network. In 2012 Third International Conference on Computing, Communication and Networking Technologies (ICCCNT'12) (pp. 1-5). IEEE.

Luo, F., Wu, J., & Yan, K. (2010, July). A novel nonlinear combination model based on support vector machine for stock market prediction. In 2010 8th World Congress on Intelligent Control and Automation (pp. 5048-5053). IEEE.

Wang, J. H., & Leu, J. Y. (1996, June). Stock market trend prediction using ARIMA-based neural networks. In Proceedings of International Conference on Neural Networks (ICNN'96) (Vol. 4, pp. 2160-2165). IEEE.

Akinwale Adio, T., Arogundade, O. T., & Adekoya Adebayo, F. (2009). Translated Nigeria stock market price using artificial neural network for effective prediction. Journal of theoretical and Applied Information technology, 1(1), 36-43.

Wang, F., Liu, L., & Dou, C. (2012, June). Stock market volatility prediction: a service-oriented multi-kernel learning approach. In 2012 IEEE ninth international conference on services computing (pp. 49-56). IEEE.

Naeini, M. P., Taremian, H., & Hashemi, H. B. (2010, October). Stock market value prediction using neural networks. In 2010 international conference on computer information systems and industrial management applications (CISIM) (pp. 132-136). IEEE.

Li, F., & Liu, C. (2009, August). Application study of BP neural network on stock market prediction. In 2009 Ninth International Conference on Hybrid Intelligent Systems (Vol. 3, pp. 174-178). IEEE.

Schierholt, K., & Dagli, C. H. (1996, March). Stock market prediction using different neural network classification architectures. In IEEE/IAFE 1996 Conference on Computational Intelligence for Financial Engineering (CIFEr) (pp. 72-78). IEEE.

Hassan, M. R., & Nath, B. (2005, September). Stock market forecasting using hidden Markov model: a new approach. In 5th international conference on intelligent systems design and applications (ISDA'05) (pp. 192-196). IEEE.

Xiang, C., & Fu, W. M. (2006, December). Predicting the stock market using multiple models. In 2006 9th International Conference on Control, Automation, Robotics and Vision (pp. 1-6). IEEE.
https://finance.yahoo.com
Devi, B. U., Sundar, D., & Alli, P. (2013). An effective time series analysis for stock trend prediction using ARIMA model for nifty midcap-50. International Journal of Data Mining & Knowledge Management Process, 3(1), 65.
Dhyani, B., Kumar, M., Verma, P., & Jain, A. (2020). Stock market forecasting technique using arima model. International Journal of Recent Technology and Engineering, 8(6), 2694-2697.
Ariyo, A. A., Adewumi, A. O., & Ayo, C. K. (2014, March). Stock price prediction using the ARIMA model. In 2014 UKSim-AMSS 16th international conference on computer modelling and simulation (pp. 106-112). IEEE.
Kavitha, G., Udhayakumar, A., & Nagarajan, D. (2013). Stock market trend analysis using hidden markov models. arXiv preprint arXiv:1311.4771.
Nguyen, N. (2017). An analysis and implementation of the hidden Markov model to technology stock prediction. Risks, 5(4), 62.