PREDICTING STOCK BEHAVIOUR USING VAR AND NNAR MODELS: ACOMPARATIVE ANALYSIS

Srinivas Gumparthi [1] D. Venkata Vara Prasad 2
Bhargavi Rentachintala[3] Attuluri Thandava Krishna[4] A Mohamed Ajwad [5]

[1] Professor, Sri Sivasubramaniya NadarCollege of Engineering (SSN School of Management)

[2] Professor, Computer Science and Engineering, Sri Sivasubramaniya NadarCollege of Engineering

[3] Profeessor, SCOPE, VIT University, Chennai

[4,5] Computer Science and Engineering, Sri Sivasubramaniya Nadar Collegeof Engineering

1 Abstract

In the modern day predicting stock market behavior has become important because data-driven decisions help traders, analysts, and policymakers a lot with their day-to-day life which impact their portfolios. This research aims to use Vector Auto-regressive (VAR) and Neural Network Auto-regressive (NNAR) models to analyze stock market benchmarks that affect the sentiment of traders and the market. By examining data and interrelated financial variables the project aims to provide forecasts that reveal trends and risks involved in investing in a particular stock. Utilization of VAR and NNAR models will help our understanding of the future ahead of us by providing insights for decision-making. This study not only contributes to the field of analysis but also has practical applications concerning investment strategies, risk management, and policy development. The VAR model is a multivariate time-series model that captures the relationships by relating current observations of a variable with the past observations of itself and previous values of other variables in the system over the lagged values. The NNAR model is a feed-forward neural network with lagged inputs and a single hidden layer with non-linear functions it is said to be one of the best tool for analyzing time-series data.

2 Introduction

The world of finance is a place that sees constant flux, with many changes happening at a time which determine the future of a particular company. The markets by design are shaped by complicated and interconnected forces that are influenced by economic policies, geopolitical events, and trader sentiment. Understanding the behavior is not only a fundamental aspect of financial management but also to be able to predict the future which determines an investor's course of actions. This pursuit has attracted the imagination of countless traders, analysts, and researchers for many years. This quest for precise and comprehensive tools which help us in forecasting has increased exponentially over the years where data-driven decisions are required. This is the challenge we try to solve in this project, we aim to provide reliable tools for predicting stock market trends which unravel the complexities in the stock market movements.

The stock market is often defined by its complicated nature and being volatile which makes it hard to predict market trends. However, the traditional methods are primitive in finding the inter-dependencies among various financial variables, making it one of the challenges to develop a holistic understanding.of market dynamics. This is where VAR and NNAR models take precedence over the traditional methods of prediction, VAR and NNAR offer a multivariate time-series approach that can determine the relationship between variables allowing us to give a more comprehensive analysis on how economic factors impact the financial markets.

Buying a stock has become one of the frequent way for an insightful investor in the current year, be it the Mutual fund or securities or buying a stock directly. The volatile nature of the stock market offers a platform to generate high returns whether it is a short-term or a long term investment.

The underlying statement is that traders in the current time need a sophisticated tool that makes predictions faster with small data provided and which can be relied upon to take calculated risks.

There is as much as downside as there is an upside, that means the price can rise or drop at any point given. At times geopolitical events that concern one country can determine the events happening in the stock market of another country. For an instance Russia-Ukraine war has affected the stock market in India for a week at least. Another instance is the Taliban takeover of Afghanistan. One might argue that there can be better chances in the long time investment whereas the fact is that there is an equal amount of risk involved in the long-term investment as it is in the short-term investment.

In this project, we concentrate on the benchmark VWAP(Volume-weighted average price), it is a combined indicator for representing average price of a security based on price and volume. It helps the traders in determining the entry and exit points of a particular stock. VWAP is widely used by traders and analysts to make informed decisions because it provides better and more accurate representation of a company’s true average price over a stipulated time. VWAP can also be used to determine if a company is undervalued or overvalued when we compare it with the current price. The direction of VWAP indicates the trends in the market, mastering the interpretation of VWAP can be very useful in enhancing trading strategies.

The foundation of VAR and NNAR models is that it can estimate how financial variables behave over time give their volatile nature. By doing so, VAR and NNAR models enable us to understand the hidden patterns that might be observed while analyzing these variables. This project uses the above features of these models to bring forth a more surrounding approach to predict stock behavior. By taking the variable VWAP into account we aim to create a forecasting system that can offer valuable insights into market movements, ultimately helping traders and other decision-makers with a more robust and reliable tool.

- In the dynamic and volatile world of stock markets, understanding the

  stock behaviour has always been crucial to the traders.

- Predicting the stock behaviour accurately has always been and will be a challenging task due to the complicated relationship between various market indicators like benchmarks, economic indicators and geopolitical events.

- We aim to leverage the ability of Vector Auto-regressive (VAR) , Neural Network Auto-regression (NNAR) models in this project to explore the complicated nature of stock market benchmarks that that influence the stock prices.

- The goal is to provide traders and financial analysts with a tool that makes reliable predictions which will help them make informed decisions.

The underlying reason behind the problem formulation is that in the modern world financial markets are getting complicated and fast-paced. Nowadays stock trading has become accessible to everyone which seems to have big changes in stock trading with new platforms coming into the market daily like computerized trading. We can clearly say that the number of people who invest in the stock market is drastically increasing which needs an algorithmic technique to provide the users a better understanding of the present scenario of stock markets.

In today's world traders face one common problem on a daily basis, which is to what to do with the money they have. The traders and other participants actively look to make informed decisions rather than depending on sheer luck or chance. The ability to predict stock behavior is not just something nice to have it is also something that helps everyone to make smart choices, know the risks involved and achieve financial goals.

The challenge we face is twofold. First, we need to empower the effective capabilities of the models in capturing the complex patterns among the stock market benchmarks. Second, it is imperative to develop a reliable forecasting tool that adapts to the volatile nature of the stock market.

Hence, we aim to address the above-mentioned challenges by employing Vector Auto-regression(VAR) and Neural Network Auto-regression (NNAR) modeling, which can study the interactions on multiple financial time series. By using VAR and NNAR, we aim to provide stakeholders with a robust tool for forecasting stock behavior and understanding the complex patterns in the time series.

3 iterature survey

Many researchers have approached this problem of predicting stock market with different approaches of both Statistical and Machine learning methods on time-series data, some of them have even tried combining two models to give a new model. All these researchers are unique in their own way, we cannot classify any results as significant results keeping the dynamic nature of our dataset in mind.

The research by Sadia et al[1] 2022 aims to draw a performance evaluation of three models TBATS, NNAR, and ARIMA on similar datasets which is the dataset of Blockchain-based Cryptocurrencies. The research extends from statistics to quantitative finance by using machine learning. Sadia et al chose RMSE and MAPE as the metrics to evaluate the models. In conclusion, they said the NNAR and TBATS models are better at predictions on time series data than the other models.

Avramov et al [2] 2019 explored the profitability of investments based on deep learning signals in the context of difficult-to-arbitrage stocks and high limits-to-arbitrage market states. They explored two different approaches, the first is a fee-forward neural network with three hidden layers and the second approach combines four neural networks including recurrent neural network and feed-forward neural network with LSTM cells and forming a generative adversarial network (GAN). The GAN approach incorporates an adversarial process where one player aims to choose the best-performing model while the other player selects conditions leading to the worst performance, aligning with the findings of previous studies by Hansen and Richard (1987). They aimed to explore profitability, downside risk, and the ability to use the considered methods to identify mispriced stocks rather than relying on accuracy as a performance metric. Avramov et al stated the economic significance of machine learning algorithms in predicting stock returns, they showcased that these methods can generate economically interpretable trading strategies given prior knowledge.

Manish Kumar et al [3] 2019 focused on comparing the effectiveness of Support vector machines and random forest algorithms. They followed a comparative approach to assess SVM and Random forest algorithms in predicting stock index movements. They aimed to evaluate their models based on accuracy, reliability, and predictive power in anticipating stock market trends. They suggested that the machine learning methods they followed could help reduce the downside risk and provide a good boundary during a market crisis.

4

Toharudin et al[4] 2021 discussed the impact of larger-scale social restrictions on COVID-19 cases in Jakarta and West Java, Indonesia. They used neural network methods to predict future movement of COVID-19 cases. They followed NNAR, MLP, and ELM approaches to assess the historical data of confirmed, recovered, and death cases of COVID-19. These models were used to forecast the movements of COVID-19 cases by using different metrics like Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Scaled Error (MASE). Their study found that the government's measures including Large Scale Social Restrictions (PSBB) and the transition to the "new normal," had a significant impact on the number of COVID-19 cases in both regions. They also concluded that after the introduction of the new normal phase, both the regions saw a significant increase in the number of cases. This rise was said to be caused by people not following the health protocols and underestimating the risk of COVID-19 transmission.

Sandesh Kancherla et al[5] 2020 explored the application of machine learning techniques specifically SVM to forecast and predict stock market trends. The study emphasized the importance of feature selection in improving the predictive capabilities of Machine learning models when analyzing stock market data, by studying relevant features the study aimed to enhance the accuracy of forecasts. The performance metrics followed in this study were Value Volatility which measures the instability in share prices over the previous n periods, calculating the daily price differences in shares, performance metric is Index Volatility this metric focuses on the volatility of the index, measuring the daily percentage changes in the index over the previous n periods. The paper discusses the role of feature selection in enhancing the predictive capabilities of machine learning models, another objective was to mitigate the risks associated with investing in stocks by developing a robust prediction model. The application of a supervised learning model particularly in SVM helped in optimizing the investment strategies.

J S Vaiz et al[6] 2017 used the capabilities of SVM and ANN to predict stock market trends, they made a hybrid model by taking advantage of both models. SVM offers feature selection by which they have cleared the noise and irrelevant features from the dataset and ANN was trained on that modified data. The performance metrics followed were Accuracy, Precision, and Specificity. J S Vaiz et al showed that ANN has predicted better over the hybrid model rather than ANN predicting over data alone. They have achieved an accuracy of 83 to 90 percent with their hybrid model.

Aydin et al[7] 2015 did a comparative analysis of Neural Networks and Machine learning models. The analysis includes the utilization of Artificial neural networks(ANN) and Vector Auto regression(VAR) models on different variables that are volatile and similar to the stock exchange data. The variables that were considered are the USD/TRY exchange rate, gold prices, and the BIST 100 index. Aydin et al found that the ANN approach has given superior prediction capabilities when compared to VAR. They further discussed the VAR model being less capable with low data and finding it difficult to make predictions over time series data than ANN can.

Almarashi et al[8] 2024 have studied and tried to enhance the accuracy of predicting player standings in future matches by using time as a factor in the analysis specifically by comparing the non-linear NNAR with ARIMA and TBATS. The study evaluates the findings by calculating Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). They were able to exhibit 95 percent confidence intervals.

Sobreiro et al[9] 2018 reviewed the various studies that have utilized machine learning algorithms, particularly SVM, and also used k-nearest Neighbours (KNN), and decision trees to forecast stock market trends and prices. Moreover, the study has introduced SVR as a regression model for predicting stock prices rather than classifying observations into groups. They followed Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) for evaluating their models.

Benrhmach et al[10] 2020 discussed the application of a Nonlinear Autoregressive Neural network and used the Extended Kalman Filter for predicting financial time series. They used mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination as their performance evaluation metrics. The study demonstrated the superiority of their proposed NAR-EKF model over existing prediction models.

Mohamad As'ad et al[11] 2020 aimed to address the uncertainty in the gold price to help investors seeking profit from their investments. They have employed the NNAR method to make predictions on the historical data available. NNAR(25,13) is the model that Mohamad As'ad et al found most accurate for their study, it involves one hidden layer in the Neural network structure. They have used mean absolute percentage error (MAPE), mean absolute square error (MASE), and root mean square error (RMSE) as their performance evaluation metrics. The researchers found that NNAR

6relatively offers a promising approach and can be implemented using open-source software.

Murthy et al[12] 2023 focused on predicting BSE Sensex, which is the stock market index of India using Machine learning techniques. They aimed to forecast the BSE Sensex accurately to help the analysts make informed decisions. The study also focused on the preprocessing of historical data, feature selection was applied to prepare it as an input for the Machine Learning models. They have employed the NNAR model to generate accurate forecasts. With the help of NNAR, they tried to capture the sequential dependencies and patterns. The research highlights that the daily price of gold is significantly influenced by the gold price from the previous day up to 24 periods ago when using the NNAR (25,13) model.

Saadon et al[13] 2024 studied to determine the behavior of Nonlinear multi-dependent variables with the help of the Neural Network AutoRegressive exogenous (NNARX) model. The study was conducted on the nonlinear behavior of riverbank erosion rates of the Sg. Bernam River in Malaysia. The model had 5 independent variables and 10 hidden layers and was stated to be the most accurate predictor for this dataset. They have considered Mean Square Error (MSE), Root Mean Square Error (RMSE), and Discrepancy Ratio (DR) to evaluate the model.

Shobande et al[14] 2021 studied the relationship between sustainable finance and the natural resource curse in Nigeria and Ghana with the help of Bayer and Hanch's cointegration test on VAR and Vector Error correction Granger causality tests. In their approach, they tested stationarity, cointegration, and Granger causality, which are statistical tests that helped to find out the relationship between sustainable finance and natural resource management. The results have significant implications for the efficient and successful natural resources management of the two countries.

Manasseh et al[15] 2018 explored the causal relationship between stock market development, financial sector reforms, and economic growth in Nigeria. They used the Vector Autoregression and Error Correction model to examine the bidirectional relationship. They have found a unidirectional causality from financial sector reform to economic growth. They followed many metrics for their model evaluation like Likelihood ratio, 5% critical ratio, and 1% critical ratio. The results suggested that stock market development and economic growth is positively co-integrated indicating a long-run equilibrium relationship.

7Hetwarz and Wang[16] 2024 introduced a novel panel approach to structural vector autoregressive analysis by using statistical identification. They have performed experiments by simulating the robustness of the proposed model under correlation and heterogeneity. The paper stated the challenges in the implementation of structural vector autoregressive models and highlighted the importance of statistical identification. They have discussed the effectiveness of their model over the traditional autoregression. They have also identified a common rotation of orthogonalized modal residuals.

Abidi and Touhami[17] 2024 investigated the safe haven property of bitcoin, crude oil, and precious metals during crucial periods like COVID-19 and the Russia-Ukraine conflict. Safe haven is a type of investment that is expected to retain or increase in value during times of market turbulence. The study employs the DCC-GARCH model to analyze the volatile correlation between oil returns and metals. Their research examined the spillover effects of volatility between the oil market and Bitcoin by using the Bayesian model namely the time-varying coefficient vector autoregression (TVC-VAR).

Ho Lee and Ho Kang[18] 2024 explored the interactions between the yield curve and macroeconomic variables within a modeling framework. Their study contributed to the field of finance by offering a detailed examination of the term structure and bond yields and their connection with macroeconomic variables. The study addressed the challenges of estimating from a high-dimensional model by utilizing Bayesian methods which resulted in efficient estimation. To assess the estimation process rolling average and standard deviation are used as evaluation metrics. The study introduced a scenario analysis that allowed an evaluation of linear combinations of observable variables in the future.

Almarashi et al focussed[19] 2024 on modeling the Gross Domestic Product(GDP) of Saudi Arabia using both linear and non-linear time series models, by using the NNAR model. They aimed to provide efficient and accurate forecasting of the GDP annual growth rate. The findings have shown the lowest values of Mean Absolute error(MAE) and Root Mean Squared error(RMSE), with that they have concluded that NNAR(5,3) is the optimal model over the other models. The model predicted the growth to be 1.3% which is closely aligned with the projection of the International Monetary Fund benchmark of the year 2023.

P.V. Thayyib et al[20] 2023 focused on the forecasting of Indian Goods and 8Service tax(GST) revenue using various time series models such as TBATS, ETS, ANN, ARIMA, NNAR, and also a hybrid time series model. All these models were trained with the historical data and a comparative analysis was drawn to identify the most accurate and effective model for forecasting monthly GST revenue. The hybrid models include ARIMA-NNAR, ARIMA-ANN, Theta-NNAR, and Theta-TBATS. They tried minimizing Mean Absolute error(MAE) and Root Mean Squared error(RMSE) which are used to evaluate the models. They have concluded that the hybrid mode Theta-TBATS has performed better than the other models in making accurate predictions.

4 Research Gaps

- Model Comparison and Evaluation: Many studies have compared different predictive models such as TBATS, NNAR, ARIMA, VAR, and ANN. Some of these studies have used metrics like RMSE and MAPE while some of the other studies have used accuracy, reliability, and economic significance. Standardization of the evaluation metrics would facilitate a transparent understanding of the users. Developing one such metric can be a challenging task because different researchers and different readers have different tastes in looking at evaluations. A technical expert may be looking for a minimum RMSE whereas a general investor might look for a better accuracy.

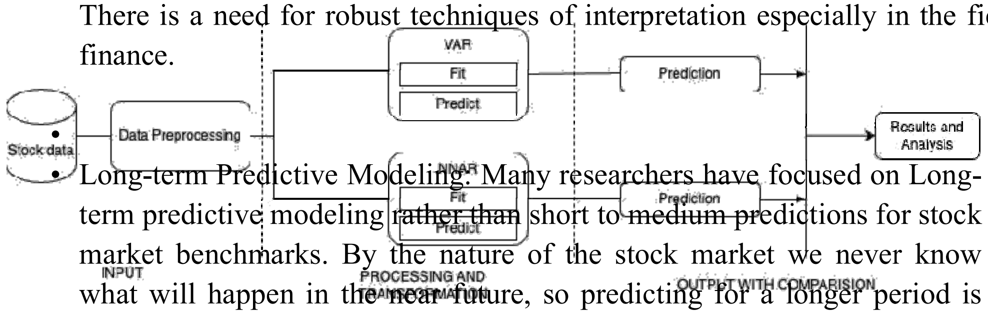
- Feature Selection and Model Performance: Several studies have stated the importance of feature selection in improving the predictive capabilities of Machine learning models. However, there is a need to study those specific features that can be relevant for most predictive tasks. Because after all, it is the characteristic of the stock market to be volatile and show abnormalities in the data. There is a need to capture and understand these abnormalities.

- Hybrid Models: Many researchers have used hybrid models that have combined the capabilities of statistical and machine learning approaches. While these hybrid models have tended to show improved performance, there is limited research that clarifies the optimal combinations and best practices for integrating.

- Data for training: While taking training data many researchers have taken historical data for more than 5 years and some have considered it for more than 20 years which may take time for the Machine learning algorithms to read and capture the patterns. Moreover especially in stock market data, and historical data for more than 8 previous quarters the predictions may be misleading in a few cases. That is because there will be a change in dividend to price ratio over time which affects the volume of the shares purchased, these metrics are not reflected in the price of the stock.

- Application of Specific Contexts: Some studies have focussed on the kind of applications such as predicting stock market trends, forecasting economic indicators like GDP and GST revenues, and the impact of external factors like COVID-19 on economic variables. However, there is always a need for research on the applications of predictive models to different contexts like sanitation, weather, and many such places.

- Interpretation and Transparency: While Machine learning models often offer high accuracy in predictions few of them lack interpretation which makes it tough for the stakeholders to understand the factors driving the predictions. There is a need for robust techniques of interpretation especially in the field of finance.



- Long-term Predictive Modeling: Many researchers have focused on Long-term predictive modeling rather than short to medium predictions for stock market benchmarks. By the nature of the stock market we never know what will happen in the near future, so predicting for a longer period is risking the accuracy of the prediction. Moreover, most traders look at a picture of 2 to 3 months to determine their entry and exit at a stock.

- Dynamic Model Adaption: Existing predictive models often assume stationary data distributions and have fixed the model parameters over time. However, the financial systems are inherently dynamic and subject to changes, and external shocks. There is a need for research on adaptive predictive modeling techniques that can adjust the model parameters dynamically and structures in response to market conditions.

- Expectations from the results: Predicting the right value at one particular time should not be the goal of a predictive model. When it comes to the prediction of stock market benchmarks the model should be able to deliver the Strength, Momentum, and Direction of the stock.

5 Proposed Architecture

This system aims to utilize Vector Auto-regression (VAR) and Neural Network Auto-regression (NNAR) models for the time series analysis to forecast and explain stock market benchmark VWAP's behavior. The mission is to offer valuable insights to traders into dynamic market trends and empower them with clear actionable information. The wishes to encompass comprehensive data processing using VAR and NNAR with transparent data visualization. In the present day understanding the market is the ultimate goal, this system aims to shed light on intricate market patterns and aid data-driven decision-making.

Figure 1: Proposed Architecture

5.1 Data Collection and Preprocessing

The foundation of the system begins with the collection of historical data for selected stock market benchmarks. These benchmarks may include indices, opening and closing prices, and other financial indicators. In this case, the benchmark is Volume-weighted Average Price (VWAP). It is calculated by the summation of the rupees traded for every transaction (price multiplied by the number of shares traded at an instance) and then divided by the total shares traded for the day. It is a good indicator because it incorporates both the price of the stock and the volume traded into consideration by providing a comprehensive view of the market activity.

$$VWAP = \frac{Price \times Volume of Price}{TotalVolume}$$

For the analysis, the training data is considered for 248 days (1 financial year) and the testing data is considered for 60 days. Initially, the dataset contains (in CSV format) many columns and it is then narrowed down to Date and VWAP.

5.2 Processing and Transformation

After completing the essential steps of data collection and preprocessing, the dataset is meticulously divided into training and testing subsets. Subsequently, both VAR and NNAR models undergo individual training processes using the training data. For VAR models, the training phase involves estimating the optimal lag order to effectively capture the intricate relationships between the variables, followed by testing procedure. Similarly, NNAR model is meticulously fine-tuned by adjusting the weights and biases of the neural network to optimise predictive accuracy before commencing training on the dataset.

5.3 Output with Comparison

After both the VAR and NNAR models are trained on their dataset, testing is done on them separately. After testing is done, results are visualised using plots separately for both VAR and NNAR models. The performance of the models are assessed using the metrics Mean Absolute Error(MAE), Root Mean Squared Error(RMSE), Mean Absolute Percentage Error(MAPE). Using the results of these metrics, both the models are analyzed on each stocks.

6 Module Split

6.1 Vector Auto-regression (VAR) Model Implementation

The heart of the system involves the application of VAR models. VAR models are multivariate time series models capable of capturing the dynamic inter-dependencies between different stock market benchmarks.

Lag order selection is typically determined through methods like AIC (Akaike Information Criterion) or BIC (Bayesian Information Criterion), which is employed to determine the optimal number of lags in the VAR model.

Cointegration analysis may be integrated if the stock market benchmarks exhibit long-term relationships. This helps in modeling both short-term fluctuations and long-term trends..

- Data Preprocessing: Clean the data by handling missing values, removing outliers, and ensuring consistency in the time series.
- Stationarity Testing: Check for stationarity in the time series data using tests like the Augmented Dickey-Fuller (ADF) test.
- Differencing: If the data is non-stationary, differentiate the variables to achieve stationarity.
- Lag Selection: Determine the appropriate lag order for the VAR model using information criteria like AIC or BIC.
- Model Estimation: Fit the VAR model to the differenced or stationary data using least squares estimation or maximum likelihood estimation.
- Forecasting: Use the fitted VAR model to make forecasts for future time periods based on past data.

13Evaluation: Evaluate the forecasting performance of the VAR model using metrics like Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).

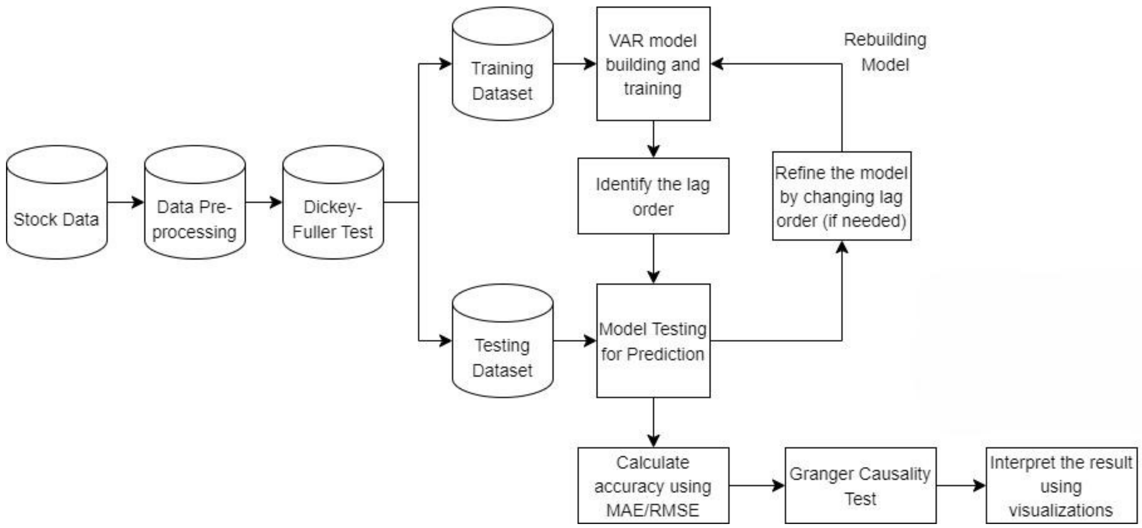


Figure 2: VAR modelling

6.2 VAR model in detail

A VAR model is a generalization of the univariate autoregressive model for forecasting a vector of time series. It comprises one equation per variable in the system. The right-hand side of each equation includes a constant and lags of all of the variables in the system. To keep it simple, we will consider a two-variable VAR with one lag. We write a 2-dimensional VAR as shown in Figure 3. In the provided context, the terms

e_1&e_2

represent white noise processes that could potentially be correlated at the same point in time. These processes are typically characterized by random fluctuations with a mean of zero and constant variance. Where:

$$y_{1,t} = c_1 + \phi_{11,1} y_{1,t-1} + \phi_{12,1} y_{2,t-1} + e_{1,t}$$
$$y_{2,t} = c_2 + \phi_{21,1} y_{1,t-1} + \phi_{22,1} y_{2,t-1} + e_{2,t},$$

Figure 3: Two variable lag in VAR

- $y_{1,t}$ and $y_{2,t}$ are the two variables in the VAR model.
- c_1 and c_2 are constants representing the intercept terms for each equation.

- $\phi_{11,1}$, $\phi_{12,1}$, $\phi_{21,1}$, and $\phi_{22,1}$ are coefficients capturing the influence of the lagged variables.

- $e_{1,t}$ and $e_{2,t}$ are white noise processes that may be contemporaneously correlated.

6.3 VAR in Forecasting

VAR models and their variations have achieved prominent status in the world of forecasting, especially in forecasting financial variables. They have played a pivotal role in the evaluation of risk and performance within investment portfolios. Many notable studies have underlined the importance of VAR modeling in quantitative assessments of risk and asset allocation strategies. There are studies that illuminated the instrumental role played by VAR models in making data-driven decisions. By using the power of the VAR model users can get a deep understanding of potential risks that are inherent in their investments while maintaining an acceptable level of risk. The evidence from multiple studies suggests that VAR models can have an impact on making forecasts thus helping in building a better portfolio.

6.4 Neural Network Auto-regression (NNAR) Implementation

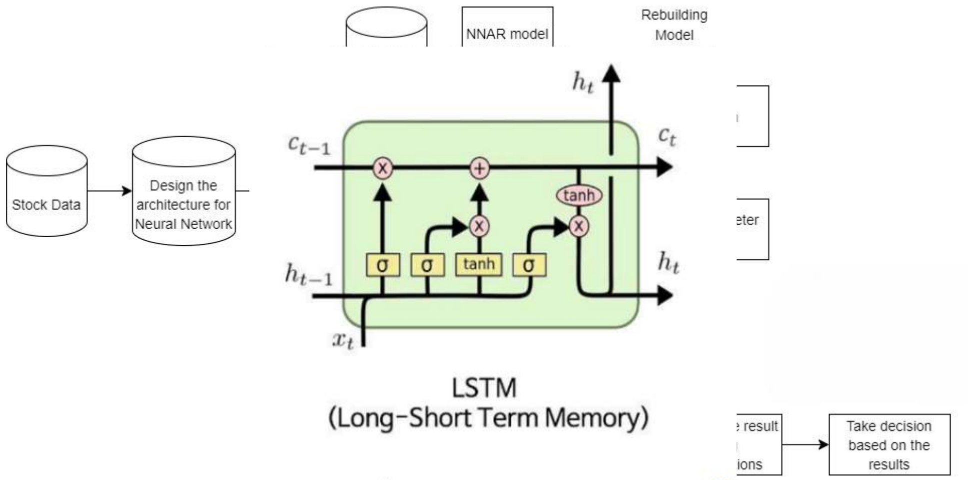
The core of this system leverages the implementation of Neural Network Auto-regression (NNAR) models. NNAR models are dynamic and adaptable tools for time series analysis, designed to capture intricate dependencies and patterns within different stock market benchmarks.

1.Network Architecture Design: The configuration of a Neural network is an important part of building the model because it involves in determining the number of hidden layers and the number of neurons in each layer. These choices affect the model's ability to make predictions.

2.Training and Learning: The NNAR model's training process is the process of making the model study the time series it uses optimization techniques like stochastic gradient descent(SGD). The model learns from the historical data and identifies the patterns and relationships that are used in making predictions.

3.Hyperparameter Tuning: Fine-tuning of hyperparameters such as learning rates, batch sizes, and regularization terms are required to get optimal performance. These adjustments are important in NNAR's predictive accuracy.

154.Time Lag Consideration: Similar to VAR, the NNAR approach also takes time lags into account. It involves selecting an optimal number of lag terms to capture short-term dependencies within the time series data.



$$y_{t-1} = (y_{t-1}, y_{t-2}, \ldots, y_{t-10})'$$

Figure 4: NNAR modelling

6.5 NNAR model in detail

- The NNAR model architecture consists of two LSTM layers followed by a Dense layer.

- Long Short-Term Memory(LSTM) units are a type of recurrent neural network(RNN) that captures long-term dependencies in sequential data. Each LSTM unit performs operations involving input, output, and forget gates to control the flow of information through the steps.

- This model is compiled using Adam optimizer and mean squared error loss function. Adam optimizer adjusts the learning rate adaptively during training, and MSE loss measures the difference between predicted and actual VWAP values.

- The mathematical operations involve matrix multiplication, activation functions tanh within the LSTM units, and optimization techniques during model training.

Figure 5 is a vector containing lagged values of the series, and f is a neural network with 6 hidden nodes in a single layer. The error series

ϵ_t

Figure 5: NNAR vector equation

is assumed to be homoscedastic (and possibly also normally distributed) We can simulate future sample paths of this model iteratively, by randomly generating a value for

ϵ_T

either from a normal distribution or by resampling from the historical values. So if

$\epsilon_T +1^*$

is a random draw from the distribution of errors at time T + 1, then

6.6 NNAR in Forecasting

NNAR models have emerged as powerful tools in the world of forecasting, particularly in predicting financial variables. The NNAR model can become one of the vital tools in risk management thus helping users to build a better portfolio. Notable researchers have highlighted the significance of NNAR modeling on quantitative risk evaluation and defining entry and exit points in stock. Studies have mentioned how NNAR models provide insights to make data-driven decisions. Multiple pieces of evidence suggest that NNAR models

$$y^*_{T+1} = f(\boldsymbol{y}_T) + \varepsilon^*_{T+1}$$

Figure 6: NNAR vector equation with Error series

7significantly impact the accuracy of forecasting and aid in building robust portfolios.

6.7 Model Evaluation

Upon thorough investigation of the existing studies, we can conveniently say that Mean Absolute Error(MAE), Root Mean Squared Error(RMSE), and Mean Absolute Percentage Error(MAPE) can be incorporated as model evaluation metrics that can assess the accuracy of predictions.

6.8 Granger Causality Testing

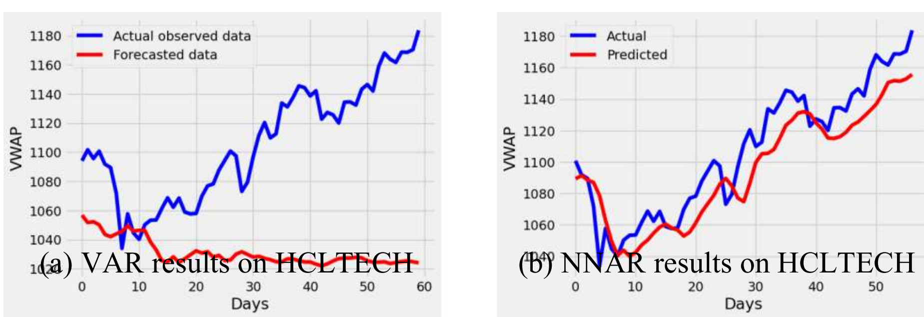
The utilization of Granger causality tests in the VAR framework is practiced within the field of time series analysis. First conceptualized by Granger in 1969, this notion searches into the intricate dynamics of temporal relationships by validating the capacity of past observations within one time series variable to predict future values with this round of predicted values. Granger causality tests provide an analytical approach besides the presence of direction and magnitude of these relationships, discerning the complex interplay of causal mechanisms laying the foundation in the volatile world of financial markets.

7 Results

We have taken one fiscal year's(1st April 2022 to 31st March 2023) data for each stock to train the models, i.e. data from the previous four quarters. The models were individually tuned to give better predictions and tested the results with the upcoming quarter(1st April 2023 to June 30th 2023).

We have conducted experiments on 25 companies across 5 sectors by at least considering 4 companies per sector. Below are some of the results from the experiments which will give an overall insight of the results.

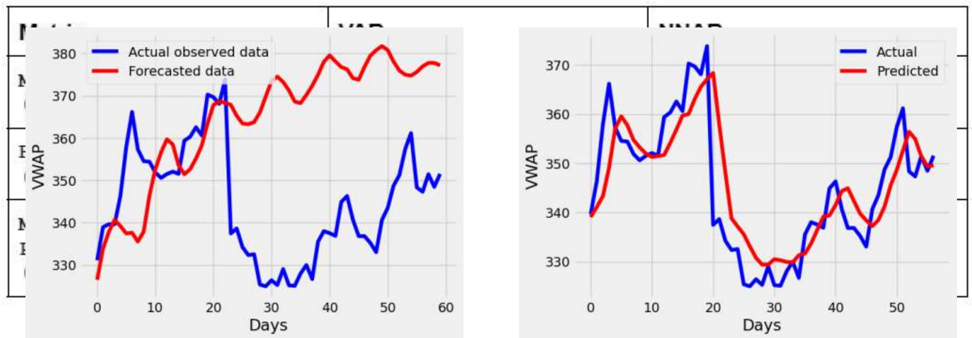
7.1 Performance Evaluation of Models with IT Sector



(a) VAR results on HCLTECH (b) NNAR results on HCLTECH

Figure 7: Predictions on HCL Tech

Although the VAR prediction has offered reliable information on the direction for a few days NNAR prediction turned out to be more reliable in terms of direction, momentum, and strength of the movement. Which reflected in the error rates of both model's predictions.



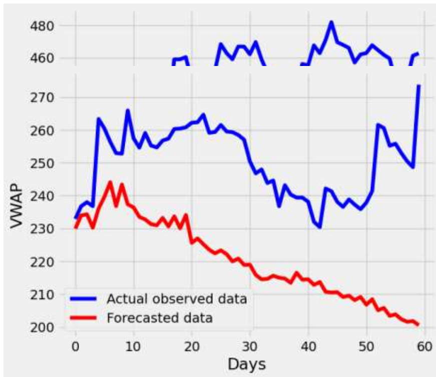
(a) VAR results on LATENTVIEW (b) NNAR results on LATENTVIEW

Figure 8: Predictions on Latent View Analytics Ltd

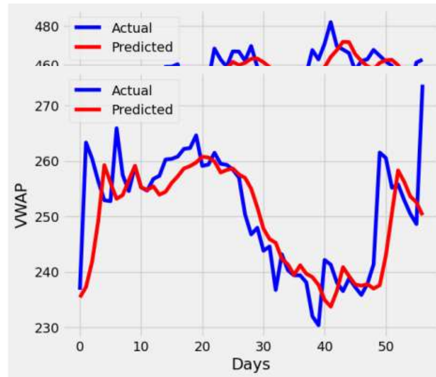
Above we can see how different the graphs are when predictions were made on the stock of Latent View Analytics Ltd and from the evaluation metrics we can see that the Error rate of VAR is way higher than NNAR and comparatively we can see that VAR has performed better on this stock than it did on other stocks but NNAR took precedence in comparison. Although the VAR prediction has offered reliable information on the direction for a few days NNAR prediction turned out to be more reliable in terms of direction, momentum, and strength of the movement.

Metric	VAR	NNAR
Mean Absolute Error (MAE)	24.9495487222	5.71327077765
Root Mean Squared Error (RMSE)	29.514462798	7.76480892639
Mean Absolute Percentage Error (MAPE):	7.37341413043	4.03678191095

7.2 Performance Evaluation of Models with Pharma Sector



(a) VAR results on AARTIDRUGS



(b) NNAR results on AARTIDRUGS

Figure 9: Predictions on Aarti Drugs Ltd

Above we can see how different the graphs are when predictions were made on the stock of Aarti Drugs Ltd, VAR offered a better prediction on both direction and momentum and gave reliable results whereas NNAR predicted the results with less error rate, NNAR also gave a better understanding on direction, momentum, and strength of the movement in comparison.

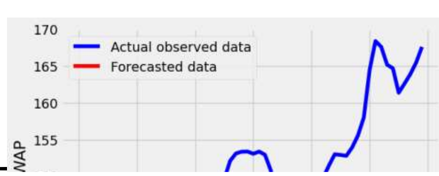
Metric	VAR	NNAR
Mean Absolute Error (MAE)	60.122931581	9.87825167472
Root Mean Squared Error (RMSE)	62.8253171224	13.6798839894
Mean Absolute Percentage Error (MAPE) :	13.4621574243	5.43678104201

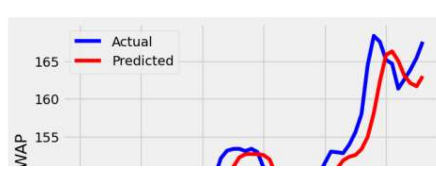
a) VAR results on SHILPAMED (b) NNAR results on SHILPAMED

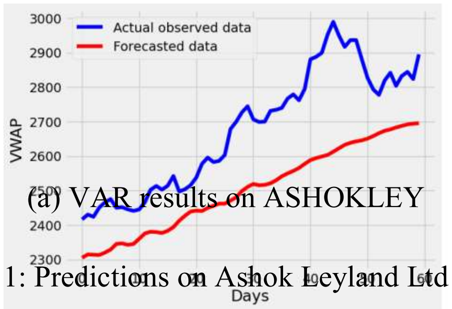
Figure 10: Predictions on Shilpa Medicare Ltd

Above we can see how different the graphs are when predictions were made on the stock of Shilpa Medicare Ltd, VAR prediction has offered good insights in terms of direction and overall movement of the stock for most of the days, but NNAR showed better results in terms of overall performance considering direction, momentum and strength of the movement.

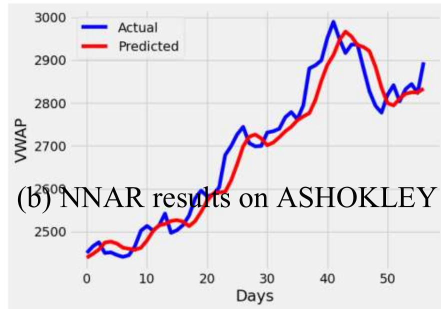
7.3 Performance Evaluation of Models with Automobile sector







(a) VAR results on ASHOKLEY



(b) NNAR results on ASHOKLEY

Figure 11: Predictions on Ashok Leyland Ltd

Above we can see how different the graphs are when predictions were made on the stock of Ashok Leyland Ltd, The VAR model has offered a good insight on the overall direction of the stock but NNAR has given better results in direction, momentum, and strength of prediction. By also looking at evaluation metrics we can say that VAR has performed better in terms of error rate than the other VAR performances but NNAR gave better results there too, thus we can say that NNAR has performed better in this particular stock.

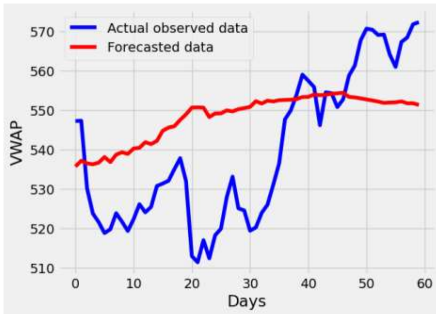
Metric	VAR	NNAR
Mean Absolute Error (MAE)	6.94413098177	2.09605042575
Root Mean Squared Error (RMSE)	9.26298095207	2.90126979515
Mean Absolute Percentage Error (MAPE):	4.451111836	6.45782566452

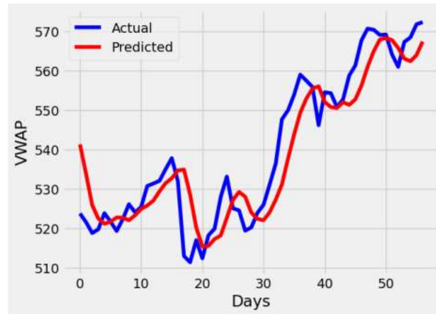
(a) VAR results on HEROMOTOCO (b) NNAR results on HEROMOTOCO

Figure 12: Predictions on Hero Motocorp Ltd

Above we can see how different the graphs are when predictions were made on the stock of Hero Motocorp Ltd, The VAR model has offered good insight into the overall direction but it has failed to predict abnormal highs of the stock after 40 days but NNAR has given better results in direction, momentum, and strength of prediction and also have captured the abnormal highs after 40 days. By also looking at evaluation metrics we can say that VAR has performed better in terms of error rate than the other VAR performances but NNAR gave better results there too, thus we can say that NNAR has performed better in this particular stock.

7.4 Performance Evaluation of Models with FMCG Sector





Metric	VAR	NNAR
Mean Absolute Error (MAE)	174.341239977	33.2458357319
Root Mean Squared Error (RMSE)	188.914845072	42.603033866
Mean Absolute Percentage Error (MAPE):	6.39682861052	7.16868025921

Metric	VAR	NNAR
Mean Absolute Error (MAE)	29.8900785407	4.93225067674
Root Mean Squared Error (RMSE)	32.6535091813	7.74269908714
Mean Absolute Percentage Error (MAPE):	11.8492720447	4.35711033033

(a) VAR results on DABUR (b) NNAR results on DABUR

Figure 13: Predictions on Dabur India Ltd

Above we can see how different the graphs are when predictions were made on the stock of Dabur India Ltd, The VAR model offered good insight on the overall direction for 48 days, and from there it failed to predict the upward movement of stock and it also failed to show the lows that were noticed during period of 15-25 days. Whereas NNAR has given results that were reliable in terms of strength, direction, and the overall movement of the stock. By also looking at the evaluation metrics we can say that NANR is a better choice for this stock even though VAR has shown somewhat better results than other VAR performances.

a) VAR results on VIPIND (b) NNAR results on VIPIND

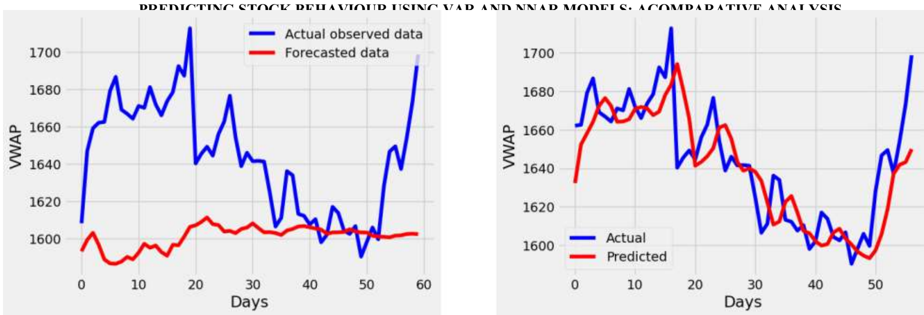
Figure 14: Predictions on VIP Industries Ltd

Metric	VAR	NNAR
Mean Absolute Error (MAE)	16.0338897909	5.83840477659
Root Mean Squared Error (RMSE)	18.7049336059	7.54471980556
Mean Absolute Percentage Error (MAPE):	3.0103808219	3.78338152851

Above we can see how different the graphs are when predictions were made on the stock of VIP Industries Ltd, The VAR model has given a good insight into the overall direction of the stock movement but NNAR has given proper insights on the overall highs and lows of the stock movement and was able to capture those abnormal movements. By also looking at the evaluation metrics we can say that NNAR has performed better than VAR even though the difference is very low for this particular stock.

7.5 Performance Evaluation of Models with Banking Sector

Metric	VAR	NNAR
Mean Absolute Error (MAE)	15.9109647721	9.3112222879
Root Mean Squared Error (RMSE)	18.8131477813	11.4650352292
Mean Absolute Percentage Error (MAPE):	2.59690135579	3.29664917169



a) VAR results on HDFCBANK (b) NNAR results on HDFCBANK

Figure 15: Predictions on HDFC Bank Ltd

Above we can see how different the graphs are when predictions were made on the stock of HDFC Bank Ltd, the VAR model has provided some insights on the overall direction of the stock for 22 days, and from there on it failed to predict the lows that were observed, on the other hand, NNAR was successful in predicting the abnormalities with precision and can be said as a better choice in terms of Strength, Momentum and the Direction of the stock movement. Also by looking at the evaluation metrics, we can say that VAR has given a higher Error rate and NNAR has given predictions with a lower Error rate, making NNAR as a better model for this particular stock.

(a) VAR results on KOTAKBANK (b) NNAR results on KOTAKBANK

Figure 16: Predictions on Kotak Mahindra Bank Ltd Fully Paid Ord. Shrs

Metric	VAR	NNAR
Mean Absolute Error (MAE)	43.8170280755	14.4410315584
Root Mean Squared Error (RMSE)	54.0717019057	18.8251839295
Mean Absolute Percentage Error (MAPE) :	2.63206025767	2.04598091843

Above we can see how different the graphs are when predictions were made on the stock of Kotak Mahindra Bank Ltd Fully Paid Ord. Shrs, the VAR model was successful in capturing the overall momentum of the stock movement but has failed to precisely capture the abnormalities in the stock movement which was done correctly by the NNAR predictions. NNAR has accurately predicted the abnormalities making it more reliable. Also by looking at the evaluation metrics, we can say that VAR predictions have given a higher error rate than NNAR predictions, thus making NNAR a better fit for this particular stock.

8 Conclusions

When it comes to the predictions three important abstract aspects attract the traders and build their trust in the predictive model, they are Direction, Momentum, and Strength. To put these terms in context

- Direction: A predictive model should determine the direction of the stock from a particular point to another point between any two given points of the graph that match the observed values.

- Momentum: A predictive model should determine the overall movement of the stock, be it upward or downward correctly.

- Strength: A predictive model is said to have good strength when the values predicted by the model are very near to the values observed.

From the results, we can infer that the NNAR model has outperformed the VAR model in every stock that was experimented with. After experimenting with 20 companies across 5 sectors, we can confidently say that the NNAR model gave predictions that can be trusted by the traders by scrutinizing the above three factors.

A predictive model's ultimate goal should be serving its stakeholders right. Here in our research, we wish to empower traders with data-driven decisions who can determine their entry and exit from a stock to maximize their profit. Usually, most traders look at a picture for a month or two in our results we have given our best to produce reliable predictions for a quarter that is three months which is a huge window. Another aspect is that traders need to make decisions faster, comparatively the NNAR model is a faster model to execute than the other Bootstrapping and Bagging algorithms in machine learning.

The benchmark VWAP is chosen in our project to predict the stock market because this particular benchmark is a combined metric of several factors. Let us say there is an important decision that was taken and has been announced to the market, it is not the price that is affected immediately after the announcement it is the volume. The volume transacted and the price at which the volume was transacted determines the behavior of the stock. Since VWAP is a combined metric of volume and price, the predicted fluctuations in VWAP can be mapped to the fluctuations of the observed prices. The financial decision that was taken cannot be fed into the model but we can feed the effect of that decision by considering VWAP. When we take the price of the stock as our variable to predict we can only determine the price volatility but when we consider VWAP we can determine price volatility and also the volume volatility.

Keeping the fact that Machine learning models have emerged to be able to capture the intricate dependencies and the volatile nature of the stock market, we never know which economic change will bring which kind of changes in the behavior of stocks at any time in the future. There is a need to develop a dynamic model that can adapt to any kind of changes and make precise predictions, considering VWAP has solved that problem to some extent but

Metric	VAR	NNAR
Mean Absolute Error (MAE)	98.9491701476	19.8630321409
Root Mean Squared Error (RMSE)	112.683198132	26.0477765509
Mean Absolute Percentage Error (MAPE):	5.15143813801	3.09989778404

there is still a need for better research in developing dynamic models.

Predictions are said to be reliable in the financial sector when an interested trader can use a particular set of predictions to determine when to buy a stock and when to sell it. A model is said to be robust when it generates predictions that can be used to make informed decisions on when to enter and when to exit.

References

[1] Iqra Sadia, Atif Mahmood, Miss Laiha Binti Mat Kiah, and Saaidal Razalli Azzuhri, Analysis and Forecasting of Blockchain-based Cryptocurrencies and Performance Evaluation of TBATS, NNAR and ARIMA, University of Malaya 2022, IEEE International Conference on Artificial Intelligence in Engineering and Technology 2022

[2] Doron Avramov, Si Cheng, and Lior Metzker, Machine Learning versus Economic Restrictions: Evidence from Stock Return Predictability, University of Maryland 2019, Midwest Finance Association Annual Meeting 2020

[3] Manish Kumar and M. Thenmozhi, Forecasting Stock Index Movement: A Comparison of Support Vector Machines and Random Forest, International Conference on Computational Intelligence and Data Science (ICCIDS 2019), IIT Madras,2019

[4] Toni Toharudin, Resa Septiani Pontoh, Rezzy Eko Caraka, Solichatus Zahroh, Antony Akbar, Bens Pardamean, Rung Ching Chen Indonesia in Facing New Normal: An Evidence Hybrid Forecasting of COVID-19 Cases Using MLP, NNAR and ELM , 2021, Volume 29, Issue 2: June 2021

[5] Sandesh Kancherla, Dr.M.Rajasekar, Dr.N.Sandhya, Forecasting and Predicting Stock Market Trends using Feature Selection,2018, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, India. Journal of Critical Reviews,Vol 7, Issue 06, 2020.

[6] J Sharmila Vaiz, M Ramaswami, A Hybrid Model to Forecast Stock Trend Using Support Vector Machine and Neural Networks ,2019, Madurai Kamaraj University, Madurai, India. International Journal Of Engineering Research And Development, Volume 13, Issue 9 (September 2017).

[7] Alev Dilek Aydin, Seyma Caliskan Cavdar, Comparison of Prediction Performances of Artificial Neural Network (ANN) and Vector Autoregressive (VAR) Models by using the Macroeconomic Variables of Gold Prices, Borsa Istanbul (BIST) 100 Index and US Dollar-Turkish Lira (USD/TRY) Exchange Rates , Halic University, 3rd Economics and Finance Conference, Rome, Italy, April 14-17, 2015 and 4th Economics and Finance Conference, London, UK, August 25-28, 2015

[8] Abdullah M. Almarashi, Muhammad Daniyal and Farrukh Jamal A novel comparative study of NNAR approach with linear stochastic time series models in predicting tennis player's performance BMC Sports Science, Medicine and Rehabilitation, Jan 25, 2024

[9] Bruno Miranda Henrique, Vinicius Amorim Sobreiro, Herbert KimuraStock price prediction using support vector regression on daily and up to the minute prices The Journal of Finance and Data Science 183-201, 27th Apr 2018,

[10] Ghassane Benrhmach, Khalil Namir, Abdelwahed Namir, and Jamal BouyaghroumniNonlinear Autoregressive Neural Network and Extended Kalman Filters for Prediction of Financial Time SeriesJournal of Applied Mathematics Volume 2020, Article ID 5057801,27 April 2020

[11] Mohamad As'ad, Sujito, Sigit SetyowibowoPrediction of Daily Gold Prices Using an Autoregressive Neural NetworkJurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi Vol. 5 No. 2, July 2020

[12] B Ramana Murthy, Shaik Nafeez Umar, K Vijaya Kumar and G Tejaswini ReddyBSE Sensex forecasting using a machine learning approach International Journal of Statistics and Applied Mathematics 2023; 8(5): 209-212, 04 Jun 2023

[13] Azlinda Saadon, Jazuri Abdullah, Ihsan Mohd Yassin, Nur Shazwani Muhammad, Junaidah Ariffin Nonlinear multi independent variables in

quantifying river bank erosion using Neural Network AutoRegressive eXogenous (NNARX) model Elsevier Ltd. 16 February 2024

[14] latunji Abdul Shobande and Joseph Onuche EnemonaA Multivariate VAR Model for Evaluating Sustainable Finance and Natural Resource Curse in West Africa: Evidence from Nigeria and Ghana Sustainability 2021, 13, 2847

[15] Charles O. Manasseh, Jonathan E. Ogbuabor, Charles N. Anumudu, Felicia C. Abada, Martins A. Okolie and Okoro E. OkoroThe Causal Effect of Stock Market Development, Financial Sector Reforms and Economic Growth: The Application of Vector Autoregressive and Error Correction ModelInternational Journal of Economics and Financial Issues, 2018, volume 8 issue 2, 357-369

[16] Helmut Herwartz, Shu WangStatistical identification in panel structural vector autoregressive models based on independence criteriaJournal of Applied Econometrics published by John Wiley and Sons Ltd. 2024 Vol 1 issue 20

[17] Ilyes Abidi, Kamel TouhamiSafe Haven for Crude Oil: Bitcoin or Precious Metals? New Insight from Time Varying Coefficient-Vector Autoregressive ModelInternational Journal of Energy Economics and Policy, 2024, 14(1), 184-195

[18] Sun Ho Lee, Kyu Ho KangA Large Vector Autoregression of

the Yield Curve and Macroeconomic Variables with No-Arbitrage RestrictionSSRN Electronic Journal Feb 2024

[19] Almarashi AM, Daniyal M, Jamal F (2024) Modelling the GDP of KSA using linear and nonlinear NNAR and hybrid stochastic time series models. PLoS ONE 19(2): e0297180

[20] P.V. Thayyib, Muhammed Navas Thorakkattle, Faisal Usmani, Ali T Yahya and Najib H.S FarhanForecasting Indian Goods and Services Tax revenue using TBATS, ETS, Neural Networks, and hybrid time series models, Cogent Economics and Finance(2023), 11:2, 2285649,