2023 4th IEEE Global Conference for Advancement in Technology (GCAT)
Bangalore, India. Oct 6-8, 2023

Hybridizing Statistical and Neural Network Models for Enhanced Stock Price Forecasting

Jisha R C
Department of Computer Science and Applications
Amrita School Of Computing
Amrita Vishwa Vidyapeetham
Amritapuri, India
jisha@am.amrita.edu

Hisana Nazeer
Department of Computer Science and Applications
Amrita School Of Computing
Amrita Vishwa Vidyapeetham
Amritapuri, India
hisananazeer1@gmail.com

Kavya R
Department of Computer Science and Applications
Amrita School Of Computing
Amrita Vishwa Vidyapeetham
Amritapuri, India
kavyakairali47@gmail.com

***Abstract*—This paper presents a hybrid prediction approach for stock price forecasting by combining Autoregressive Integrated Moving Average (ARIMA) and Seasonal Autoregressive Integrated Moving Average (SARIMA) with Long Short-Term Memory (LSTM) or Gated Recurrent Unit (GRU) models to create ensemble models. The objective is to enhance the precision and dependability of stock price predictions by using the benefits of each individual model. The ensemble models proficiently manage trends, seasonality, non-stationarity, and identify long-term correlations in the TITAN stock price data. This research uses the Volume Weighted Average Price (VWAP) as a key factor to accurately depict the behavior of the market. The ensemble models are structured using the residual method resulting in improved predictions. The superior performance of the ensemble models is confirmed through evaluation using Mean Squared Error (MSE) as the metric. Visualizations in the form of graphs are employed to compare actual and predicted values, demonstrating the accuracy and performance of the models in capturing stock market trends and patterns. The models also exhibit high accuracy in forecasting future stock prices, highlighting their practicality in real-life scenarios.**

***Index Terms*—ARIMA, SARIMA,LSTM, GRU, Residual approach, stock market price**

I. Introduction

Foreseeing stock prices correctly holds immense importance for financial institutions, traders, and investors. It allows them to gather helpful knowledge for informed decisions and effective optimization of investment plans. Although, due to the non-linear and unsure attributes of the financial market, it is an extremely tough task to predict stock prices as it is affected by numerous factors like worldwide incidents, market sentiments, and macroeconomic indicators.

Traditional time series models like Autoregressive Integrated Moving Average (ARIMA) and its seasonal variation, Seasonal ARIMA (SARIMA), have been extensively used for predicting stock prices because they can analyze linear relationships and temporal dependencies. Despite their effectiveness, these models have limitations in terms of their assumptions and inability to capture complex data patterns.

The potential for improving the prediction of stock prices has been expanded by the introduction of machine learning and deep learning techniques. A hybrid approach that combines traditional time series models such as ARIMA with contemporary deep learning models such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks has become a promising solution for addressing the limitations of conventional methods.

In order to enhance the precision of stock price prediction, our research proposes a novel hybrid model that unites SARIMA and ARIMA together with LSTM or GRU models. This innovative ensemble method aims to leverage the unique advantages of each technique. Our proposed ARIMA-SARIMA-LSTM and ARIMA-SARIMA-GRU ensemble models outperform existing hybrid models. Our assessment comprised an extensive examination of the effectiveness of our ensemble model. We conducted a thorough study by comparing its predictive precision with ARIMA, SARIMA, LSTM and GRU models. This research utilizes the Titan stock market dataset to develop the ensemble models. This resulted in the demonstration of the ensemble model's superiority over leading techniques. Our findings are a valuable addition to the existing research on the application of machine learning in finance.

II. Related Work

Several studies have investigated the application of deep learning models in predicting stock prices. For instance, Balaji et al. scrutinized the potential of deep learning in forecasting BANKEX data[11]. Jaiswal and Singh further expanded this line of research by developing a hybrid Convolutional Recurrent (CNN-GRU) Model for stock price prediction[15]. These works underscore the remarkable capacity of deep learning models in learning complex patterns from stock market data, thereby enabling accurate predictions.

In the domain of financial forecasting, substantial advancements have been made over the years, particularly with the emergence of machine learning models. One notable contribution was made by Emmanuel Dave et al., who proposed a hybrid ARIMA, LSTM model to forecast Indonesian exports[6]. Their model effectively combined traditional statistical methods with modern deep learning techniques, establishing the viability of such hybrid systems[16]. This approach has been further extended by Li and Yang, who applied a similar hybrid model, specifically a SARIMA, LSTM to forecast air temperature[7]. [17]presented an innovative approach to short-term stock price forecasting by incorporating sentiment analysis and Artificial Neural Networks. This work demonstrated how non-traditional data sources like market sentiment data could be integrated with machine learning techniques to enhance prediction performance. In a significant contribution to the field, K. Chen, Y. Zhou, and F. Dai conducted a case study on the China stock market using an LSTM-based method for predicting stock returns which exemplifies the use of recurrent neural networks, particularly Long Short-Term Memory (LSTM) networks, in predicting stock market trends[12]. Further attention has been drawn towards the Indian stock market with researchers such as Jisha et al. analyzing networks using correlation and association mining-based networks with different centrality measures[1]. This type of analysis enables a deeper understanding of the underlying structure of the stock market, providing essential insights for prediction models. Complementing this, researchers like Hiransha et al. and Sreelekshmy et al. have endeavored to predict NSE stock prices specifically using deep learning models[19][21].

Data-driven models also play an integral role in stock price prediction. Deepthi P. K. and colleagues proposed a data-driven model utilizing dynamic mode decomposition[18], while Unnithan et al. introduced a data-driven model for day-wise stock prediction[20]. These methodologies harness the inherent patterns and trends within the data to make predictions, making them an essential component of any comprehensive review of stock prediction methodologies.

Finally, the extensive survey by Binoy B. Nair and Mohandas on artificial intelligence applications in financial forecasting[22] warrants attention. Their work offers an overview of the various artificial intelligence applications used in this domain, thereby illuminating the wide-ranging methods available and the breadth of research already conducted in this area. This kind of meta-analysis is invaluable for identifying the state of the art and potential future directions in the field of stock price prediction.

III. Proposed Work

The objective of this research endeavor is to engineer an innovative predictive model for stock market prices for the TITAN stock market dataset. The proposed research project, in essence, seeks to combine the predictive capabilities of sophisticated models: Auto Regressive Integrated Moving Average (ARIMA), Seasonal AutoRegressive Integrated Moving Average (SARIMA), Long Short-Term Memory networks (LSTM), and Gated Recurrent Units (GRU).

The intention is to construct two distinct ensemble models:

- ARIMA-SARIMA-LSTM
- ARIMA-SARIMA-GRU

At the heart of this research project lies a residual-based methodology for the amalgamation of the models. This involves employing ARIMA and SARIMA models for initial training and prediction of stock market prices, followed by the computation of the residuals resulting from the original value and average of initially predicted values using ARIMA and SARIMA model. Subsequently, these residuals will be utilized as the training data for the LSTM in ARIMA-SARIMA-LSTM and GRU in ARIMA-SARIMA-GRU models. This innovative approach aims to harmonize the advantages of both statistical and deep learning models into a robust ensemble.

The principal hypothesis is that this proposed ensemble methodology would outperform the individual models in terms of prediction accuracy. This hypothesis will be validated through a comprehensive assessment of the Mean Squared Error (MSE) of the predictions yielded by both the ensemble and the individual models. The model will then be used to predict the subsequent day's stock price.

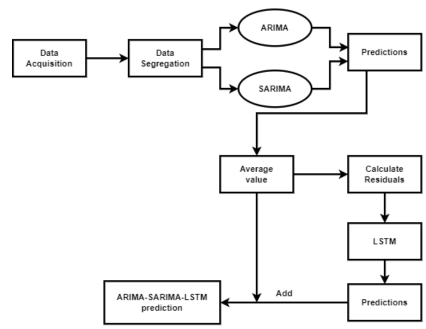

Fig. 1. Architectural diagram of ARIMA-SARIMA-LSTM model

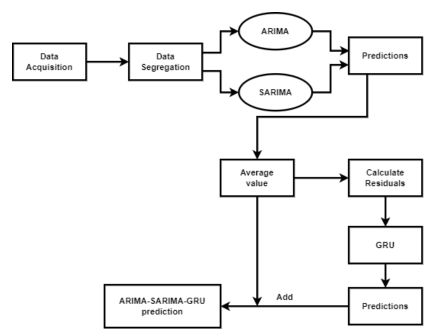

Fig. 2. Architectural diagram of ARIMA-SARIMA-GRU model

We aim to advance the frontiers of machine learning in financial forecasting, specifically by enhancing prediction accuracy. This study uniquely combines the strengths of different models to offset their respective weaknesses, creating a synergistic ensemble that performs more effectively than any of the individual models alone. In this way, we hope to demonstrate the power and potential of such multi-faceted, collaborative modeling approaches in tackling complex prediction tasks.

IV. Methodology

A. Data Acquisition and Preprocessing

The preliminary phase of this research is centered around the collection of the TITAN stock market dataset. With over two decades of data, ranging from the year 2000 to 2021, the models will have a substantial volume of information to learn from.The dataset was harnessed as a cornerstone to construct and train models oriented towards forecasting future stock prices with substantial precision and dependability.The acquisition of the dataset marks the commencement of the research. However, raw data typically contains inconsistencies, missing values, or other elements that may lead to inaccuracies in model training and prediction. Therefore, it is of the utmost importance to preprocess the data before feeding it into the models.

Following the data cleaning process, the dataset was then scaled using Min-Max scaling. Stock market data can have large variations in magnitude, which can affect the performance of the machine learning models, particularly deep learning models like LSTM and GRU. Therefore, normalization methods are done to scale the data. Normalization rescales the data into a range of [0,1].

The dataset encapsulates various parameters, including the Volume Weighted Average Price(VWAP), Open, High, Low, and Close prices, among others. Central to our research, we introduce ensemble models purposed for stock market prediction with a particular emphasis on VWAP (Volume Weighted Average Price) as our target variable. VWAP integrates both price and volume data, yielding a more comprehensive depiction of market activity over a designated time frame. Contrarily, Open, High, Low, and Close prices merely supply fragmentary glimpses of price movement at discrete instances, without accounting for the trading volume. Moreover, VWAP's characteristic of being less susceptible to fleeting price fluctuations compared to individual price attributes, due to its essence as a volume-weighted average price, offers significant advantages. It minimizes the influence of noise, thereby delivering a smoother delineation of price trends. This attribute of VWAP has the potential to enhance model performance by providing a more stable and reliable basis for prediction.

In order to make effective use of time-series forecasting models such as ARIMA, SARIMA, LSTM, and GRU, it is crucial to transform the continuous stream of data into structured sequences. In this study, we have selected a window size of 50, meaning that the models are trained to use the past 50 days of VWAP values to predict the 51st day's value. This window size represents a balance between including sufficient historical data for the model to identify trends and patterns, while also not overwhelming the model with an excessively large input size.These form the input and output pairs for training the models.

In the context of predictive modelling, an essential preliminary step is the segregation of available data into distinct training and validation sets. Data splitting, on the other hand, is performed based on a predefined cutoff point. By default, the initial 5000 data points are allocated to the training set, with the remaining ones forming the validation set. It is crucial to reshape the data to conform to the preferred three-dimensional arrangement of the LSTM and GRU models for optimal results. The conversion involves changing the data from a two-dimensional array (samples, features) to a three-dimensional structure (samples, time steps, features) to ensure that the input data aligns with the model's structure.

B. Model Overview and Selection

B1. ARIMA (AutoRegressive Integrated Moving Average)

The AutoRegressive Integrated Moving Average, or ARIMA model, is an extension of the autoregressive moving average model. It is employed to make predictions about future points in a time series data. ARIMA models are used when the data exhibit signs of non-stationarity and require the application of a differencing step one or more times to eliminate such non-stationarity[2]. The ARIMA model efficiently captures trends, noise, and autocorrelations in time-series data[3,4]. We incorporate the ARIMA model into our ensemble forecasting approach to enhance the accuracy of our future stock price predictions

B2. SARIMA (Seasonal AutoRegressive Integrated Moving Average)

SARIMA is designed to cater specifically to univariate time series data that includes seasonal patterns. SARIMA introduces three additional hyperparameters for defining the autoregression (AR), differencing (I), and moving average (MA) of the seasonal part of the series[5]. These new parameters complement the usual ARIMA parameters.By including a seasonal component in the SARIMA model, we add a valuable supplement to the ARIMA model. This allows our ensemble to better incorporate seasonal variations, which the ARIMA model may not be able to fully accommodate.

B3. LSTM (Long Short-Term Memory)

LSTM (Long Short-Term Memory)'s unique edge lies in its potential to store and transmit information across prolonged sequences. Memory cells and gating mechanisms ensure this ability is achieved.The functioning of the LSTM cell comprises a number of elements such as a memory-acting cell state, a regulator of information that must be abandoned known as the forget gate, a tracker for new information to assimilate recognized as the input gate, alongside an output gate capable of determining the output based on present input and past hidden state.LSTM can learn and remember important data for extended periods by utilizing gating mechanisms, allowing for the creation of accurate models for long-term dependencies[8].Its ability to comprehend intricate temporal dependencies and store significant data over longer sequences renders it indispensable for elaborate sequential modeling undertakings[9,10].

B4. GRU (Gated Recurrent Units)

GRU implements gating mechanisms through the use of an update gate and a reset gate, which regulate the network's information flow. The update gate dictates how much of the prior hidden state should be retained for the present time step, whereas the reset gate decides to what extent the preceding information must be ignored. GRU networks achieve a harmonious blend of computational efficiency and representational prowess through the integration of gating mechanisms. [14]The flexibility of GRU networks lies in their ability to incorporate gating mechanisms, which selectively update and erase information based on its importance to the task at hand. This adaptability is what contributes to their computational efficiency and makes them effective learners even in situations where data is scarce[13]. Such networks are particularly adept at capturing dependencies in sequential data, enabling them to unlock hidden insights and patterns.

C. Ensembling the Models

The combination of ARIMA, SARIMA with LSTM or GRU models in an ensemble approach can provide numerous advantages compared to using each model independently. Primarily, each model is able to capture distinct features of the data. While ARIMA and SARIMA concentrate on linear connections and sequential dependencies, LSTM is better equipped to discern non-linear patterns and extended dependencies.

Although SARIMA is an extended version of ARIMA, the base ARIMA model might still be better at modeling the underlying non-seasonal components in the data because of its simplicity. The complexity of SARIMA makes it more likely to overfit the data, particularly when dealing with non-seasonal data or data with weak seasonality. By incorporating ARIMA in the ensemble, the risk of overfitting may be reduced, providing a more generalized and robust forecasting model.In an ensemble, having redundancy (i.e., multiple models that can capture similar patterns) can be beneficial by making the ensemble more robust to potential changes in the data or errors in individual model forecasts.

ARIMA and SARIMA models assume a linear relationship between input features and the output variable, unpredictable fluctuations or non-linear relationships could prove challenging for these models to capture in certain scenarios. Inaccurate predictions and restricted model performance may occur, and in such situations, neural networks are efficient at identifying non-linear relationships and patterns in the data. They achieve this by using several hidden layers with nonlinear activation functions, which allows the model to capture complex interactions and dependencies that cannot be identified by linear models alone. Using TensorFlow, the LSTM or GRU layers, with their nonlinear activation functions, are trained to tease out these complex patterns, serving as a rectification mechanism for linear model's shortcomings.

With the TITAN dataset, ARIMA and SARIMA, implemented via statsmodels, were first trained and their predictions then averaged – a process designed to amalgamate their individual insights.The combined understanding we had was not flawless, and that's where the residuals play a role. By subtracting the average prediction values from the real stock prices, we successfully extracted the imperfections or residuals. We utilized TensorFlow to train LSTM and GRU models with these residuals, essentially guiding them to fix the mistakes made by their linear counterparts. We utilized a range of libraries for different purposes in this project. TensorFlow was employed for constructing and training neural networks, pandas was utilized for manipulating and processing data. Numpy supported us in performing numerical operations and transforming data. We relied on statsmodels for implementing ARIMA and SARIMA time series forecasting models. Sklearn aided us in data preprocessing and evaluating model performance. Lastly, matplotlib was used to visualize the results and performance of the model.

In conclusion, the ensemble models used in this study were mathematically formulated as follows:

- **ARIMA-SARIMA-LSTM :**

$$\text{Ensemble}(t) = \left(\frac{\text{ARIMA}(t) + \text{SARIMA}(t)}{2}\right) + \text{LSTM}\left\{y(t) - \left[\frac{\text{ARIMA}(t) + \text{SARIMA}(t)}{2}\right]\right\}$$

- **ARIMA-SARIMA-GRU :**

$$\text{Ensemble}(t) = \left(\frac{\text{ARIMA}(t) + \text{SARIMA}(t)}{2}\right) + \text{GRU}\left\{y(t) - \left[\frac{\text{ARIMA}(t) + \text{SARIMA}(t)}{2}\right]\right\}$$

Here, t denotes each time step, ARIMA(t) and SARIMA(t) represent the predictions of the ARIMA and SARIMA models, respectively, and LSTM(t) or GRU(t) signifies the predictions of the LSTM or GRU models, respectively, trained on the residuals. These formulas embody our approach of leveraging diverse predictive models to yield more robust and accurate stock market predictions.

The overarching aim was to reduce the prediction error and enhance the accuracy and reliability of the stock market price forecasts. This method capitalized on the strengths of both statistical models and neural networks, generating ensemble models with a higher potential for performance than any individual model.

After the completion of the training phase, our research proceeded to the validation and performance measurement of the models. This critical juncture involved the computation of the Mean Squared Error (MSE), a customary metric in the field of machine learning and data science for assessing the accuracy of a model's predictions, specifically in regression problems.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Here, n represents the total number of observations or data points. y_i denotes the actual value for the i-th observation, and $\hat{y}_i$ is the model's predicted value for the i-th observation.

MSE computes the average squared difference between the predicted and actual values, essentially quantifying the discrepancy between the model's forecasts and the genuine stock prices within the TITAN dataset. A lower MSE indicates a more accurate fit of the model to the data. The calculated MSE facilitated a comprehensive evaluation of the performance of the proposed ensemble models, ARIMA-SARIMA-LSTM and

ARIMA-SARIMA-GRU, when applied to the TITAN dataset. Additionally, we visually assessed the performance of the models by plotting graphs that compare the actual stock prices with the corresponding predicted values.

Our study conducted an experiment to compare the performance of individual models: ARIMA, SARIMA, GRU, and LSTM with that of the proposed ensemble models: ARIMA-SARIMA-LSTM and ARIMA-SARIMA-GRU. Each model, inclusive of the ensemble, was independently constructed and tested. The testing phase involved comparing the model's predictions with the actual data, calculating the MSE to offer a quantifiable measure of each model's predictive accuracy. Our findings indicated that the proposed ensemble models outperformed the individual models in forecasting stock prices, thereby validating the superiority of the ensemble approach, as it leverages the strengths of different models while mitigating their weaknesses.

In the subsequent phase of our study, we ventured to forecast future stock prices, utilizing the TITAN stock price dataset comprising the past year's data up to the present day. The models were then tasked to predict the following day's stock price. This exercise is aligned with standard methodologies in time-series forecasting and presents an opportunity to examine the model's utility in practical, real-world scenarios.

Our models performed admirably during this phase, with the predictions aligning well with actual data. The superior performance of our models underpins their potential applicability in forecasting future stock prices. These promising results provide compelling evidence of the model's capabilities, further strengthening the case for the use of the ensemble models in financial forecasting.

V. Result

Our experimental analysis of both individual and ensemble models has yielded some significant results. The graph represents a comparison between the VWAP (Volume Weighted Average Price), plotted on the y-axis, and the i-th prediction values, plotted on the x-axis. It allows for visual analysis of how well the predictions align with the actual VWAP values. Specifically, we sought to determine the accuracy of each model in predicting stock prices and utilized the Mean Squared Error (MSE) as our performance metric.

A. Ensemble models

A.1 ARIMA-SARIMA-LSTM

The graph offered an intriguing comparison between the actual stock prices from the test dataset and the predictions generated by the ensembled ARIMA-SARIMA-LSTM model. The predicted stock prices demonstrated a remarkable alignment with the original values, exhibiting the model's proficiency in capturing both the broader trend and sharp fluctuations of the market. The Mean Squared Error (MSE) for the ARIMA-SARIMA-LSTM model was significantly low as 4.4316 x 10^{-5}, indicating a high degree of predictive accuracy. The ARIMA-SARIMA-LSTM model exhibited a high degree

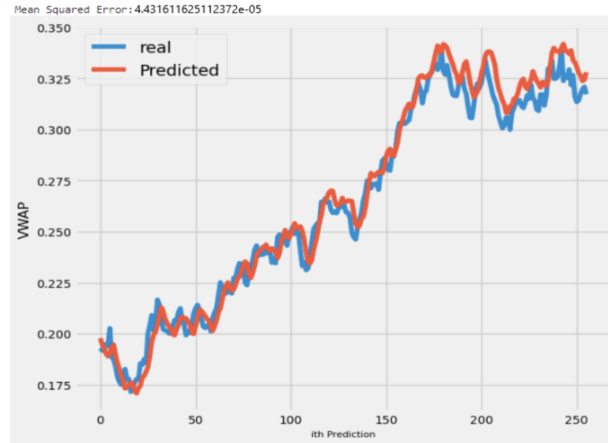

Fig. 3. Performance of the ARIMA-SARIMA-LSTM on test dataset

of effectiveness in predicting stock prices, capably modeling both the macro trends and intricate short-term fluctuations in the market.

A.2 ARIMA-SARIMA-GRU

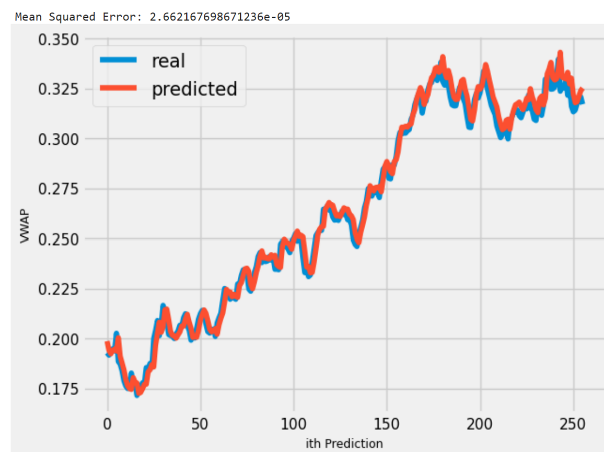

Fig. 4. Performance of the ARIMA-SARIMA-GRU on test dataset

The graph presented provided an insightful comparison between the actual stock prices from the test dataset and the predictions formed by the ensemble ARIMA-SARIMA-GRU model. Upon meticulous examination, the ARIMA-SARIMA-GRU model exhibited outstanding predictive performance among all the other models discussed in the paper. The close alignment of the prediction data with the original data implied that the model had effectively learned both the overall trend of the stock market and a substantial part of its short-term volatility. The Mean Squared Error (MSE) for the ARIMA-SARIMA-GRU model was calculated as an extraordinarily low 2.6621 x 10^{-5}, implying a superior level of predictive accuracy. This exceedingly low error signified that the ensemble models's predictions aligned very closely with the actual stock prices. The ARIMA-SARIMA-GRU model demonstrated a significant degree of proficiency in predicting stock prices, aptly modeling both the macroscopic trends and microscopic fluctuations in the stock market.

B. Individual models

B.1 ARIMA

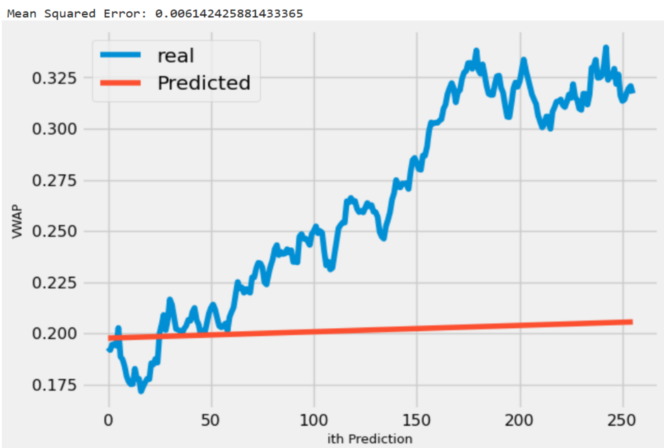

Fig. 5. Comparison of ARIMA model predictions with actual values

The graph presented showed the actual stock prices from the test dataset against the predictions made by the ARIMA model. Upon close inspection, it was clear that while the ARIMA model captured a general increasing trend in the stock prices, it struggled to emulate the intricate fluctuations present in the original data. The fact that the line of the ARIMA model was smoother and less volatile indicated a certain degree of oversimplification, revealing its limitations in capturing sudden and sharp changes in stock prices. The mean squared error for the ARIMA model was reported as 0.00614. In conclusion, the ARIMA model exhibited some strengths in predicting the overarching trend in stock prices but fell short in accurately modeling the market's volatility.

B.2 SARIMA

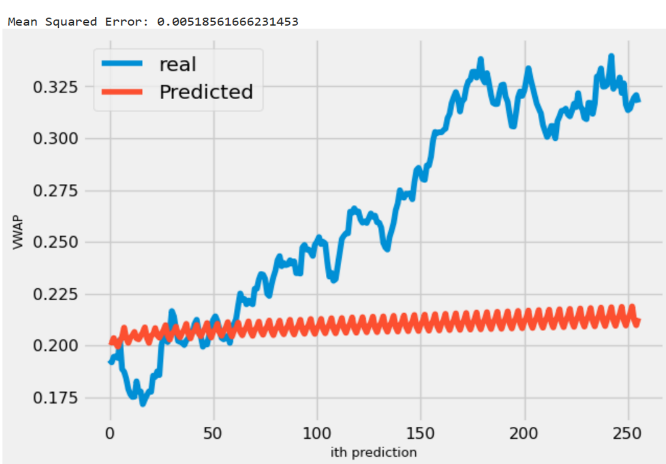

Fig. 6. Comparison of SARIMA model predictions with actual values

The graph showcased a side-by-side comparison of the predictive performance of the SARIMA model versus the actual stock prices from the test dataset. The actual stock prices were depicted as a line with numerous peaks and troughs, reflecting the inherent volatility of the stock market. On the contrary, the graph of the SARIMA model exhibited recurring patterns at regular intervals. The mean squared error for the SARIMA model was calculated as 0.00518. Despite being marginally lower than the ARIMA model, suggesting a slightly improved predictive accuracy, the model still exhibited noticeable discrepancies when it came to capturing the stock market's volatility. This showed that there is a scope of enhancement in its capability to predict with precision.

B.3 LSTM

The graph offered a comparative analysis of the actual stock prices from the test dataset versus the forecasts produced by the Long Short-Term Memory (LSTM) model. The LSTM model's predictions followed the general path of the original values but exhibited less variability, indicating the model's proficiency in capturing the overarching trend but somewhat lacking in tracing the fluctuations. The Mean Squared Error (MSE) for the LSTM model was calculated as 0.00023. Upon closer inspection, it was clear that while the LSTM model was successful in predicting the general trend of the stock prices, it struggled to fully reproduce the volatility evident in the actual data. This suggested that the LSTM model, though competent in grasping longer-term market movements, showed room for improvement when it came to accurately predicting short-term fluctuations.

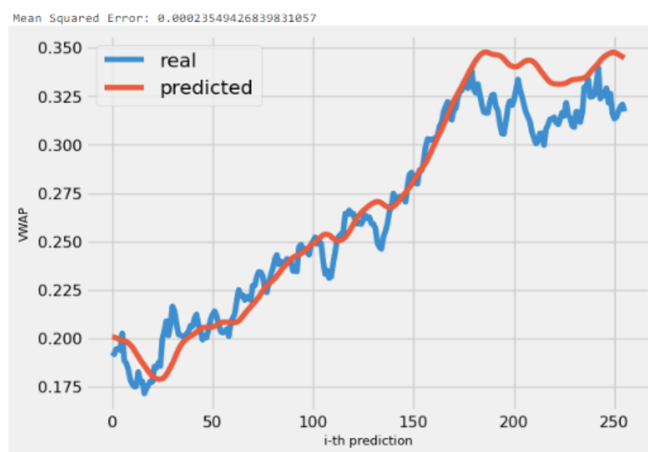

Fig. 7. Comparison of LSTM model predictions with actual values

B.4 GRU

The graph presented provided a visual comparison between the actual stock prices from the test dataset and the predictions yielded by the Gated Recurrent Unit (GRU) model. The GRU model's predictions made an earnest effort to follow and align with the original values, signifying a degree of success in modeling the stock market's complex dynamics. The Mean

Squared Error (MSE) for the GRU model was notably low at 0.00019, indicating a high degree of predictive accuracy. However, minor discrepancies were observable during periods of intense volatility. In conclusion, the GRU model showed strong predictive capabilities, successfully capturing both the broader trends and a significant amount of the market's volatility. However, the visualization suggested that there is room for further refinement. The challenge, therefore, remained to develop an even more robust model that could accurately predict the entire breadth of stock market behaviors.

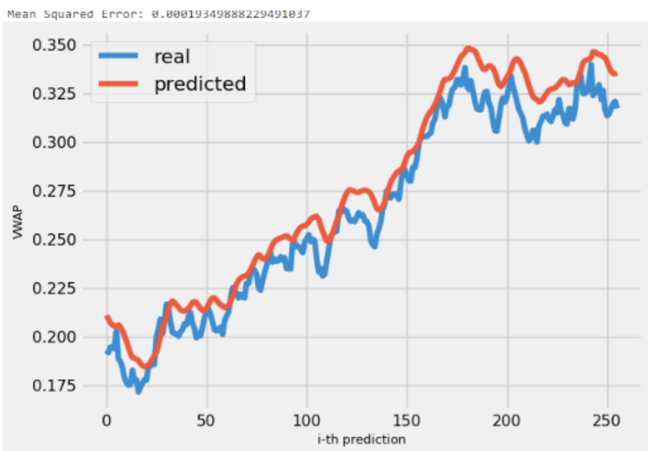

Fig. 8. Comparison of GRU model predictions with actual values

TABLE I
MSE VALUES - COMPARISON

Model	*MSE*
ARIMA	0.00614
SARIMA	0.00518
LSTM	0.00023
GRU	0.00019
ARIMA-SARIMA-LSTM	4.4316×10^{-5}
ARIMA-SARIMA-GRU	2.6621×10^{-5}

Models and their mean squared error values

In conclusion, through the detailed analysis, it has been established that ensemble models, ARIMA-SARIMA-LSTM and ARIMA-SARIMA-GRU outperformed individual models, ARIMA, SARIMA, LSTM and GRU in predicting stock market prices using the TITAN dataset. The ARIMA-SARIMA-GRU model has achieved the highest predictive accuracy, as evidenced by the lowest Mean Squared Error (MSE) of 2.6621 x 10^{-5}. Not far behind in performance was the ARIMA-SARIMA-LSTM model, which also exhibited significant predictive power, with a Mean Squared Error (MSE) of 4.4316 x 10^{-5} demonstrating that both the ensemble models significantly outperformed the individual models in predicting stock market prices.

C. Stock Value Forecasting

Leveraging the robust performance of the ARIMA-SARIMA-LSTM and ARIMA-SARIMA-GRU ensemble models, we tasked the ensemble models with predicting the subsequent day's stock value. If today's date was chosen, the models provided a forecast for tomorrow's stock value based on the available data up until today. The goal was to forecast the TITAN stock value for the next day, June 9, 2023, based on the data available up until June 8, 2023. The models predicted the values to be 2884.6873 and 2887.229 for the ARIMA-SARIMA-LSTM and ARIMA-SARIMA-GRU, respectively. Remarkably, when cross-referencing the predicted values with the actual value of 2889.14 obtained from the NSE website for the specific date, the predictions proved to be highly accurate with a mere difference of 4.4527 for ARIMA-SARIMA-LSTM and 1.911 for ARIMA-SARIMA-GRU model from the actual data. This clearly underscored the high level of accuracy that these ensemble models achieved in their market price forecasts, thereby highlighting their efficacy and potential as powerful predictive tools.

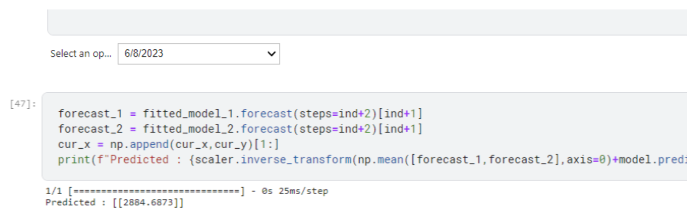

Fig. 9. Forecast of TITAN stock value for June 9, 2023, by proposed ARIMA-SARIMA-LSTM model using input data from the past year until June 8, 2023.

Fig. 10. Forecast of TITAN stock value for June 9, 2023, by proposed ARIMA-SARIMA-GRU model using input data from the past year until June 8, 2023.

TABLE II
FORECASTED VALUES BY ENSEMBLED MODELS

Model	*Actual*	*Predicted*	*Difference*
ARIMA-SARIMA-GRU	2889.14	2887.229	1.911
ARIMA-SARIMA-LSTM	2889.14	2884.6873	4.4527

Actual vs. Predicted Stock Values : June 9,2023

VI. Conclusion

This study substantiates the potential of ensemble models in the arena of stock market price prediction. By combining the distinct strengths of multiple models, we were able to generate predictions that surpassed the accuracy achievable by any of the individual model discussed here. The use of the ARIMA and SARIMA models, known for their prowess in handling trends, seasonality, and non-stationarity, in combination with LSTM and GRU neural networks, recognized for their proficiency in deciphering non-linear patterns and long-term dependencies, proved to be highly effective.

Through our research, we successfully constructed two ensemble models: ARIMA-SARIMA-LSTM and ARIMA-SARIMA-GRU, both of which showcased robust performance and outperformed individual models in predicting future stock prices. This outcome validated our hypothesis regarding the viability and superiority of the ensemble approach, which harnesses the strengths of different models while counteracting their individual weaknesses.

VII. Future works

Incorporating a wider range of parameters, such as market sentiment, news sentiment, and economic indicators, through feature engineering has the potential to redefine the granularity and accuracy of forecasts. Such an approach not only promises to enhance model effectiveness but also paves the way for more comprehensive financial forecasting models. An efficiency-optimized refinement of our ensemble models could significantly push the boundaries of real-time stock price forecasting, a development with profound implications for algorithmic trading. Envisaging the seamless integration of these advanced models into existing financial platforms, we anticipate a transformative shift in user experience and utility. While our current model focuses on near-term stock price forecasting, its extension to a broader forecast range promises models that can provide strategic market insights spanning days or even weeks, potentially ushering the field into a new era of predictive accuracy and reliability.

References

[1] Jisha. R. C., A. Pushpan, A. A. Suryawanshi, H. N., and M. Manu, Analysis of Correlation and Association Mining-based Networks using different Centrality Measures – Indian Stock Market, Grenze International Journal of Engineering and Technology, vol. 9, no. 1, pp. 38-XX, Jan. 2023, Art. no. 01.GIJET.9.1.38.

[2] Ariyo, A.A., Adewumi, A.O., Ayo, C.K. (2014). Stock Price Prediction Using the ARIMA Model. 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, 106-112.

[3] Poongodi, M., V. Vijayakumar, and Naveen Chilamkurti. Bitcoin price prediction using ARIMA model." International Journal of Internet Technology and Secured Transactions 10.4 (2020): 396-406.

[4] Fattah, J., Ezzine, L., Aman, Z., El Moussami, H., Lachhab, A. (2018). Forecasting of demand using ARIMA model. International Journal of Engineering Business Management, 10, 1847979018808673.

[5] Divisekara, R. W., Jayasinghe, G. J. M. S. R., Kumari, K. W. S. N. (2020). Forecasting the red lentils commodity market price using SARIMA models. SN Business Economics, 1(1), 20.

[6] Emmanuel Dave, Albert Leonardo, Marethia Jeanice, Novita Hanafiah, Forecasting Indonesia Exports using a Hybrid Model ARIMA-LSTM, Procedia Computer Science, Volume 179, 2021, Pages 480-487, ISSN 1877-0509

[7] Li, Guoqiang Yang, Ning. (2022). A Hybrid SARIMA-LSTM Model for Air Temperature Forecasting. Advanced Theory and Simulations. 10.1002/adts.202200502.

[8] U. M. Sirisha, M. C. Belavagi and G. Attigeri, "Profit Prediction Using ARIMA, SARIMA and LSTM Models in Time Series Forecasting: A Comparison," in IEEE Access, vol. 10, pp. 124715-124727, 2022, doi: 10.1109/ACCESS.2022.3224938.

[9] Y. Yu, X. Si, C. Hu and J. Zhang, A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures in Neural Computation, vol. 31, no. 7, pp. 1235-1270, July 2019.

[10] S. Selvin, Vinayakumar, R., Dr. E. A. Gopalakrishnan, Menon, V. K., and Dr. Soman K. P., "Stock Price Prediction using LSTM, RNN and CNN-sliding Window Model", in 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2017.

[11] A. Jayanth Balaji, D.S. Harish Ram, Binoy B. Nair,Applicability of Deep Learning Models for Stock Price Forecasting An Empirical Study on BANKEX Data, Procedia Computer Science,Volume 143,2018,Pages 947-953,ISSN 1877-0509.

[12] K. Chen, Y. Zhou and F. Dai, "A LSTM-based method for stock returns prediction: A case study of China stock market," 2015 IEEE International Conference on Big Data (Big Data), Santa Clara, CA, USA, 2015, pp. 2823-2824, doi: 10.1109/BigData.2015.7364089.

[13] ArunKumar, K.E., Kalaga, D.V., Mohan Sai Kumar, C., Kawaji, M., Brenza, T.M. (2022). Comparative analysis of Gated Recurrent Units (GRU), long Short-Term memory (LSTM) cells, autoregressive Integrated moving average (ARIMA), seasonal autoregressive Integrated moving average (SARIMA) for forecasting COVID-19 trends. Alexandria Engineering Journal, 61, 7585 - 7603.

[14] Farah Shahid, Aneela Zameer, Muhammad Muneeb, Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM, Chaos, Solitons Fractals, Volume 140, 2020, 110212, ISSN 0960-0779.

[15] R. Jaiswal and B. Singh, A Hybrid Convolutional Recurrent (CNN-GRU) Model for Stock Price Prediction, 2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT), Indore, India, 2022, pp. 299-304, doi: 10.1109/CSNT54456.2022.9787651.

[16] L. S. Chong, K. M. Lim and C. P. Lee, Stock Market Prediction using Ensemble of Deep Neural Networks, 2020 IEEE 2nd International Conference on Artificial Intelligence in Engineering and Technology (IICAIET), Kota Kinabalu, Malaysia, 2020, pp. 1-5, doi: 10.1109/IICAIET49801.2020.9257864.

[17] Dr. Binoy B. Nair, Forecasting short-term stock prices using sentiment analysis and Artificial Neural Networks, Journal of Chemical and Pharmaceutical Sciences , vol. 9, no. 1, pp. 533 - 536, 2016.

[18] Deepthi P. K, Dr. E. A. Gopalakrishnan, Vijay Krishna Menon and Soman K. P, Stock price prediction using dynamic mode decomposition, International Conference on Advances in Computing, Communications and Informatics. September 13-16, 2017, Manipal, India.

[19] Hiransha. M, Dr. E. A. Gopalakrishnan, Vijay Krishna Menon, and Dr. Soman K. P, NSE Stock Market Prediction Using Deep-Learning Models, in Procedia Computer Science, 2018, vol. 132, pp. 1351 - 1362.

[20] Unnithan, N. A., Gopalakrishnan, E. A., Vijay, K. M. and Soman, K. P. (2019), "A data-driven model approach for day wise stock prediction" in "Emerging research in electronics, computer science and technology", 149 – 158, Lecture Notes in Electrical Engineering,Springer,2019.

[21] Sreelekshmy, S., Vinayakumar, R., Gopalakrishnan, E. A., Vijay, K. M. and Soman, K. P. "Stock price prediction using LSTM, RNN and CNN-sliding window model". International Conference on Advances in Computing, Communications and Informatics, September 13-16, 2017,Manipal,India.

[22] Dr. Binoy B. Nair and Mohandas, V. P., "Artificial intelligence applications in financial forecasting-a survey and some empirical results", Intelligent Decision Technologies, vol. 9, 99-140,2015.