IJIRIS: International Journal of Innovative Research in Information Security
Volume 09, Issue 03, June 2023
E-ISSN: 2349-7017
P-ISSN: 2349-7009
https://www.ijiris.com/archives

Stock Market Price Prediction Using Machine Learning and Deep Learning

Asha Kumari A.
Assistance Professor, Department of Computer Science and Engineering,
The Oxford College of Engineering, Bengaluru – 560068
asha.hebbar@gmail.com

Publication History
Manuscript Reference No: IJIRIS/RS/Vol.09/Issue03/JNIS10103
Research Article | Open Access | Double-blind Peer-reviewed |Article ID: IJIRIS/RS/Vol.09/Issue03/JNIS10103
Received: 02, June 2023 | Accepted: 20, June 2023 | Published: 23, June 2023
Volume 2023 | Article ID JNIS10103 http://www.ijiris.com/volumes/Vol09/iss-03/24.JNIS10103.pdf
ArticleCitation: Asha(2023).Stock Market price prediction using maching kearning and deep Learning. International Journal of Innovative Research in Information Security (IJIRIS), 10, 189-194
doi: https://doi.org/10.26562/ijiris.2023.v0903.24
BibTex key: Asha2023Stick

Copyright: ©2023 This is an open access article distributed under the terms of the Creative Commons Attribution License; which Permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

ABSTRACT: This project aims to address the long-standing challenge of stock market prediction by leveraging the power of machine learning. The stock market is influenced by a multitude of factors, making it difficult for investors to accurately predict trends and make informed decisions. To mitigate this risk, we have chosen to analyse NSE real-time series data using the ARIMA model, a popular forecasting method. Our ultimate goal is to develop a user-friendly web application using Flask, which will allow investors to select a specific firm and receive a forecast for its future stock price. This application will utilize web scraping techniques to retrieve real-time data from the NSE, and the ARIMA model will be applied to predict future outcomes. By combining these powerful tools, we hope to significantly reduce the potential danger associated with stock market prediction and enable investors to make more informed decisions.
Key Words: ARIMA(Auto Regressive Integrated Moving Average), Flask framework,Web scraping.

INTRODUCTION:

Stock market prediction is a challenging task for statisticians and financial professionals, typically involving buying equities with high potential for growth and selling those with high potential for decline. Two main methods for predicting the stock market are fundamental analysis, which focuses on a company's strategy and fundamental data, and technical analysis, which examines prior stock prices and values. As machine learning techniques advance, data scientists are increasingly using these methods to improve the efficiency and precision of prediction models, with the ultimate goal of creating models that can accurately forecast stock market movements. In this study, we compare the accuracy of nine machine learning models for predicting stock market movements, including Decision Tree, Random Forest, Adaboost, XGBoost, SVC, Naive Bayes model, KNN, Logistic Regression, and ARIMA. We feed these models with ten technical indicators and use two different input methods to evaluate the impact of pre-processing. The first approach uses continuous data such as open, close, high, and low values, while the second involves a pre-processing procedure to convert continuous data into binary data. We evaluate the performance of each model using three classification metrics and present the optimal tuning parameter for each model except Naive Bayes and Logistic Regression. Each technical indicator has a distinct potential for upward or downward movement based on the underlying characteristics of the market, and our study provides insights into the best machine learning models and technical indicators for accurate stockmarket prediction.

RELATED WORKS:

HaiderMaqsood, IrfanMehmood, MuazzamMaqsood, MuhammadYasirdSitara. AfzalcFarhan Aadilc Mahmoud and Mohamed SelimeKhan Muhammad **[1]** proposed a hybrid random forest with a linear model (HRFLM) methodology was published in 2019. It used sentiment analysis on Twitter data to forecast stock exchange using linear regression, support vector regression, and deep learning. They evaluated the system's performance using RMSE and MAE. The approach improves accuracy but has long execution time and only suitable for binary classification.

WenLong, ZhichenLu and LingxiaoCui **[2]** proposed end-to-end multi-filters neural network (MFNN) published in 2018. It combines convolutional and recurrent neurons for feature extraction on financial time series samples and stock price movement prediction. It outperforms traditional machine learning models, statistical models, and single-structure networks, achieving higher accuracy, profitability, and stability in extreme market prediction and signal-based trading simulation tasks.

SuryodayBasak, SaibalKar, SnehanshuSaha, LuckysonKhaidem and Sudeepa **[3]** proposes a novel approach to predicting the direction of stock market prices published in 2019. It uses tree-based classifiers such as random forests and gradient boosted decision trees. The paper highlights the importance of selecting technical indicators and their use as features to achieve high accuracy in medium to long-term stock price direction prediction. Although the paper achieves high accuracy for long-term predictions, its overall accuracy is still less compared to other approaches.

Amit Kelotra and Prateek Pandey **[4]** proposed stock market prediction system that uses anoptimized Deep-ConvLSTM model trained by the Rider-based monarch butterfly optimization algorithm published in 2022. The paper employs feature selection using Sparse- Fuzzy C-Means clustering to obtain robust features for accurate prediction. The proposed method achieves a minimal mean squared error and root mean squared error, minimizing the error rate in stock price prediction, and helping to earn good profits. However, the method is not suitable for future prediction.

JiaweiLong, ZhaopengChen, WeibingHe, TaiyuWub and JiangtaoRen**[5]** proposed an integrated framework that utilizes deep learning and knowledge graph techniques publishedin 2020. It predicts the stock price trends in the Chinese stock exchange market. The proposed technique forecasts stock price movements with excellent accuracy using desensitised transaction records and open market data. Although the method is fast and accurate, it is only suitable for the Chinese stock market and requires a longer training time.
M. Nabipour, P. Nayyeri, H. Jabani,A. Mosavi,E. Salwana and Shahab S **[6]** proposeda system that concentrates on the future prediction of stock market groups published in 2020. This paper compares various machine learning algorithms including decision tree, bagging, random forest, adaptive boosting (Adaboost), gradient boosting, eXtreme gradient boosting (XGBoost), artificial neural networks (ANN), recurrent neural network (RNN), and long short-term memory (LSTM) for predicting the future values of stock market groups. LSTM provides the most accurate results, while tree-based models show intense competition between Adaboost, Gradient Boosting, and XGBoost.

G. P. Zhang **[7]** proposed model combines the strengths of both models to improve accuracy in linear and nonlinear modelling published in 2022. This paper proposes a hybrid ARIMA and neural network model for time series forecasting. ARIMA and ANNs are often compared in terms of forecasting performance, but their superiority varies. The Experimental results with real data sets show that the combined model is effective in improving forecasting accuracy achieved by either of the models used separately.

P.-F. Pai and C.-S. Lin **[8]** proposed Support vector machines (SVMs) that have been as a promising alternative for solving nonlinear regression problems published in 2022. Traditional linear models such as autoregressive integrated moving average (ARIMA) have limitations in capturing nonlinear patterns, leading to less accurate time series forecasting. To exploit the strengths of both models, a hybrid methodology combining ARIMA and SVMs was proposed for forecasting stock prices. Experimental results with real-world data showed promising improvements in forecasting accuracy compared to using each model separately.

Ju-Jie Wang, Jian-Zhou Wang, Zhe-George Zhang and Shu-Po Guo**[9]** proposed model to forecast the stock market price index was published in 2012.A hybrid approach combining exponential smoothing model (ESM), autoregressive integrated moving average model (ARIMA), and back propagation neural network (BPNN). Genetic algorithms are usedto choose the hybrid model's weight. The Shenzhen Integrated Index and Dow Jones Industrial Average Index are used to evaluate the model. The suggested model beats all conventional models, including the equal weight hybrid model, the random walk model, the ESM, ARIMA, BPNN, and others, according to the results.

A. M. Ashik and K. S. Kannan [10] proposed model by applying the Box-Jenkins method to forecast the future closing price of Nifty 50 using an ARIMA model published in 2017. The National Stock Exchange in India is a widely used and fully automatic trading system. Time series models have gained significant interest in analysing and predicting stock market trends. In this paper, Ashik and Kannan apply the Box-Jenkins method to forecast the future closing price of Nifty 50 using an ARIMA model. The study shows that the energy sector has a high R-squared value of 99% and a tiny mean absolute percentage error of 0.745,making it a low-risk investment with increasing fluctuation trends for upcoming trading days.

PROPOSED WORK:

We propose the concept of time series analysis and forecasting in the perspective of Indian economy. The recent sharp decline in the value of the Indian rupee has made stock market forecasting imperative in order to protect investors' interests. This approach aims to create an effective ARIMA model to forecast the volatility of the Indian stock market. For this study, time series data from the Indian stock market that is available to the public was utilized. The difference between the projected and actual time series, which has been compared, is around 5% mean percentage error for both the Nifty and Sensex combined.

METHODOLOGY:

Auto Regressive Integrated Moving Average is referred to as ARIMA. It belongs to a family of models that may represent a variety of common temporal patterns in time series data.

Components of ARIMA:

- ARIMA is made up of three parts: the autoregressive term (AR), the differentiating term (I), and the moving average term (MA). Let's examine each of these elements individually.
- The term "AR" refers to historical data that is used to predict future values. The 'p' parameter in ARIMA defines the AR term.
- The MA word is used to describe the quantity of historical prediction mistakes that were utilised to estimate future values. The MA term is represented by the 'q' parameter in ARIMA. The ideal 'q' value is found using an ACF plot.
- The number of times the differencing operation must be applied to a series in order for it to become stationary is specified by the order of differencing. It is possible to find the d value and evaluate if the series is stationary by using tests like ADF and KPSS.

Steps of ARIMA

1. Load the data: Load the data into your notebook.
2. Pre-processing data: The input should be univariate, hence drop the other columns
3. Fit Auto ARIMA: Fit the model to the univariate series using Auto ARIMA.
4. Predict values on validation set: Make predictions regarding the validation set'svalues.
5. Calculate RMSE: compare the model's performance between the predicted and actualvalues and Calculate the RMSE.

SYSTEM ARCHITECTURE

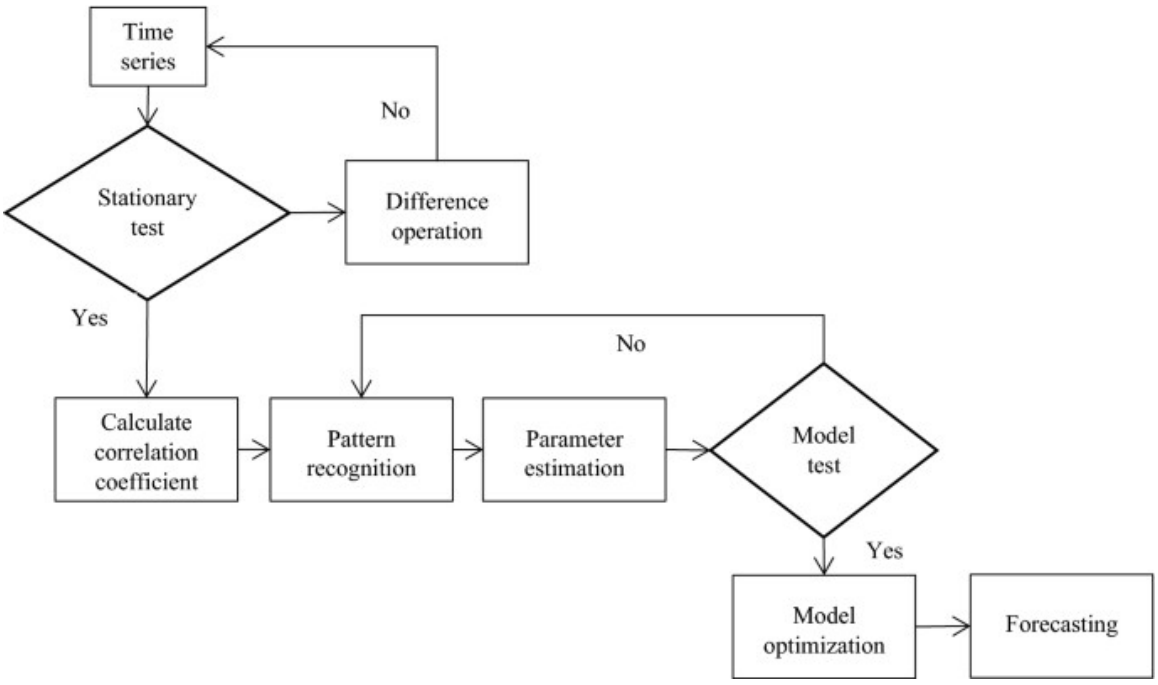

Figure 1: Complete system's architecture flow chartThe system architecture shows the Stock market Price prediction.

1. Verify stationarity: Ensure the time series lacks trend or seasonality components touse ARIMA for forecasting.
2. Apply differencing: If non-stationary, difference the time series until it is stationary,checking for seasonality as well.
3. Reserve validation sample: Set aside a portion of the data to assess model accuracy,via train-test split.
4. Choose AR and MA terms: Utilize ACF and PACF to select AR, MA, or both terms.
5. Construct the model: Create the model and determine the forecast periods, N, basedon your requirements.
6. Validate model: compare the predicted values to the actuals in the validation sample.

MODULES

1. Data Collection:

The first step in any data-driven project is collecting relevant data. In this project, we have collected historical and live data about all stocks from the National Stock Exchange (NSE) of India, including Sensex and Nifty indices. The data includes information about opening price,closing price, highest price, lowest price, adjusted closing price, and volume against each date. The data has been divided into two parts, training and testing, where the training data will be used to formulate the model, and the testing data will be used to validate the model's predictions. we can visualize the data using line charts to identify trends and patterns in the data.

2. Stationarity of a Time Series:

In order to use the ARIMA model, the time series must be stationary, meaning it should lack trend and seasonality components. To check stationarity, we used statistical tests like the Augmented Dickey-Fuller (ADF) test, which determines the presence of a unit root in the time series. If the time series is non-stationary, we used differencing to make it stationary. Computing the differences between successive observations is known as differencing. After differencing, again check the stationarity using statistical tests and visualizations like the autocorrelation function (ACF) and partial autocorrelation function (PACF).

3. Determination of ACF and PACF:

Once we made the time series stationary, next is to determine the ideal parameters for the ARIMA model. The ACF and PACF plots are statistical methods that signify the relation between observations in a time series. The ACF measures the correlation between the series and its lags, while the PACF measures the correlation between the series and its lags after adjusting for the correlations at shorter lags. By analysing these plots, we can identify the order of the autoregressive (AR) and moving average (MA) terms, which are essential parameters in the ARIMA model.

4. Price Forecasting:

The final step is to use the ARIMA model to forecast the stock prices. The ARIMA model uses the historical data to predict future values of the time series. We can set the number of periods to forecast based on your needs, and the model will generate a forecast. We can evaluate the accuracy of the model by comparing the forecasted values with the testing data. To visualize the forecasted prices, wecan use line charts or candlestick charts that show the opening, closing, highest, and lowest prices of the stock for a specific period. Additionally,
we can use statistical metrics like root mean square error (RMSE) and mean absolutepercentage error (MAPE) to evaluate the accuracy of the model's predictions.

RESULT

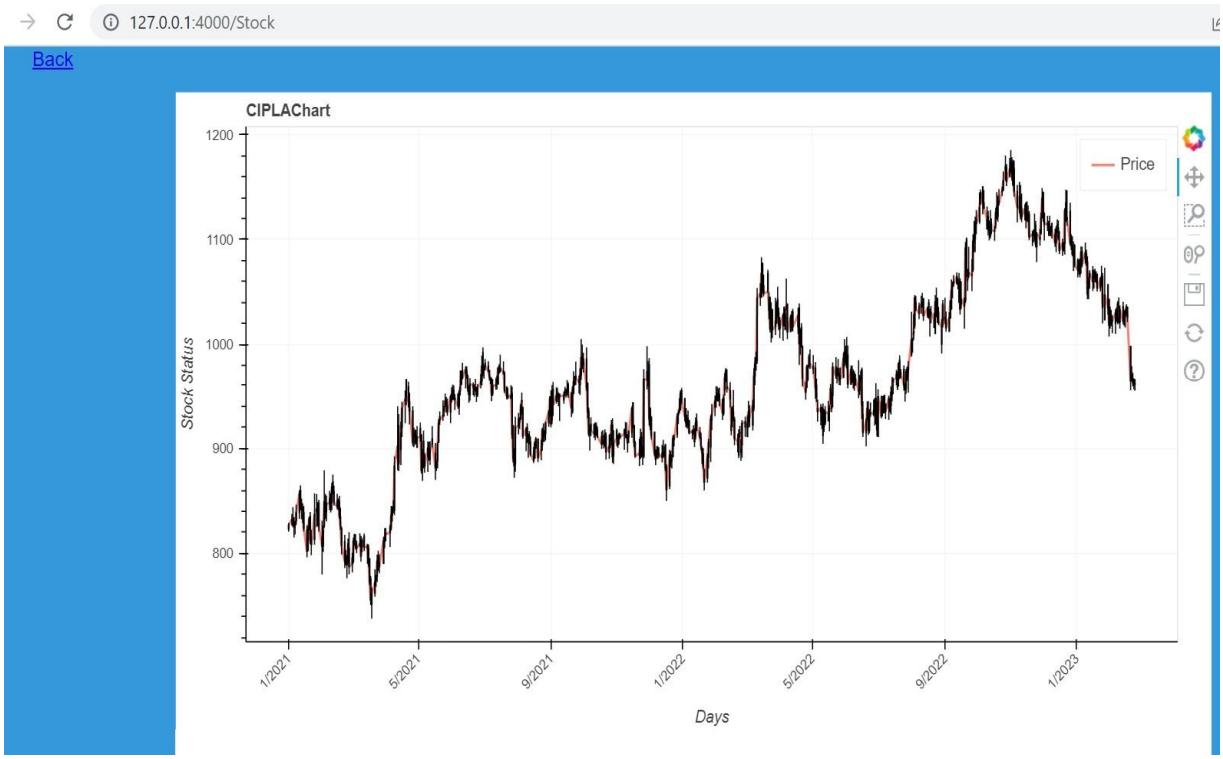

Figure 2: Graph of Current Stock Status for CIPLA Company

IJIRIS: International Journal of Innovative Research in Information Security
Volume 09, Issue 03, June 2023
E-ISSN: 2349-7017
P-ISSN: 2349-7009
https://www.ijiris.com/archives

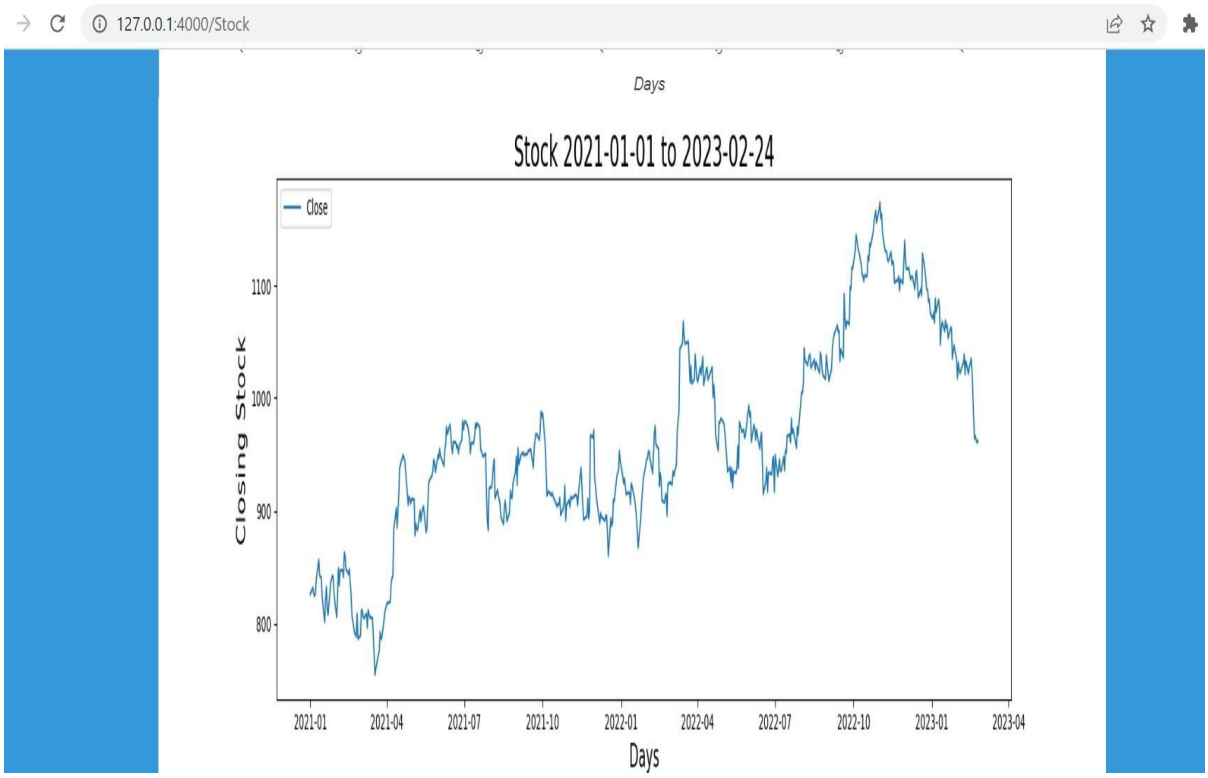

Figure 2: Graph of ClosingStock Prices Over Time

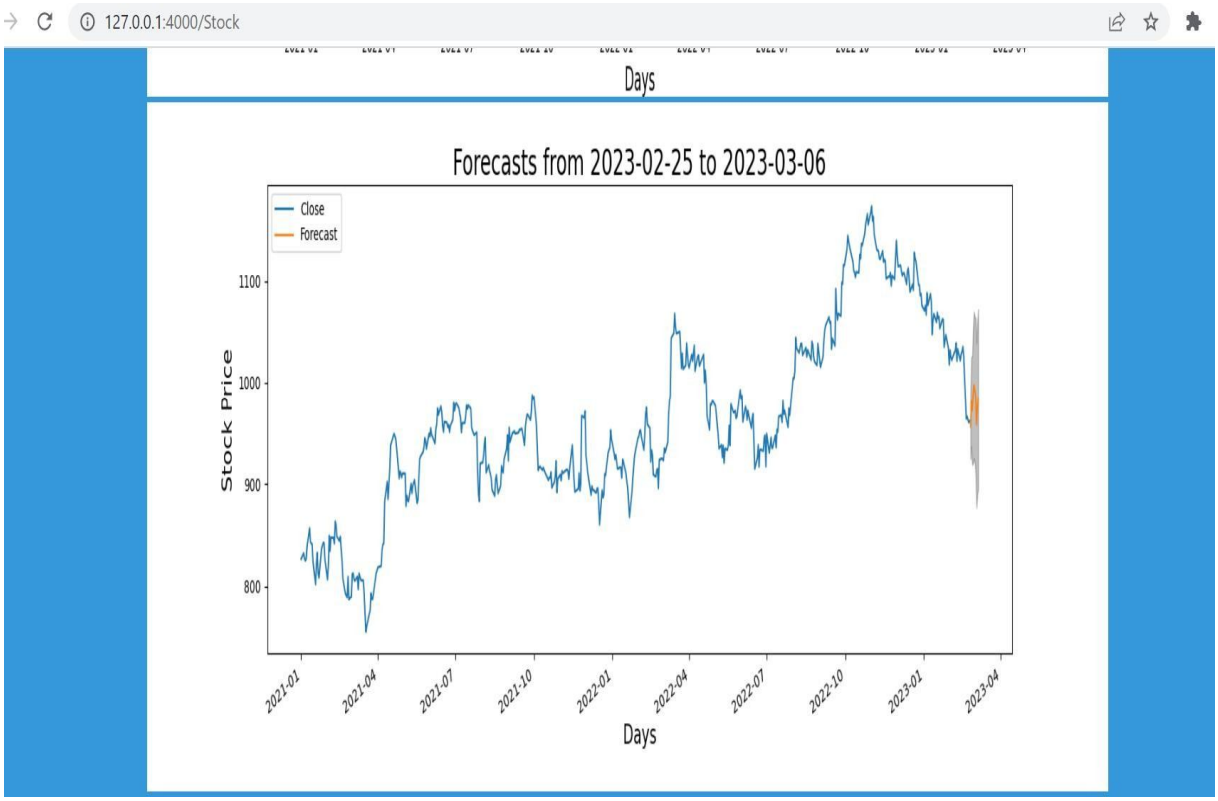

Figure3: Graph thatForecasted Closing Stock for Next 15 Days

CONCLUSION

This project introduces the concept of time series analysis and forecasting in the perspective of Indian economy. The recent sharp decline in the value of the Indian rupee has made stock market forecasting imperative in order to protect investors' interests. In order to forecast the volatility of the Indian stock market, this study aims to develop an effective ARIMA model. The publicly available time series data of Indian stock market has been used for this study. The difference between the projected and actual time series, which has been compared, is around 5% mean percentage error for both the Nifty and Sensex combined. The anticipated time series may be validated using a variety of methods. However, for the purpose of validation in this work, we employed the "ADF test and the L-jung box tests." We argue that the ARIMA technique is suitable for dealing with time series data and, as such, may be highlyhelpful in a variety of real-world issues, such as those in the health, education, finance, and other practical areas for prediction.

REFERENCES:

[1] G. González' Rivera and T. H. Lee's article on "Nonlinear time series in financialforecasting" from the Encyclopedia of Complexity and Systems Science.

[2] P.-F. Pai and C.-S. Lin's research paper on "A hybrid ARIMA and support vector machinesmodel in stock price forecasting" published in Omega in 2005.

[3] J.-J. Wang, J.-Z. Wang, Z.-G. Zhang, and S.-P. Guo,"Stockindexforecasting based on ahybridmodel," Omega, vol. 40, pp. 758–766, Dec. 2012.

[4] L.-Y. Wei's work on "A hybrid model based on ANFIS and adaptive expectation geneticalgorithm to forecast TAIEX" in Econ. Model.

[5] D. C. Montgomery, C. L. Jennings, and M. Kulahci, Introduction to Time Series Analysisand Forecasting. Hoboken, NJ, USA: Wiley, 2015.

[6] R. J. Hyndman. Forecasting Overview. Accessed: Nov. 8, 2009. [Online]. Available: https://robjhyndman.com/papers/forecastingoverview.pdf

[7] Data Collection for Demand Forecasting. Accessed: Jul. 2018. [Online]. Available: http://www.economicsdiscussion.net/demand-forecasting/data-collection-for- demandforecasting /3583

[8] Forecasting. Accessed: Aug. 2018. [Online]. Available: https://en.wikipedia.org/wiki/Forecasting

[9] L. Kurzak, "Importance of forecasting in enterprise management," Adv. Logistic Syst.,vol. 6, no. 1, pp. 173–182, 2012.

[10] A. Khodabakhsh, I. Arí, M. Bakír, and A. O. Ercan's article on "Multivariate sensor data analysis for oil refineries and multi-mode identification of system behavior in real-time" published in IEEE Access in 2018.

[11] An Introduction to Indian Stock Market. Accessed: Jul. 2018. [Online]. Available: https://www.selfgrowth.com/articles/an-introduction-to-theindian-stock-market

[12] An Introduction to Indian Stock Market. Accessed: Jul. 2018. [Online]. Available: https://www.investopedia.com/articles/stocks/09/indian-stockmarket.asp

[13] A. M. Ashik and K. S. Kannan, "Forecasting Indian stock prices using time seriesmodel," in Logistics, Supply Chain and Financial Predictive Analytics. Singapore: Springer, 2019, pp.221–231.

[14] T. Raicharoen, C. Lursinsap, and P. Sanguanbhokai, "Application of critical support vector machine to time series prediction," in Proc. Int. Symp. ISCAS, vol. 5, May 2003, p. 5.

[15] G. P. Zhang's study on "A neural network ensemble method with jittered training data fortime series forecasting" in Inf. Sci.vol.177, no.23, pp.5329–5346, 2007.