Time-series Forecasting of Stock Prices using ARIMA: A Case Study of TESLA and NIO

Sidharth Tiwary (Corresponding Author)
Student of MBA (Business Analytics)
School of Management Studies, University of Hyderabad, Hyderabad
Sidharthtiwary73@gmail.com

Pramod K. Mishra
Asst. Prof. of Operations Management and Decision Sciences
School of Management Studies, University of Hyderabad, Hyderabad
Pramod.mishra@uohyd.ac.in

Abstract

There has been a rapid increase in the demand for electric vehicles across the globe and so are their stock prices. The companies, like TESLA and NIO, have seen a rise and fall in their stock prices over a period of time. Various models have been used by researchers/analysts in this field for the stock price prediction using econometrics models, deep learning techniques using LSTM, RNN, etc. The purpose of this paper is to predict the stock prices of these two electric vehicles (along with their stock closing prices). The econometric model, ARIMA (p, d, q) in particular, has been fitted to predict the stock prices of electric vehicles. The ARIMA (p, d, q) model helps in forecasting by converting the non-stationary data into stationary one using the differencing technique. Further, with the help of the ML algorithms, the model appropriately uses the data (training data) and then validates (testing data) in a fixed proportion. In this paper, we will predict the stock prices of electric vehicles by extrapolating the data to a future time period and then compare the forecasting accuracy. In terms of the managerial implications, the prediction is expected to help the case companies for better optional planning and execution.

1. Introduction

The rise in the consumption of fossil fuels has caused several environmental concerns, causing temperature rise, melting of icebergs at both the poles, rising sea levels, etc. Usage of fossil fuels is limited and constrained but there is unlimited demand for them. In this scenario, many automobile companies have started producing Electric vehicles replacing petrol and diesel vehicles. There has been a surge in the demand for EVs (electric vehicles). TESLA has become one of the largest EV-producing companies in the world causing a surge in its demand and stock. NIO a China-based EV maker has also risen as one of the competitors of TESLA in recent days. So, predicting the stock price of these companies will fetch numerous interesting outputs for the stock traders. Because of the uncertain nature of the stock market, predicting stock prices become difficult. To minimize the investment risk and make a profit, several forecasting techniques are used. Deep learning methods like long short-term memory (LSTM) is used to learn patterns from data and infer solutions from unknown data (Ghosh et al., 2004). For this paper Econometrics models like ARIMA has been used. ARIMA models are some of the widely used techniques in forecasting (Box and Jenkins., 1970) and have come well before machine learning-based models like LSTM. For short-term prediction of financial time series data, ARIMA modeling is efficient (Chatfield, 2000). It is an open model, it helps the researchers in understanding the nature of the dataset at each stage of the processing (Jagwani et al., 2018). This paper attempts to predict Closing stock prices using ARIMA, to analyze the prediction with the actual value. The datasets consist of 7 attributes, out of which we will be interested in predicting the closing price of the stocks as it defines the company's trading performance on day to day basis whether the stock prices have gone up (Mahadik et al.,2021). TESLA and NIO stocks data from 10-12-2018 to 03-06-2022 have been taken for analysis and have been obtained from Yahoo finance. The total number of observations during this period is 878. Further, we split the data, training 80% and testing 20% into two parts. Training data helps in model building and then with the help of testing data we predict whether it fits the data for accuracy. While this section has introduced the paper, section II will briefly discuss the literature, section III is based on the methodology and the data pre-processing, section IV will discuss the result of the analysis. In the section V we will conclude with direction for future research.

2. Literature Review

The researchers have been always interested to find an efficient model to forecast stock market trends. The paper presented shows how the ARIMA model helps the forecaster in predicting the stock prices and has proven to be a valuable model for investors too (Mahadik et al., 2021). A statistical model like ARIMA is widely used for predicting stock price. ARIMA has been used to make various predictions including agri-food prices among others and has given appropriate output to infer how the model deals with the trend and seasonal fluctuation with or without independent variables (Mishra, 2021). Before applying the fitted ARIMA model, ADF (Augmented Dickey Fuller, Dickey and Fuller (1979)) and KPSS (Kwiatkowski–Phillips–Schmidt–Shin) tests have been performed to test the stationarity issues in the dataset. Further, heteroskedasticity and normality checks have been performed using skewness and kurtosis values along with Jarque - Bera tests (Jarque,and Bera, 1980).To analyze the complete data lagged values have been used. Then we present a fitted ARIMA model to predict the stock prices of both the makes. The statistical model, that is identified using ARIMA to forecast the stock prices, was assessed by MSE, RMSE, MAPE etc. In addition to the basic forecast accuracy parameters, statistics like Akaike Information Criteria (AIC), Log-likelihood (LL) and Bayesian Information Criteria (BIC) are normally taken into consideration in forecasting literature (Mishra, 2021).

According to Box and Jenkins (1970), the ARIMA (Auto-regressive integrated moving average) model examines time series observations to make the predictions. It refers to three terms p, d, and q. It may be noted that p, q and d are, respectively, indicate auto-regressive, moving average and differencing terms. The flow chart, as suggested by Box and Jenkins (1970) has been presented in Figure 1.

ARIMA consists of three parameters p, d, and q where p refers to the auto-regressive i.e. AR(p), "I" refers to the integrated i.e. differencing, and q refers to the moving average i.e. MA(q). Three steps iterative procedure is used to build an ARIMA model (Montgomery et al. 2015):

i. Analyzing the historical data, an ARIMA modelfitted.
ii. Estimating the unknown parameters (p, d, q) of the model.
iii. Diagnostic checks are performed to determine the adequacy of the model.

If the specified model is adequate and giving good results in terms of accuracy the model is reported and predictions have been generated (Figure 1 and Figure 2).

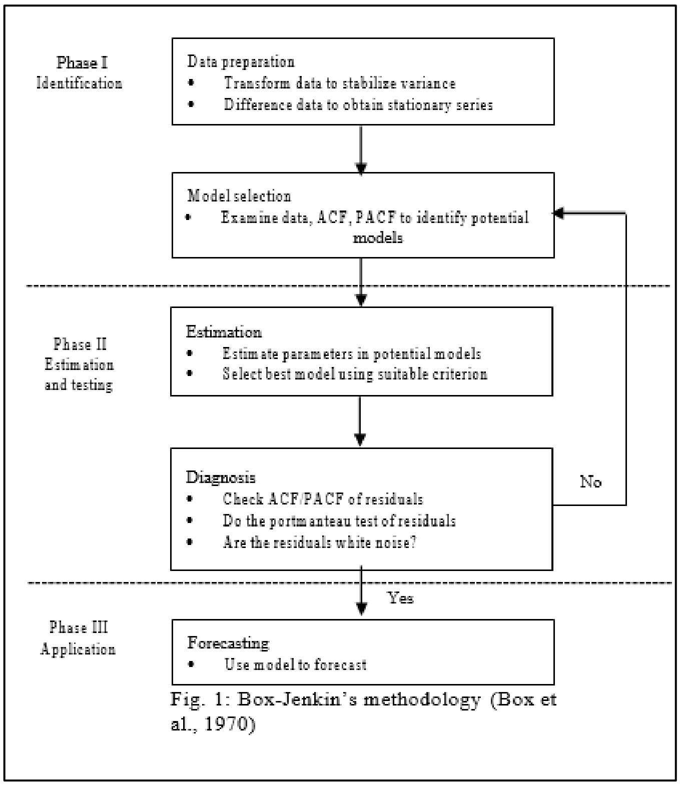

Figure 1. Box and Jenkins methodology (1970)

3. Methodology and the Data Pre-processing

The time-series data were collected from Yahoo finance which had 878 values of TESLA and NIO stock details from 10-12-2018 to 03-06-2022. Data from December 2018 to 2022 were used because of the rise in the sale of EVs (refer to the Bloomberg report for more information https://about.bnef.com/electric-vehicle-outlook/). ARIMA, a supervised machine learning model, used to forecast the stock prices. Using python's several packages, we have done the procedures like importing data from CSV format to data frame and building the model to forecast the Closing stock prices for both the companies. The readers of the paper can use (Mishra, 2021) for more information on forecasting methods and models. The ARIMA model has been fitted twice - once with auto–ARIMA and then by manually setting the hyperparameters (p, d, q) based on the accuracy levels. The paper compares the findings of all the models based on the accuracy to forecast stock prices.

The ARIMA Models (Box and Jenkins methodology)

According to the models presented by Box and Jenkins (1970) (as said in Makridakis et al., 1998) on the ARIMA framework, we have:

$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \ldots + \beta_p y_{t-p} + \varepsilon_t$ (AR)……………………………….....(1)

$y_t = \beta_0 + \beta_1 e_{t-1} + \beta_2 e_{t-2} + \ldots + \beta_p e_{t-p} + e_t$ (MR)………………………………...(2)

$y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots + \varphi_p y_{t-p} + \varepsilon_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots \theta_q e_{t-q}$ (ARMA)….(3)

$(1-\varphi_1 B)(1-B) y_t = c + (1-\theta_1 B) e_t$ (ARIMA (1,1,1))……………………………….....(4)

Step 1

To observe variance and trend in the data, time series data has been plotted below as seen in Figure 2.

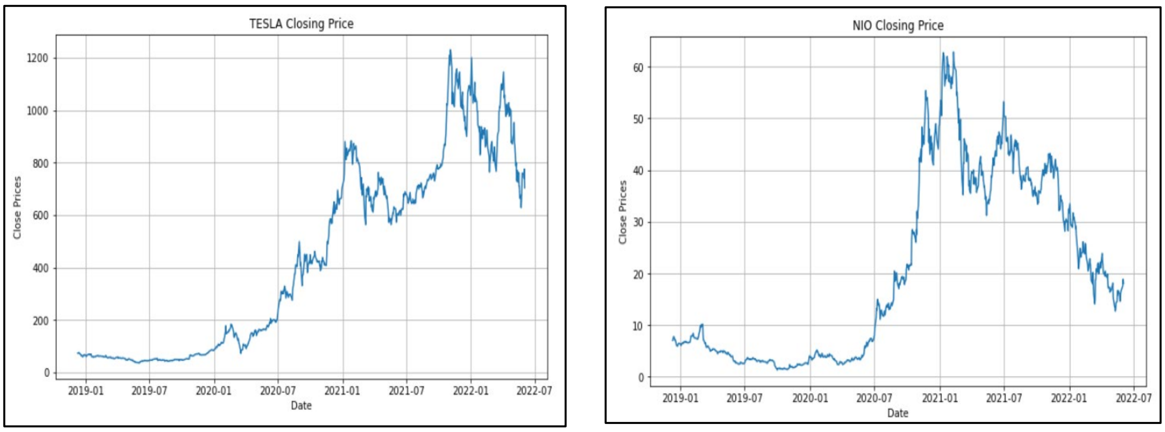

Figure 2. TESLA (1st) and NIO (2nd) closing stock prices

First, both the companies' stock's closing price graphs have been plotted to check the stationarity of the data i.e. whether the data fluctuates with respect to time or not. The graphs given above show fluctuation with respect to time i.e. the mean and standard deviations keep on fluctuating with respect to time, hence we assume that the data is non-stationary. Further, we decompose the data into 4 plots, i.e. plotting raw data, a trend plot to depict an upward or downward trend in data, a seasonal plot to depict the seasonality and a residual plot which shows the noise. In non-stationary data the features like Trends, Cycles, Random walks or a combination of these three are common and need to be addressed. These non-stationary time series processes have been analyzed for applying the correct transformation method ((Pratheeth et al., 2021; Dhadhuk,2021).

a) Pure random walk ($Y_t = Y_{t-1} + \varepsilon_t$): It shows the tendency of fluctuation as a non-mean reverting process i.e. the prices can fluctuate either in a positive or negative direction.
b) Random walk with drift ($Y_t = \alpha + Y_{t-1} + \varepsilon_t$): Along with the random walk, drift fluctuates the prices in an irregular movement i.e. drift has a tendency to move irregularly, not in one direction.
c) Deterministic trend ($Y_t = \alpha + \beta t_t + \varepsilon_t$): Since the trend moves in one direction, which fluctuates the prices around a fixed trend.
d) Random walk with drift and deterministic trend ($Y_t = \alpha + Y_{t-1} + \varepsilon_t + \beta_t$)

Step2

Here, we first check whether the time series data shows stationarity or not, to check this, two tests have been used as **A)** ADF test: It is used to check whether the time series data has unit root i.e. whether the series is grossly under or over-differenced, hence checking the stationarity. Setting up the Null Hypothesis $H_{O:}$ Series has a unit root and alternative hypothesis $H_{1:}$ Series has no unit root. The series is said to be Stationary if the alternative hypothesis is rejected (Dickey, Fuller, 1979). **B)** KPSS test:To test the stationarity of the data, the unit root test is used. Setting up the Null Hypothesis as $H_{O:}$ Series does not have a unit root and Alternative Hypothesis H_1: Series has a unit root. Failing to reject the Null hypothesis suggests series is trend stationary (Vijay, 2021). Further, Normality and Heteroskedasticity have been checked for time series data:

a) **Normality check**: To check the normality of the time series data Jarque-Bera test has been used:
H_O: Skewness and excess Kurtosis, both are equal to zero and p-value > 0.05 i.e. no lack of fit.
H_1: Skewness and excess kurtosis, both are not equal to zero and p-value <=0.05 i.e. lack of fit.

b) **Heteroskedasticity check**: Breusch – Pagan test has been used to check the heteroskedasticity, to fit the residual, the regression model is depicted using OLS. H_O: The model shows constant error variance (p-value > 0.05) i.e. homoskedasticity.
H_1: The model does not show constant error variance (p-value <= 0.05) i.e. heteroskedasticity (Breuschand Pagan, 1979).

Step3:
Autocorrelation and Partial autocorrelation Functions (ACF and PACF): We have used ACF and PACF plots to identify Autoregressive (AR) and Moving average (MA) i.e. p and q orders. ACF plots the coefficients of correlation between the current value and its lagged values. Simply, we can explain that predicting the stock prices' current value depends on all the lagged values in the past. In the ACF plot, the correlation coefficients are represented on the X-axis and the number of lags is represented on the y-axis. PACF is explained using a linear regression to predict y_t from the coefficients of y_{t-1}, y_{t-2}, ...y_{t-p} (Montgomery et al., 2015). In the following analysis ACF and PACF plots, before and after differencing have been presented to assess the parameters of the model. It may be noted that one of the key issues in ARIMA model is that the data should be stationary and to make the data stationary there is a need to use differencing the two consecutive values, normally denoted by an operand delta.

a) **Differencing**: It is used to convert the series from non-stationary to stationary. It rules out the varying mean.
$\Delta_t = y_t - y_{t-1}$; where y_t is the value at time t and y_{t-1} is the lagged value of y_t.

b) **Seasonal differencing**: This method is also used to transform the series from non-stationary to stationary especially when there is seasonality in the data. In this method, we calculate the difference between the observations from the same season to rule out the seasonality from the series.

For TESLA and NIO, lag of 30 i.e. 30 days has been used and demonstrated in the following Figures 3:

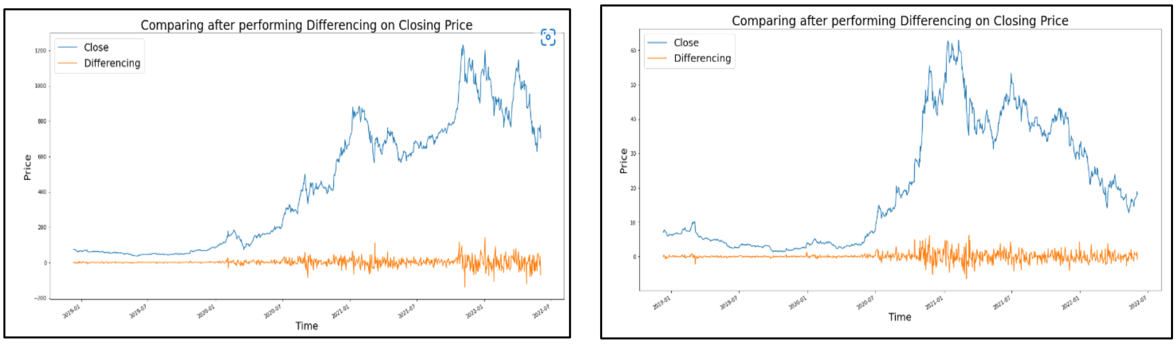

Figure 3. TESLA (1st) and NIO (2nd) closing price (pre and post differencing)

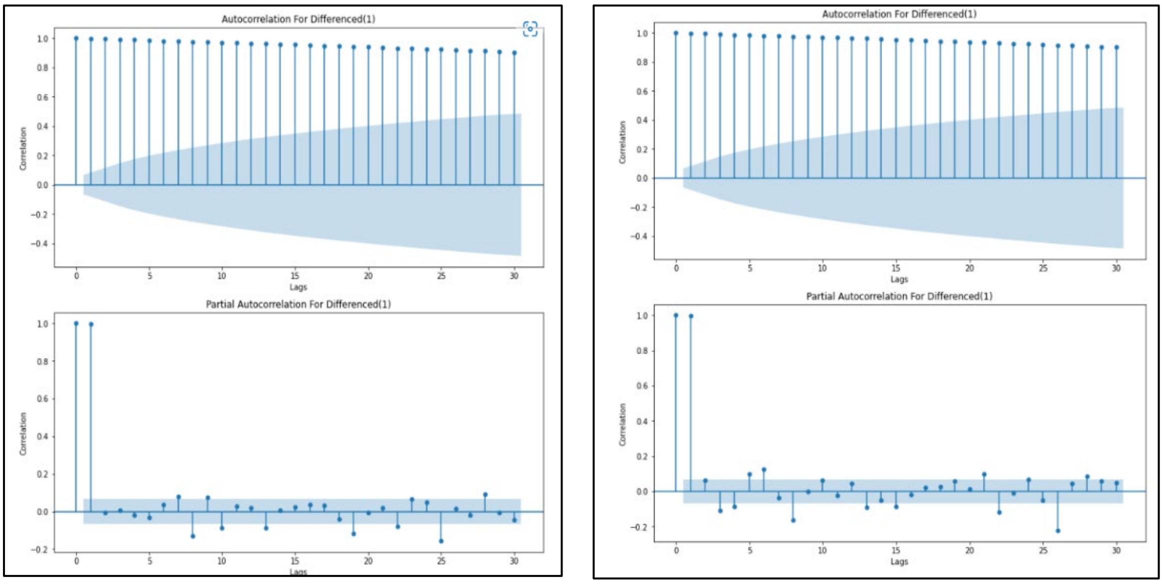

Figure 4. ACF and PACF plots of TESLA (1st) and NIO (2nd)

Hence, after transforming the data into stationary, we again use the ADF test to check whether the test statistics for the closing price, differencing, and seasonal differencing are less than the p-value or not (Figure 4).

Step4:
After transforming the time series data from non-stationary to stationary, we build an ARIMA model and train it on the training data (Saha, 2020). We split the training and test data into 80% and 20% respectively (Mishra, 2021). To choose the ARIMA model's p, q, and d parameters, auto-arima has been used. We have also tuned some ARIMA models after getting the range from ACF and PACF plots i.e. we choose the p and q values through ACF and PACF plots. We then make the model predictions on the test data. We also check for the model accuracy using MSE, RMSE, MAE, and MAPE for both companies. Then we forecast the Closing stock prices on the test dataset within 95% confidence level.

4. Results and Discussions

The model has been designed to predict the closing stock prices for the next 15 days. The ARIMA model forecasts closing stock prices on the test data with a 95% confidence level (Figure 5).

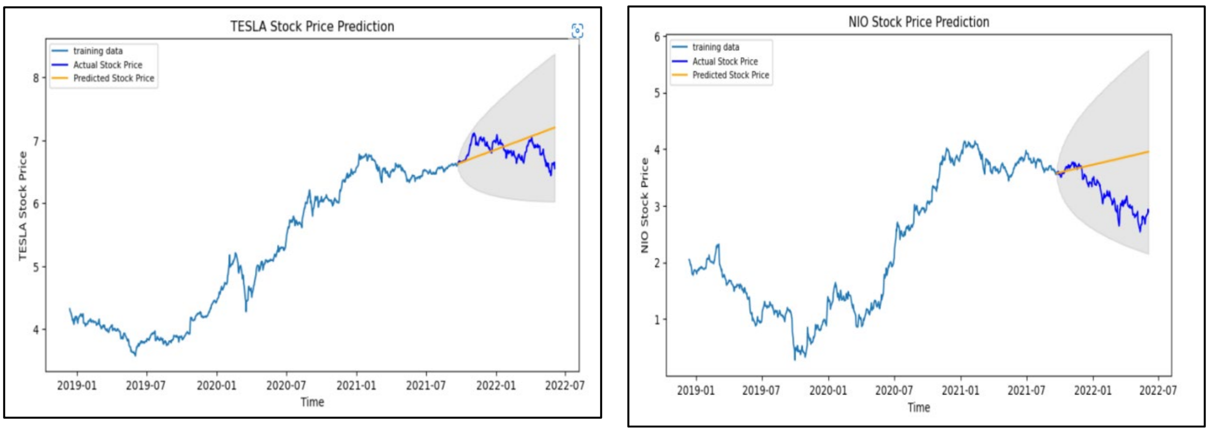

Figure 5. TESLA (1st) and NIO (2nd) closing stock prices

For NIO, the trend goes upward till 01-2021 then starts going downward. For TESLA, the trend keeps on going upward, however, the seasonal plot indicated no seasonal effect for both the companies. For TESLA, with a MAPE of 0.03%, the model is 97.0% accurate and for NIO, with a MAPE of 0.18%, the model is 82.0% accurate in predicting the closing stock prices. The model parameters, accuracy, mean absolute error (MAE), root mean squared error (RMSE) and mean absolute percentage error (MAPE) of the model are shown in Table 1, 2, 3.

Table 1. Coefficients of the model and other key statistics

TESLA models						
Optimum model:	ARIMA(0, 1, 0)					
Log Likelihood	1209.21					
AIC	-2411.92					
BIC	-2402.83					
HQIC	-2408.43					
	coef	std err	z	P>|z|	[0.025	0.975]
--						
const	0.003	0.002	2.031	0.042	0.000	0.007

Note: the model is white noise depicting

Table 2. Coefficients of the model and other key statistics

NIO models						
Optimum model	ARIMA(1, 1, 1)					
Log Likelihood	1003.44					
AIC	-1998.878					
BIC	-1980.685					
HQIC	-1991.844					
	coef	std err	z	P>|z|	[0.025	0.975]
--						
const	0.002	0.003	0.833	0.405	-0.003	0.007
ar.L1.D.Close	0.549	0.234	2.348	0.019**	0.091	1.007
ma.L1.D.Close	-0.459	0.248	-1.852	0.064***	-0.944	0.027
*10%, **5% level of significance						

Table 3. Model accuracy

Statistics	NIO	TESLA
Accuracy	82.0	97.0
MSE	0.46	0.07
MAE	0.53	0.20
RMSE	0.68	0.26
MAPE	0.18	0.03

After observing ACF and PACF plots we set up the range as 1 and 2 for MA and AR respectively i.e. p=1 and q=2. We tune Six ARIMA models for both TESLA and NIO which are ARIMA (1,1,2), ARIMA (0,1,2), ARIMA (0,1,0), ARIMA (1,1,1), ARIMA (1,1,0) and ARIMA (2,1,3). Comparing statistical estimation including their AICs, BICs and log-likelihood, it is found that the ARIMA (0,1,0) for TESLA and ARIMA (1,1,1) for NIO are the best fitted model given the data. The following Closing price predictions have been made for both the companies TESLA and NIO for the year 2022 in the given Table 4.

Table 4. Prediction of closing stock prices

Date	TESLA ($)	NIO ($)
03-06-2022	706.66	18.20

04-06-2022	703.32	18.05
05-06-2022	709.10	18.16
06-06-2022	706.15	18.19
07-06-2022	706.42	18.19
08-06-2022	705.73	17.98
09-06-2022	708.64	18.09
10-06-2022	707.42	18.08
11-06-2022	710.48	18.20
12-06-2022	715.38	18.26
13-06-2022	714.93	18.33
14-06-2022	714.66	18.25
15-06-2022	715.31	18.35
16-06-2022	712.47	18.20
17-06-2022	717.75	18.31

5. Conclusion and direction for future research

This paper attempts to predict the future stock prices of TESLA and NIO using ARIMA. Using ARIMA modeling has been simple and the results have been promising. Further, more advancement can be done using SARIMAX, and various other ML (machine learning) algorithms like logistic regression, KNN, and Decision tree (CART), etc. The stock prices of companies depend upon current events, tweets, political news, and other factors. Hence, we can overcome these limitations by using Sentiment Analysis. Thus, we can develop a model which can be more accurate than the proposed model. Throughout the whole analysis and after forecasting the Closing prices for both TESLA and NIO, we observe that TESLA stock prices have been increasing at a fast rate than the NIO stock prices, whereas NIO has seen a plunge in its stock, which rings an alarm for NIO to do the several modifications required to be done. If TESLA keeps going with this increase in its stock prices, certainly it will give it more edge over its EV rivals like NIO and other companies globally.

References

Box, P. E. G., andJenkins, M. G., Timeseriesanalysis:forecastingand control,SanFrancisco:Holden-Day,1970.

Breusch, T. S., and Pagan, A. R., Simple test for heteroscedasticity and random coefficient variation. *Econometrica, 47*(5), 1287-1294, 1979.

Chatfield, C., Time-series forecasting (1st ed.), Boca Raton, FL, USA:ChapmanandHall/CRC, 2000.

Dhaduk, H. Stock market forecasting using time series analysis with ARIMA model, accessed online at https://www.analyticsvidhya.com/blog/2021/07/stock-market-forecasting-using-time-series-analysis-with-arima-model/#h2, 2022.

Dickey, A. D., andFuller, A. W., Distribution of estimation of auto-regressive time series with a unit root," Journal of the American StatisticalAssociation,vol.74,pp.427-431,1979.

Electric vehicle consumption from Bloomberg accessed at https://about.bnef.com/electric-vehicle-outlook/ 2022.

Ghosh, P., Ariel, N., and Jajati, S. K., Forecasting directional movements of stock prices for intraday trading using LSTM and random forests, arXiv: 10178, 2004.

Jagwani, J., Hardik, S., Gupta, M., and Alka, S., Stock price forecasting using data from Yahoo Finance and analyzing the seasonal and nonseasonal trend, Proceedings of the Second International Conference on Intelligent Computing and Control Systems (ICICCS), IEEE Xplore Compliant Part Number: CFP18K74-ART; ISBN:978-1-5386-2842-3, 2018.

Jarque, C. M., and Bera, A. K., Efficient tests for normality, homoscedasticity and serial independence of regression residuals. Economic Letters, 6(3), 255-259, 1980.

Mahadik, A., Devyani, V., and Amrapali, M., Stock price prediction using LSTM and ARIMA in Proceedings of the Second International Conference on Electronics and Sustainable Communication Systems IEEE Xplore Part Number: CFP21V66-ART; ISBN: 978-1-6654-2867-5, 2021.

Makridakis, S., Wheelwright, C. S., andHyndman, J. R., Forecastingmethodsandapplications(3rded.),USA:JohnWileyandSons,1998.

Mishra, P. K., Predicting agri-food prices with time-series and data-mining based methods, Proceedings of the Fifth International Conference on Intelligent Computing and Control Systems , IEEE Xplore Part Number: CFP21K74-ART; ISBN: 978-0-7381-1327-2., 2021.

Montgomery, D. C., Cheryl, J. L., and Murat, K. Introduction to time series analysis and forecasting, Second edition, New Delhi: John Wiley and Sons, 2015.

Pratheeth, S., and Vishnu, R. P., Stock price prediction using machine learning and deep learning, 2021, IEEE Mysore Sub Section International Conference, 2021.

Saha, A., Time series analysis. Accessed from https://abhinabasaha.in/blog/Time-Series/ on 2022.
Vijay, G. K. Statistical tests to check stationarity in time series – part 1, accessed from https://www.analyticsvidhya.com/blog/2021/06/statistical-tests-to-check-stationarity-in-timeseries-part-1, 2022.

Biographies

Sidharth Kumar Tiwary, an MBA graduate in Business Analytics from the University of Hyderabad, and a graduate in Statistics from Banaras Hindu University (BHU). Have a keen interest in Statistics and Economics and applying these applications in the Data Science field. He has been vigorously working under the guidance of Dr. Pramod Mishra in the operations and Supply chain analytics field.

Pramod K. Mishra, an alumnus of NIT Rourkela and the University of Hyderabad, has been working as an Assistant Professor in the area of Operations Management and Quantitative Techniques in the School of Management Studies, University of Hyderabad. Prior to his appointment at the University of Hyderabad, he has worked as faculty at GITAM School of International Business, GITAM University, Visakhapatnam. Dr. Mishra has worked as Post Doctoral Research Fellow at IIM Bangalore for about three years. He has published several research papers in various national and international journals.