BIG BULL

MINOR PROJECT – 2

Submitted By:-

Ayush Singh Chauhan (20104047)

Surya Pratap Singh (20104049)

Rishabh Srivastava (20104055)

Under the Supervision of:-

Ms. Deepti Singh

Department of CSE/IT

Jaypee Institute Of Information Technology, Noida

May 2023

Table of Content

ACKNOWLEDGEMENT

We would like to place on record our deep sense of gratitude to **Ms Deepti Singh** under whom we have carried out the project work. Her incisive and objective guidance and timely advice encouraged us with constant flow of energy to continue the work.

We also wish to extend our thanks to our friends and other classmates for their insightful comments and constructive suggestions to improve the quality of this project work.

Name of Students

Name : **Ayush Singh Chauhan**

Enrolment No.: **20104047**

Name : **Surya Pratap Singh**

Enrolment No.: **20104049**

Name : **Rishabh Srivastava**

Enrolment No.: **20104055**

DECLARATION

We hereby declare that this submission is our own work and that, to the best of our knowledge and beliefs, it contains no material previously published or written by another person nor material which has been accepted for the award of any other degree or diploma from a university or other institute of higher learning, except where due acknowledgment has been made in the text.

Place : Jaypee Institute Of Information Technology

Date : 12 May, 2023

Name : **Ayush Singh Chauhan**

Enrolment No.: **20104047**

Name : **Surya Pratap Singh**

Enrolment No.: **20104049**

Name : **Rishabh Srivastava**

Enrolment No.: **20104055**

CERTIFICATE

This is to certify that the work titled "**Bulls Eye**" submitted by Ayush Singh Chauhan(20104047), Surya Pratap Singh (20104049) and Rishabh Srivastava (20104055) of B.Tech of Jaypee Institute of Information Technology, Noida has been carried out under my supervision. This work has not been submitted partially or wholly to any other University or Institute for the award of any other degree or diploma.

Signature of Supervisor	……………………....
Name of Supervisor	**Ms. Deepti Singh**
Designation	Assistant Professor (GRADE-II)
Date	12 May 2023

ABSTRACT

Stock price prediction is a classic and important problem in today’s world. Due to recent developments in Stock Market forecasting algorithm, prediction of the future values has become a quite easy task. By using Stock Market analysis, one can easily predict the future closing values of the stock at the end of the day.

This project applies various time bound linear regression algorithms to forecast the upcoming stock prices for easier buying and selling of the stocks and it can also help the various companies to maintain a steady position in the stock market. The objective is to predict the stock prices in order to make more informed and accurate investment decisions. In Stock Market Analysis, we are using the AR, MA, ARMA, ARIMA, SARIMA and LSTM algorithm.

There are two types of stocks. You may know of intraday trading by the commonly used term "day trading." Inter day traders hold securities positions from at least one day to the next and often for several days to weeks or months.

This is important in our case because the previous price of a stock is crucial in predicting its future price. While predicting the actual price of a stock is an uphill climb, we can build a model that will predict whether the price will go up or down. In this way we can predict the stock prices very easily.

Keywords: AR, MA, ARMA, ARIMA, SARIMA and LSTM

LIST OF FIGURES

LIST OF SYMBOLS AND ANCRONYMS

AR	Auto - Regression
MA	Moving Average
ARMA	Autoregressive Moving Average
ARMIA	Auto-Regressive Integrated Moving Average
SARIMA	Seasonal Auto-Regressive Integrated Moving Average
RMSE	Root mean square deviation
LSTM	Long Short-term Memory

CHAPTER 1 INTRODUCTION

Today’s financial market is an ever evolving dynamic and composite system where people can buy and sell currencies, equities and derivatives over virtual platforms supported by brokers. One of the vital elements of a market economy is stock market. The reason behind this is mainly because of the foundation it lays for public listed companies to gain capital via investors, who invest to buy equity in the company.

1.1 STOCK MARKET

Stocks are bought and sold on stock markets, which bring together buyers and sellers of shares in publicly traded companies. Stock markets operate kind of like auctions, with potential buyers naming the highest price they’re willing to pay (“the bid”) and potential sellers naming the lowest price they’re willing to accept (“the ask”).

The stock market allows investors to own shares of public companies through trading either by exchange or over the counter markets. This market has given investors the chance of gaining money and having a prosperous life through investing small initial amounts of money, low risk compared to the risk of opening new business or the need of high salary career. Stock markets are affected by many factors causing the uncertainty and high volatility in the market

1.2 TREND PREDICTION

A correct prediction of stocks can lead to huge profits for the seller and the broker. Recently, it is brought out that prediction is chaotic rather than random, which means it can be predicted by carefully analysing the history of respective stock market. There are many studies on the prediction and analysis of financial time series. In 1970, the Effective Market Hypothesis indicated that the stock price is an immediate reflection of stock market information. Therefore, researchers use traditional statistical methods such as regression methods like exponential average and ARIMA to predict the stock price based on historical prices.

1.3 TIME SERIES FORECASTING

Time series forecasting is an important area of forecasting in which past observations of the same variable are collected and analysed to develop a model describing the underlying relationship. Time-series analysis is an important subset of prediction algorithms and functions. It is regarded as an apt tool for predicting the trends in stock market and logistics. Before making any investment, an investor gathers intel on the past stock trends, periodic changes and various other factors that affect the capital of a company.

1.4 LINEAR REGRESSION

A linear regression model is a statistical model that relates variables in a linear way. This means that the expected value of the response variable is a linear function of the predictor variables. Linear regression is the analysis of two separate variables to define a single relationship and is a useful measure for technical and quantitative analysis in financial markets.

Linear regression could be applied to Stock market dataset to model the relationship between a particular date and the closing price of stock on that date (or any other two values). Once the trader computes the linear regression model, it can then be used to predict the trend followed by that stock based on the original data.

1.5 MOTIVATION FOR WORK

Stock market prediction is the act of trying to determine the future value of a company stock or other financial instrument traded on an exchange. The successful prediction of a stock's future price could yield significant profit. The stock market prediction has extra advantages for novice traders as they are the kind of traders who are more prone to making mistakes and facing severe losses in the market compared to experienced traders. You can better analyze and predict the stock market by gaining a complete understanding of the same.

Motivate to use time-series models rather than analysts’ forecasts stems from recent research that reports time-series predictions to be superior to analysts’ forecasts in predicting earnings for longer periods and for small firms that are hardly followed by financial analysts.

1.6 PROBLEM STATEMENT

Financial analysts investing in stock market usually are not aware of the stock market behaviour. They are facing the problem of trading as they do not properly understand which stocks to buy or which stocks to sell in order to get more profits. In today’s world, all the information pertaining to stock market is available. Analysing all this information individually or manually is tremendously difficult. As such, automation of the process is required. Thus, we tried to find a way to predict the trend followed by a stock to mitigate the risk faced by an individual or a firm while investing in stock market.

CHAPTER 2 LITERATURE SURVEY

There are many kinds of research works in the area of forecasting using time series analysis. Some of the important tasks are mentioned here.

- A study by Prof. Tinni Chaudhari and Dr. Abhijit Pandit researched on the share prices of 6 major companies of India in 2020 for the next 9 months. A significant rise in the share price of TCS was observed in the near future, which calls for special mention. It also depicts the importance of not taking very long-term prediction of such volatile data as several socio-economic factors are involved which play a major role in the rise and fall of market share prices [4].

- Sheikh Mohammad Idrees, M. Afshar Alam and Parul Agarwal in the fields of Time series analysis and Stock market forecasting also focused on various regression models. Their research paper discusses an ARIMA (Auto Regressive Integrated Moving Average) model for prediction of stock market movement. The paper discusses the different steps involved in prediction starting from gathering information to choosing appropriate model to usage and analysis of forecasting models and all the intermediate steps involved. The main datasets used are Nifty and Sensex of Bombay Stock Exchange. This paper introduces the concept of time series analysis and forecasting in the perspective of the Indian economy. This paper tries to build an efficient ARIMA model to predict the Indian stock market volatility [2].

- An Effective Time Series Analysis for Stock Trend Prediction Using ARIMA Model for Nifty Midcap-50 Journal by B. Uma Devi, D. Sundar, and Dr. P. Alli predicts the trend in stock prices which is proposed. This research discovered that the stock values are changes random and the past price values are not dependent on current values. ARIMA model was applied in this research paper and the data has been collected from NSE.com. The historical data for the period of five year since 2007 to 2011 were taken in to account for analysis. The AIC-BIC test criteria is applied against the data represented in the previous to select the best the model. The best model equation is derived for all indices. The MAPE, PMAD and % Error accuracy is applied to determine the discrimination between the actual historical data and the forecast data. This paper inferences a new investment decisions or guidelines based on the minimum error percentage obtained through the above said performance measures. The future forecast of each index for next few years also highlighted in this paper. It is hoped that more innovative approaches will be conducted to bring the hidden information about the stock market [7].

- Paulo Rotela Junior, Fernando Luiz Riêra Salomon and Edson de Oliveira Pamplona focused on ARIMA as a forecasting model to predict the variations in a time series of the Brazilian stock market Index (Ibovespa). In order to compare results with other smoothing models, the parameter of evaluation MAPE (Mean Absolute Percentage Error) was used. Historical data of monthly Bovespa quotations from January 1995 to January 2013 were used. The results showed that the model obtained lower MAPE values, thus indicating greater suitability. This study sought to obtain short-term forecasts for the next month (one step ahead) in order to minimize prediction errors [8].

- A research paper on Stock Market Forecasting Technique using Arima Model by Bijesh Dhyani, Manish Kumar, et at proposed that there are many techniques and analysis technique that can be used with times series data like ARIMA model, exponential Smoothing, Neural Networks or Simple Moving average. However, ARIMA Model is commonly used to understand time series analysis in order to extract meaningful characteristics of the data and help in the prediction of the stock prices. This research paper tells us that a stationary series has constant mean and variance over time, a stationary dataset will allow our model to predict that the mean and variance will be same in future trends. There is mathematical test which can be used to test stationarity in dataset one commonly used is ADF (Augmented Dicky Fuller Test). It's important to test the stationarity of data to apply fitting ARIMA model. In this research paper, Autocorrelation and Partial Autocorrelation graph is plotted and it talks about the two categories of ARIMA model: Seasonal ARIMA and non-seasonal ARIMA. The seasonal ARIMA is for seasonal data model incorporates both non-seasonal and seasonal factors in a multiplicative model. Finally, this research paper hope to analyze the dynamic data of the stock market and to propose an optimized ARIMA model for the dynamic data in the future [6].

- Stock price prediction using the ARIMA model published on March 2014, by Ayodele Ariyo Adebiyi, Charles Ayo, et at demonstrated the potential of ARIMA models to predict stock prices satisfactory on short-term basis. This paper presents extensive process of building stock price predictive model using the ARIMA model. The data composed of four elements, namely: open price, low price, high price and close price respectively. The model checking was done with Augmented Dickey Fuller (ADF) unit root test on "DCLOSE" of Nokia stock index. The result confirms that the series becomes stationary after the first-difference of the series. Another stock data i.e., of Zenith bank used in this study covered the period from 3rd January, 2006 to 25th February, 2011 with total of 1296 observations. From the graph, the performance of the ARIMA model selected is quite impressive as there are some instances of closely related of actual and predicted values. This research paper concludes by stating that results obtained revealed that the ARIMA model has a strong potential for short-term prediction and can compete favourably with existing techniques for stock price prediction. With the results obtained, ARIMA models can compete reasonably well with emerging forecasting techniques in short-term prediction [5].

CHAPTER 3 REQUIREMENT ANALYSIS

This project can run on commodity hardware. We ran entire project on an Intel I5 processor with 8 GB Ram, 2 GB Nvidia Graphic Processor, it also has 2 cores which runs at 1.7 GHz, 2.1 GHz respectively. First part of the is training phase which takes 10-15 mins of time and the second part is testing part which only takes few seconds to make predictions and calculate accuracy.

3.1 HARDWARE REQUIREMENT

- RAM: 4 GB
- Storage: 500 GB
- CPU: 2 GHz or faster
- Architecture: 32-bit or 64-bit

3.2 SOFTWARE REQUIREMENT

- Python 3.5 in Google Colab is used for data pre-processing, model training and prediction.
- Operating System: Windows 7 and above or Linux based OS or MAC O

CHAPTER 4 DETAILED DESIGN

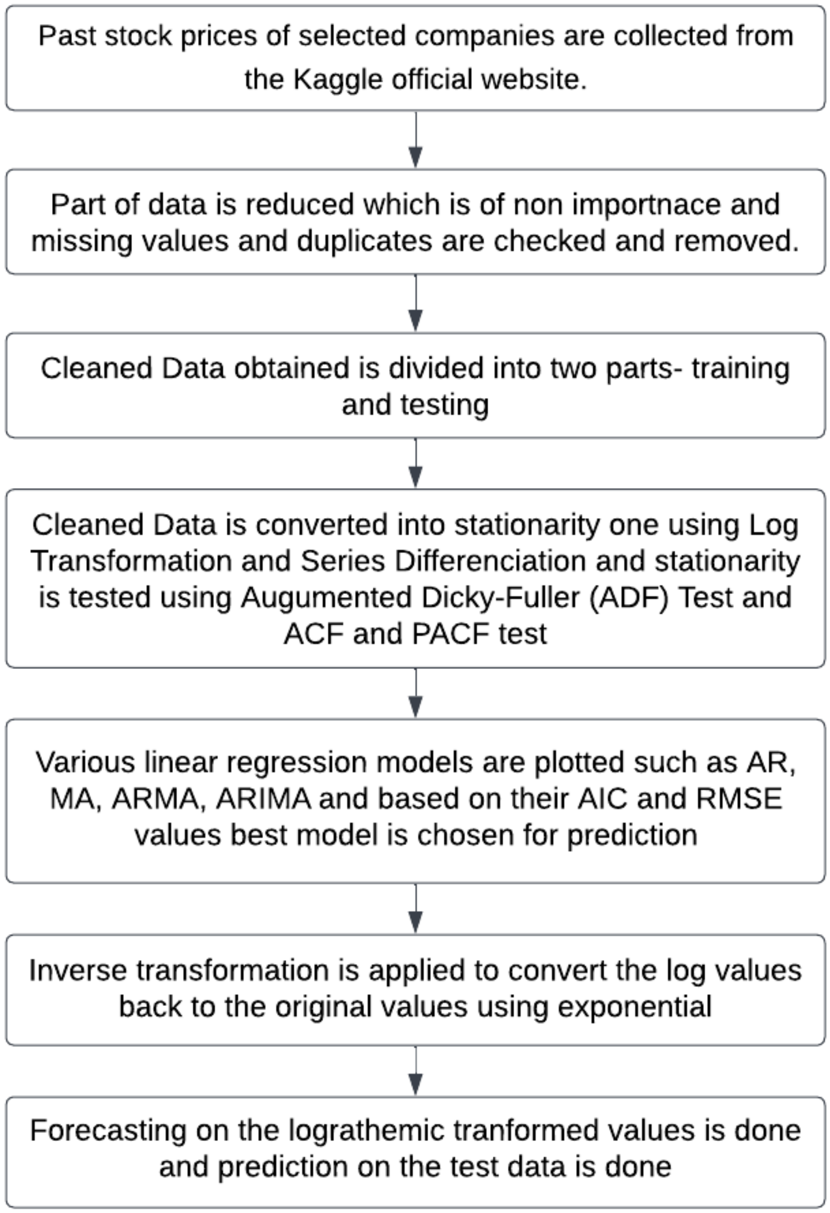

Fig1: Flow Chart

CHAPTER 5 IMPLEMENTATION

5.1 Raw Stock Price Dataset: Past stock prices of selected companies are collected from the Kaggle official website.

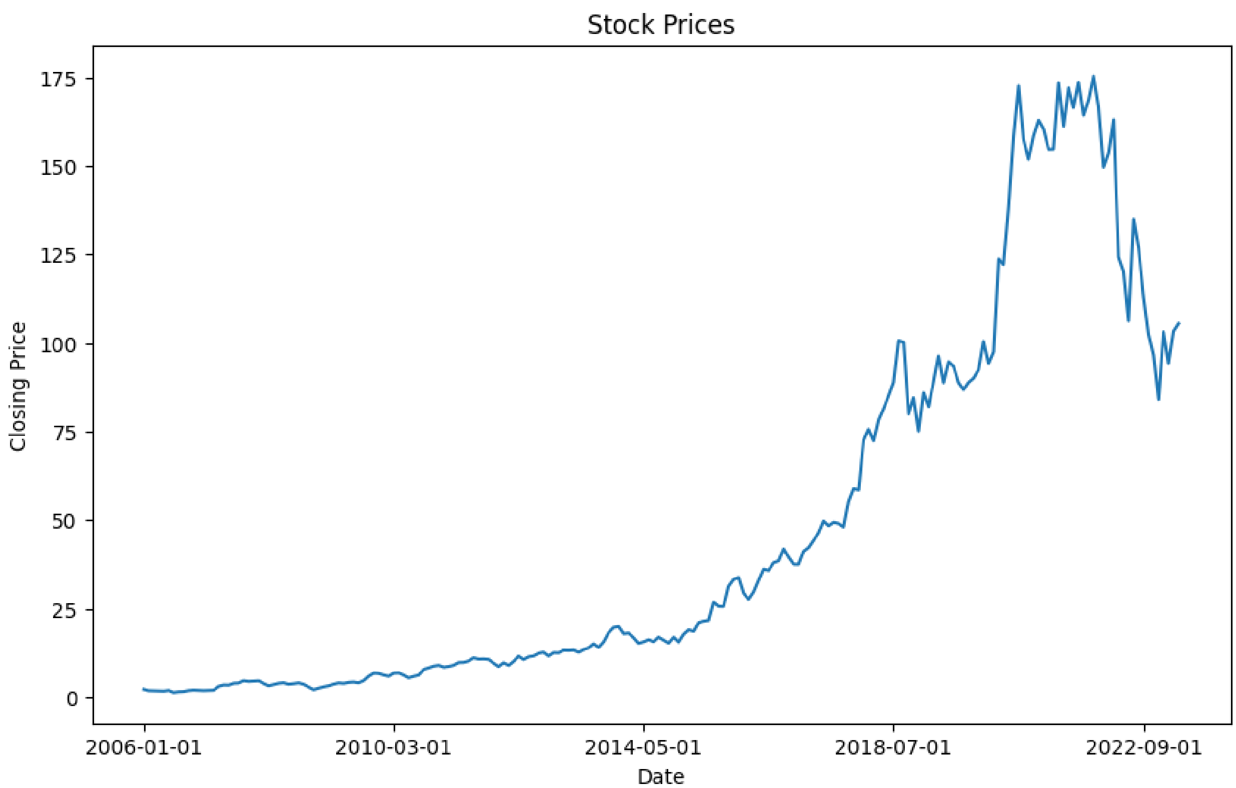

Fig2: Time Series Data Visualization

5.2 Pre-processing: This step incorporates the following:

1. **Data discretization**: Part of data reduction but with particular importance, especially for numerical data.
2. **Data Cleaning**: Fill in the missing values if empty and remove the duplicate entry and resetting the index.
3. **Data integration:** Integration of data files. After the dataset is transformed into a clean dataset, the dataset is divided into training and testing sets so as to evaluate.

5.3 Identify (Stationarity or not):

- Stationarity is tested using **Augmented Dickey-Fuller (ADF) Test and ACF and PACF.**

If non-stationary, it is converted to stationary by:

1. Log Transformation
2. By differencing the series (lagged series)

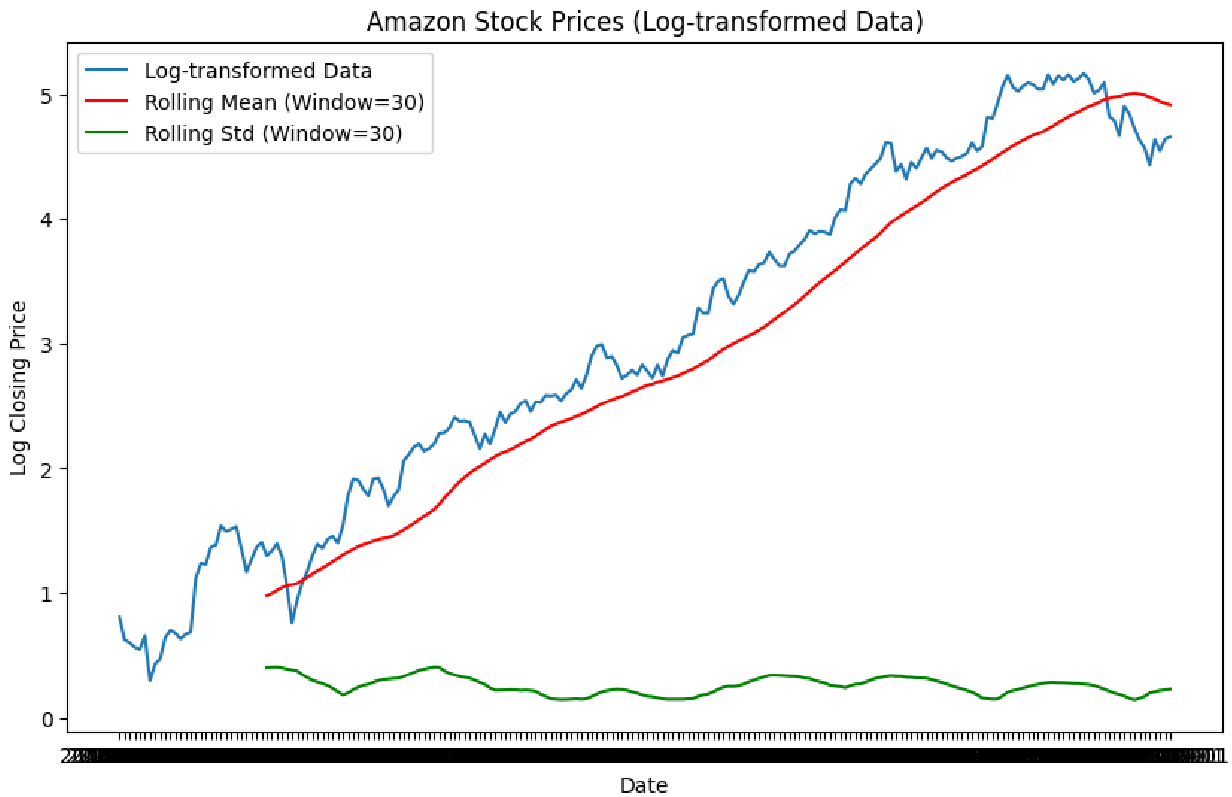

Fig3: Rolling mean and Rolling standard Deviation

The standard deviation is now nearly constant as seen from Fig3 which is a good sign that the differencing method worked somewhat and implies that the time series now has a constant variance.

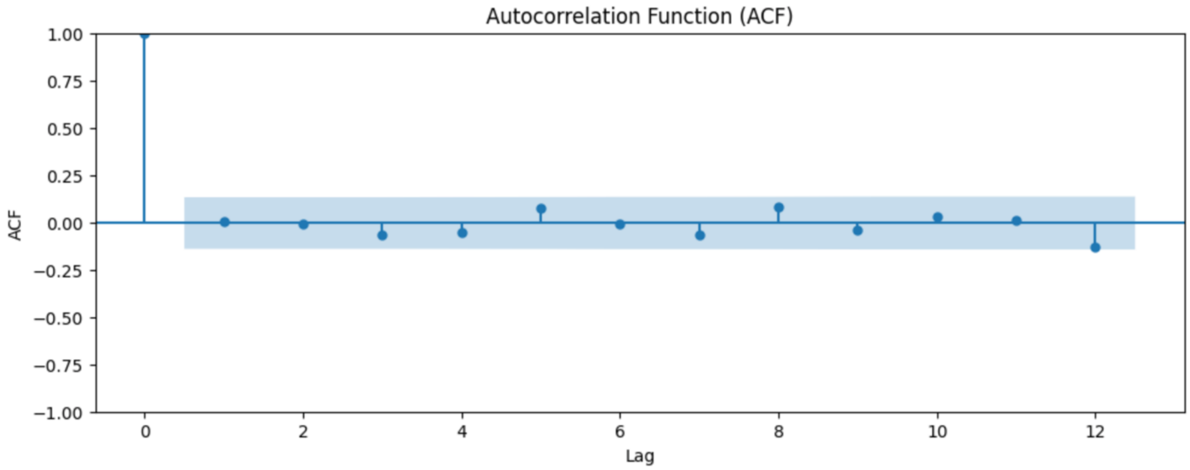

Fig5: ACF Test

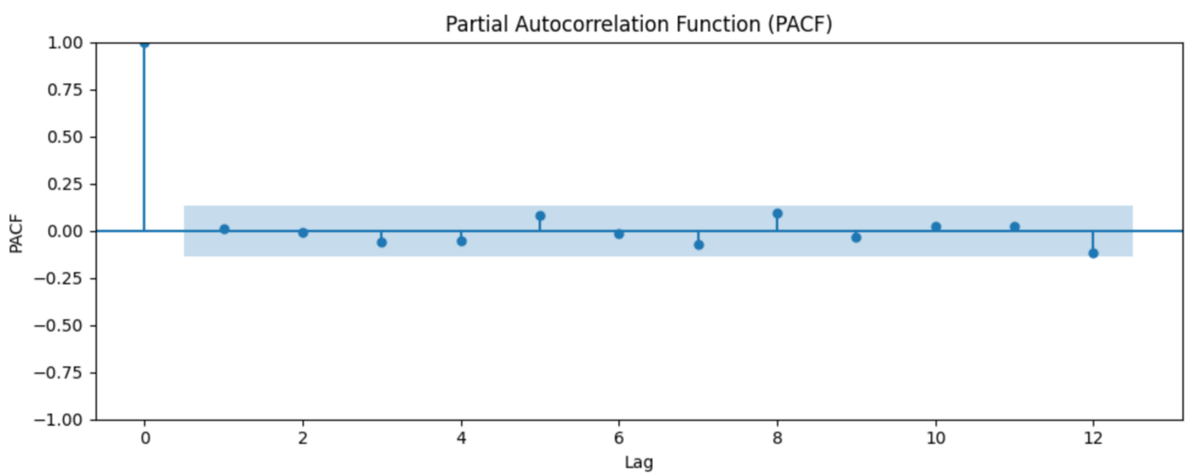

Fig6: PACF Test

In Fig5, the highest lag at which the plot extends beyond the statistically significant boundary is lag 1. This indicates that an AR Model of lag 1 ($p = 1$) should be sufficient to fit the data. Similarly, from Fig5, we can infer that $q = 1$.

5.4 Model Visualisation:

a) AR Model:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t,$$

where ε_t is white noise. This is like a multiple regression but with *lagged values* of y_t as predictors. We refer to this as an **AR(p) model**, an autoregressive model of order p.

b) MA Model:

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q},$$

where ε_t is white noise. We refer to this as an **MA(q) model**, a moving average model of order q. Of course, we do not *observe* the values of ε_t, so it is not really a regression in the usual sense.

c) ARIMA Model:

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t, \quad (8$$

where y'_t is the differenced series (it may have been differenced more than once). The "predictors" on the right hand side include both lagged values of y_t and lagged errors. We call this an **ARIMA(p, d, q) model**, where

$p =$	order of the autoregressive part;
$d =$	degree of first differencing involved;
$q =$	order of the moving average part.

d) SARIMA Model:

A seasonal ARIMA model is formed by including additional seasonal terms in the ARIMA models we have seen so far. It is written as follows:

$$\text{ARIMA} \quad \underbrace{(p, d, q)}_{\substack{\uparrow \\ \text{Non-seasonal part} \\ \text{of the model}}} \quad \underbrace{(P, D, Q)_m}_{\substack{\uparrow \\ \text{Seasonal part of} \\ \text{of the model}}}$$

where $m =$ number of observations per year. We use uppercase notation for the seasonal parts of the model, and lowercase notation for the non-seasonal parts of the model.

The seasonal part of the model consists of terms that are similar to the non-seasonal components of the model, but involve backshifts of the seasonal period. For example, an ARIMA$(1,1,1)(1,1,1)_4$ model (without a constant) is for quarterly data ($m = 4$), and can be written as

$$(1 - \phi_1 B)\ (1 - \Phi_1 B^4)(1 - B)(1 - B^4) y_t = (1 + \theta_1 B)\ (1 + \Theta_1 B^4)\varepsilon_t.$$

The additional seasonal terms are simply multiplied by the non-seasonal terms.

e) LSTM Model:

$$\begin{aligned} a^{(t)} &= Wh^{(t-1)} + Ux^{(t)} \\ h^{(t)} &= a^{(t)} \\ o^{(t)} &= Vh^{(t)} \\ \hat{y^{(t)}} &= Loss(o^{(t)}) \end{aligned}$$

In the above equations, we ignored the non-linearities and the biases. Adding those in the equations look like the following.

$$\begin{aligned} a^{(t)} &= b + Wh^{(t-1)} + Ux^{(t)} \\ h^{(t)} &= tanh(a^{(t)}) \\ o^{(t)} &= c + Vh^{(t)} \\ \hat{y^{(t)}} &= softmax(o^{(t)}) \end{aligned}$$

5.5 Model Selection: Best fitted linear regression model is selected based upon the RMSE values

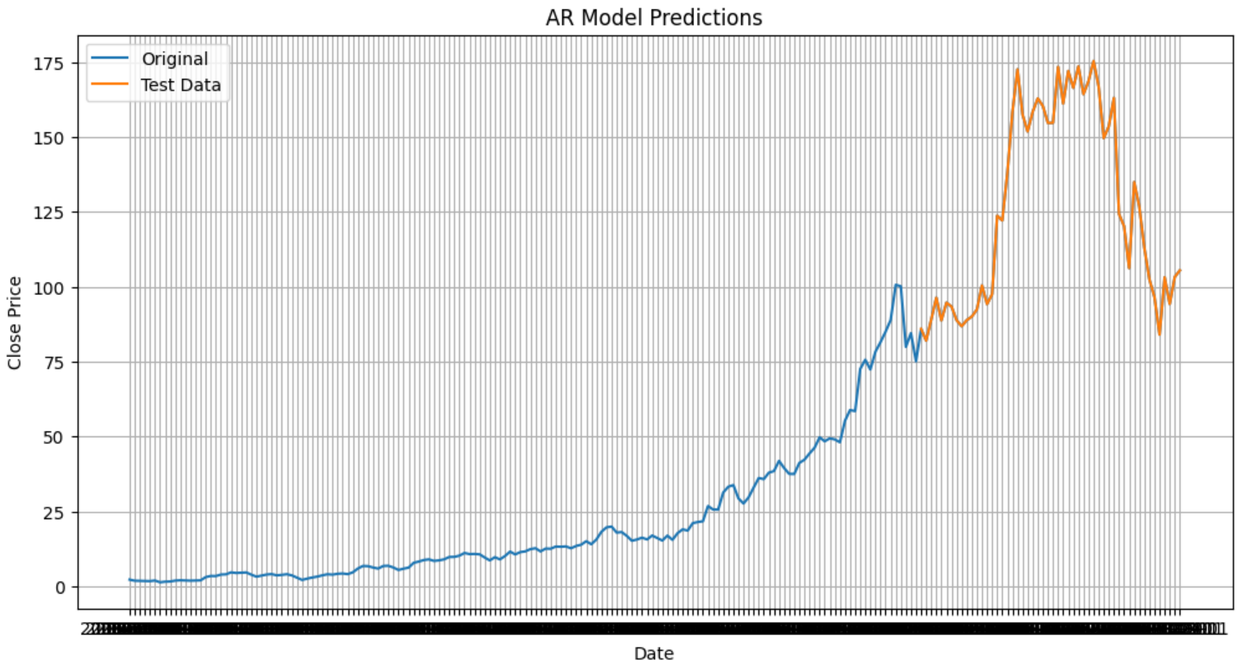

Fig7: AR Model

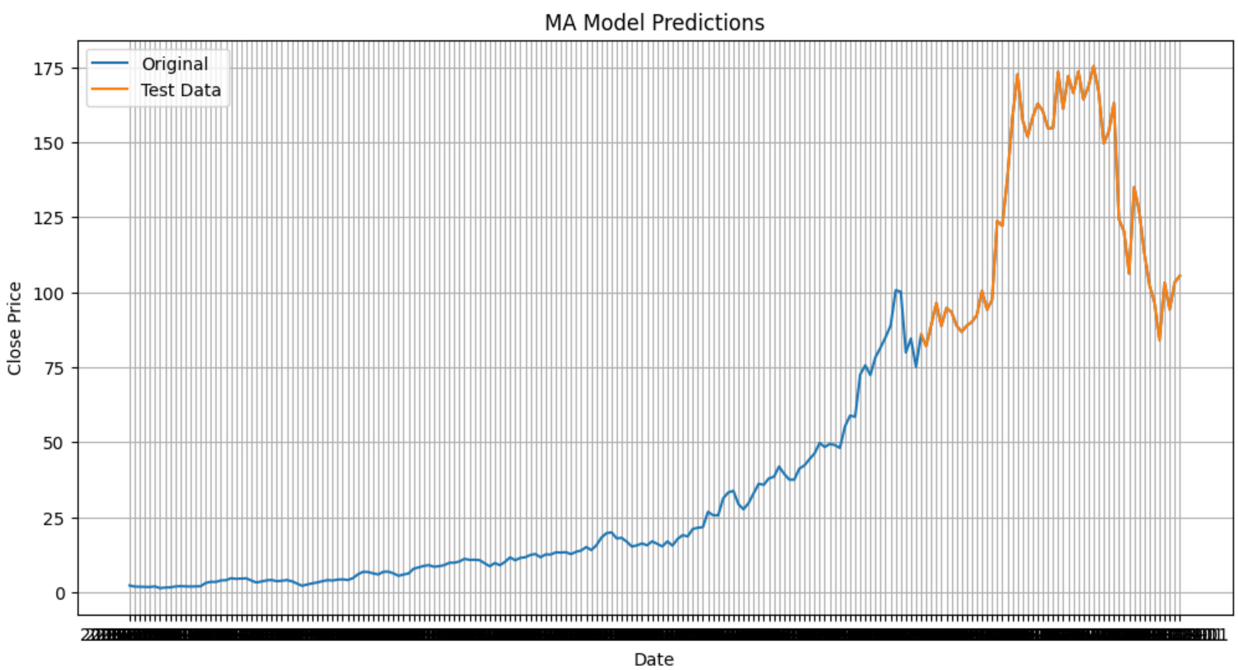

Fig8: MA Model

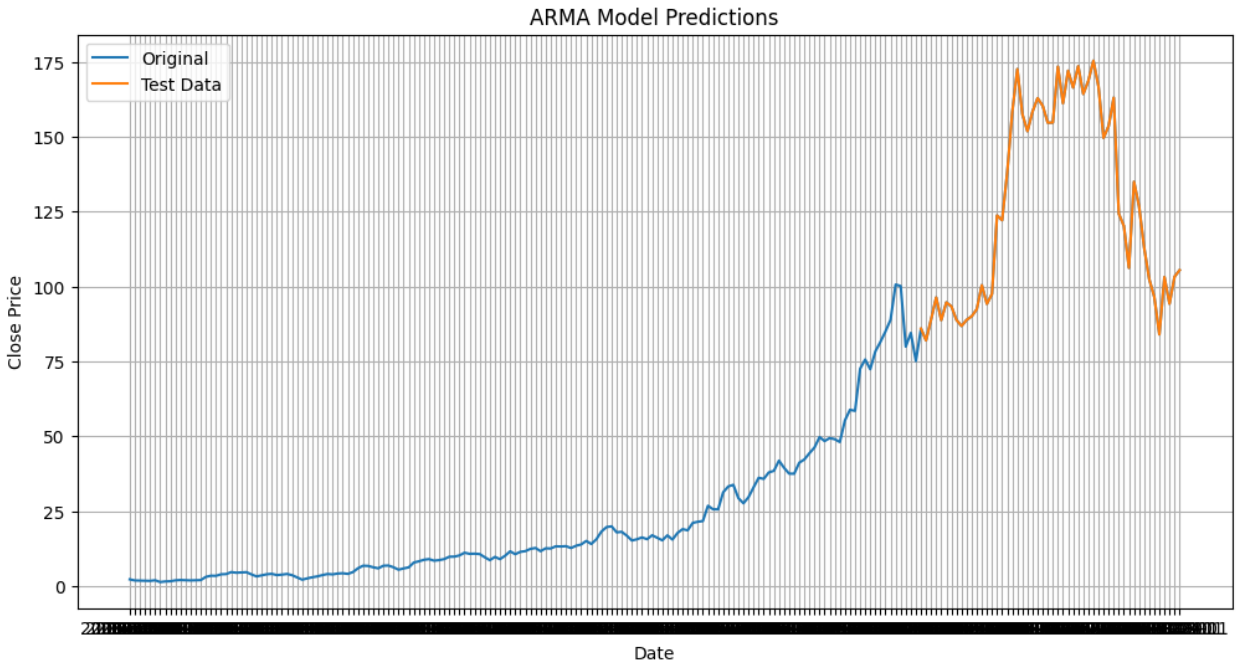

Fig9: ARMA Model

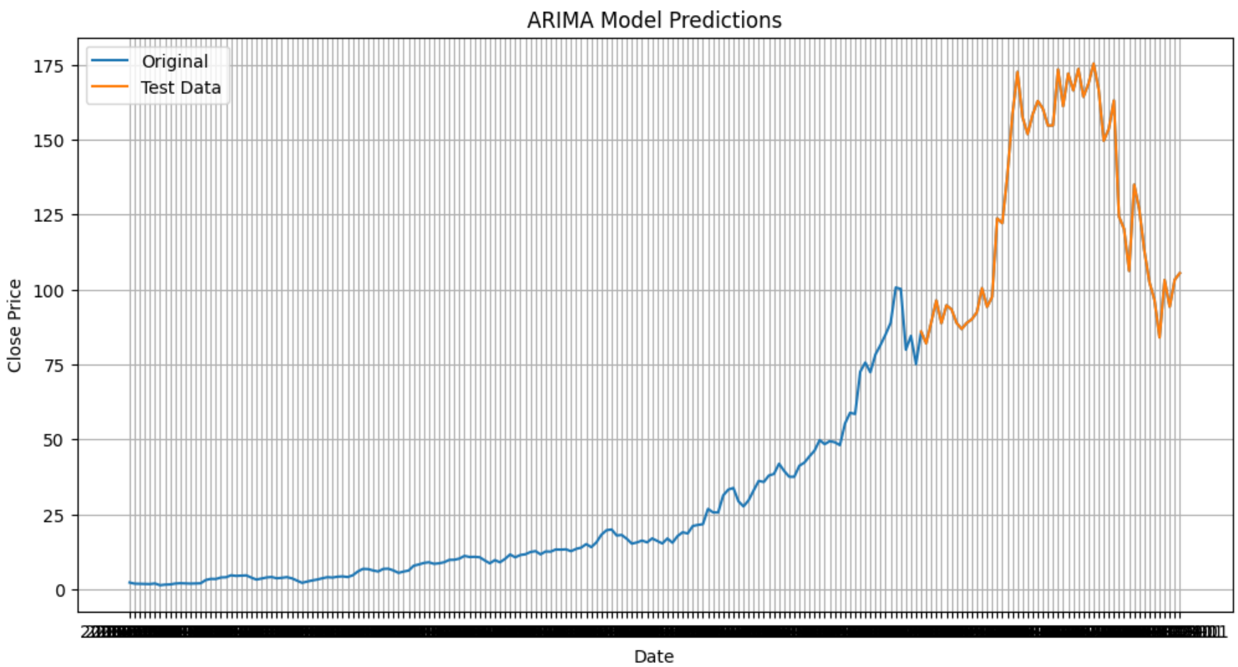

Fig10: ARMIA Model

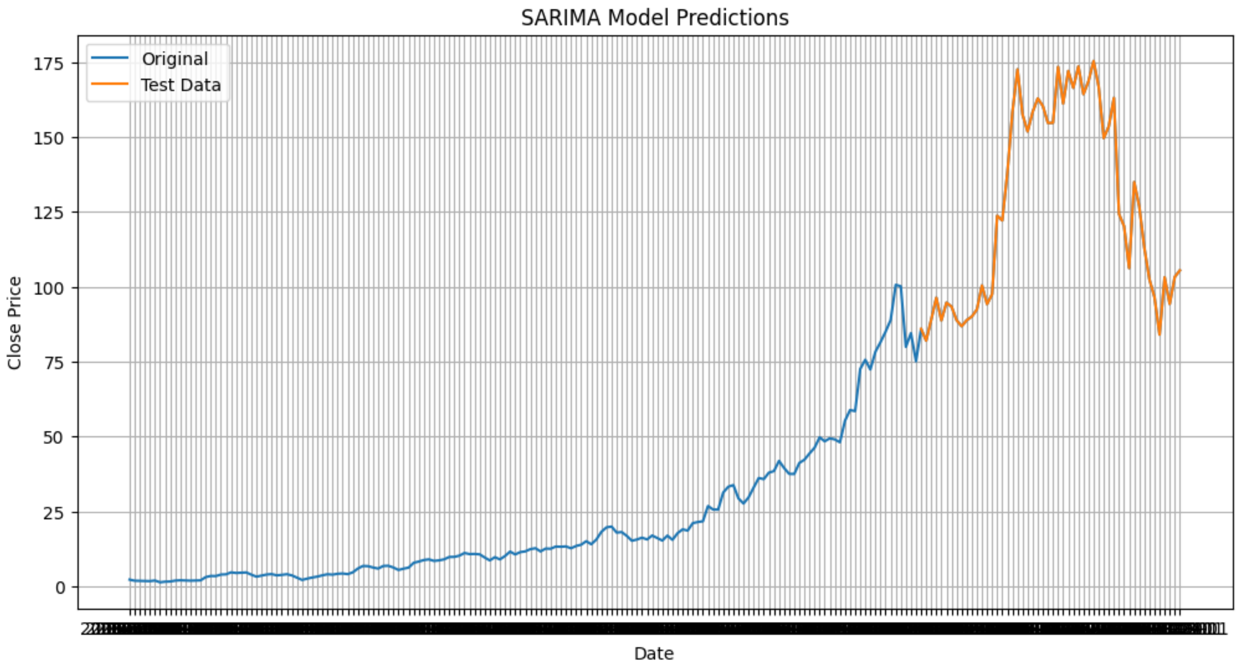

Fig11: SARIMA Model

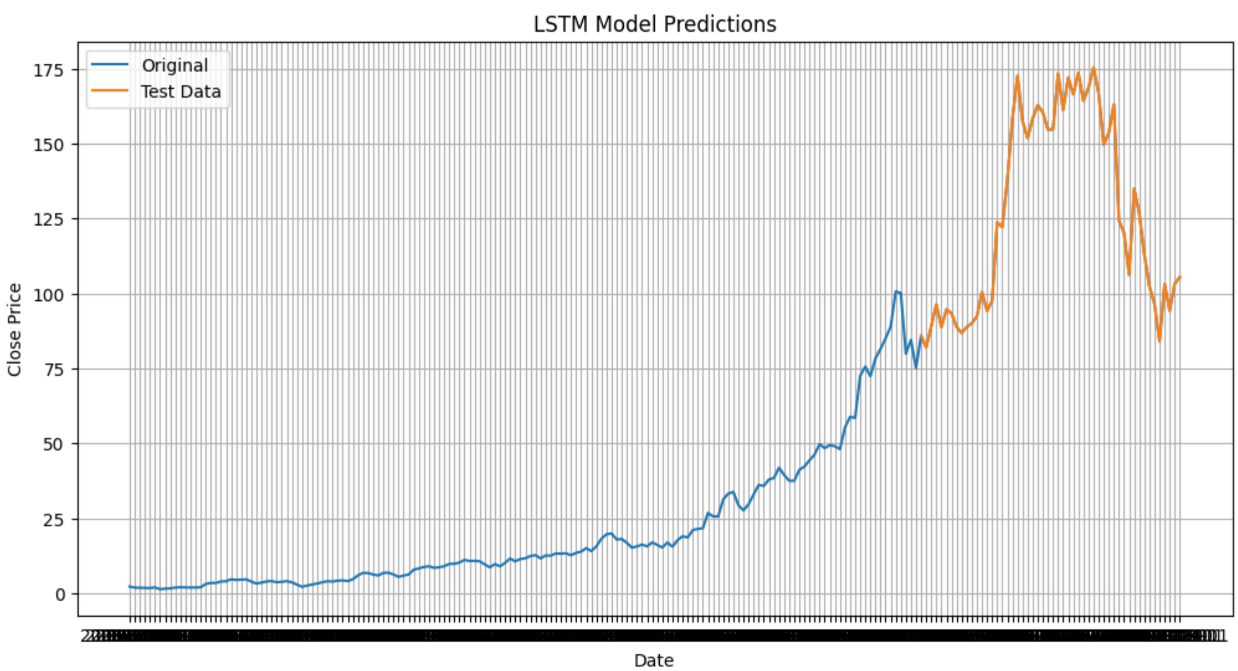

Fig12: LSTM Model

The RMSE value of ARIMA is much higher than other models as shown in Fig.12 implying ARIMA may be not a good model choice due to this increased error. ARMA has least AIC value and RMSE value as depicted in Fig.11 and Fig.12 implying ARMA as a better model.

CHAPTER 6 EXPERIMENTAL RESULTS AND ANALYSIS

6.1 PREDICTED VALUES ON TRAINING DATA

The original time series data with the predicted values on the training data and forecasted data (2014-2020) are shown below:

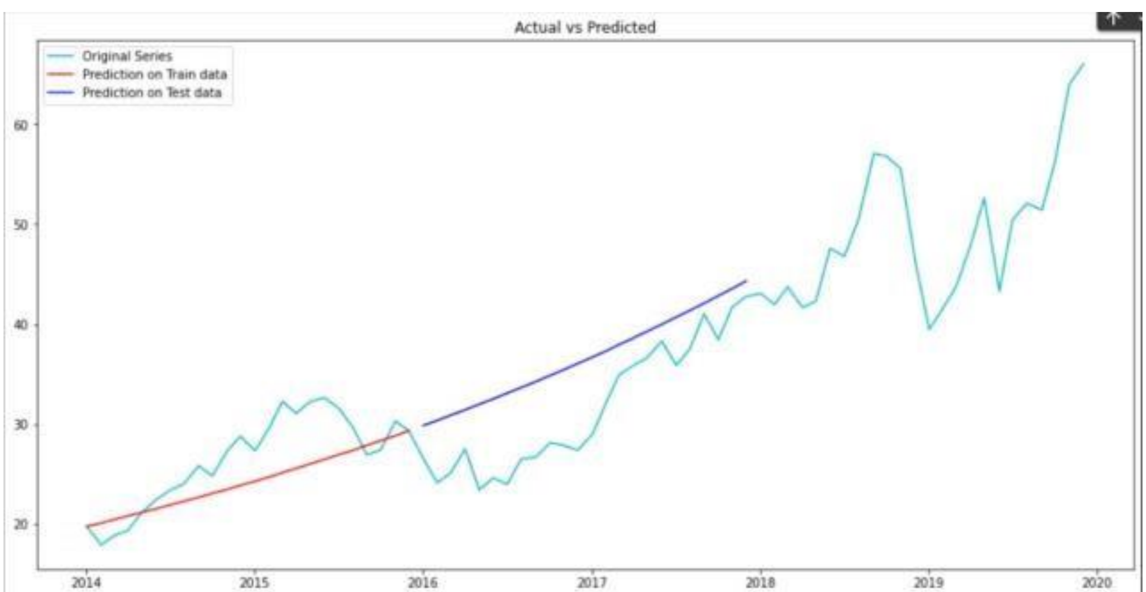

Fig15: Stock Price Prediction of Apple

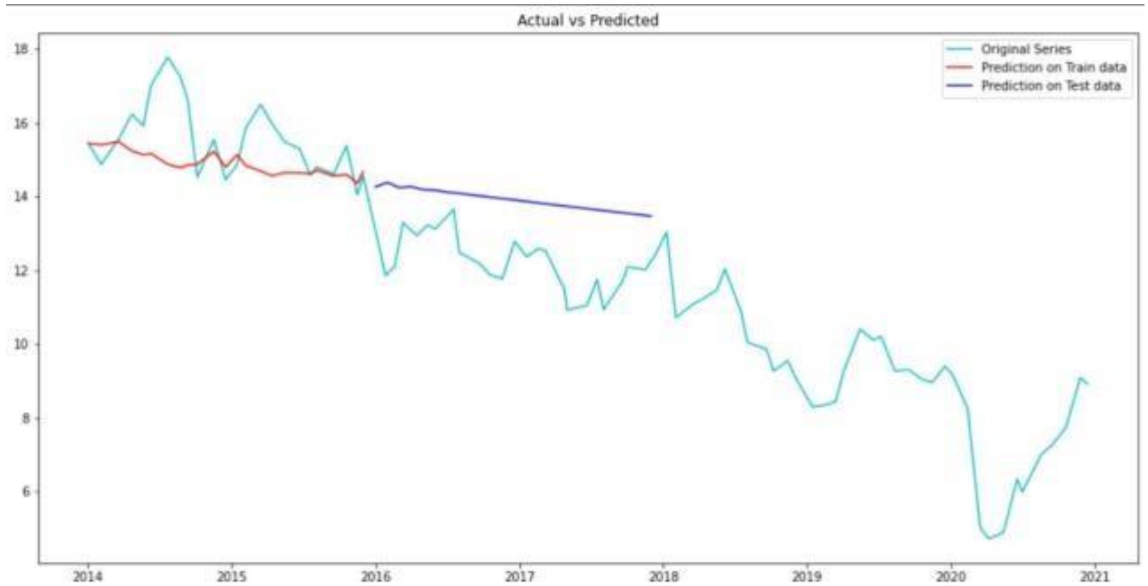

Fig16: Stock Price Prediction of Ford

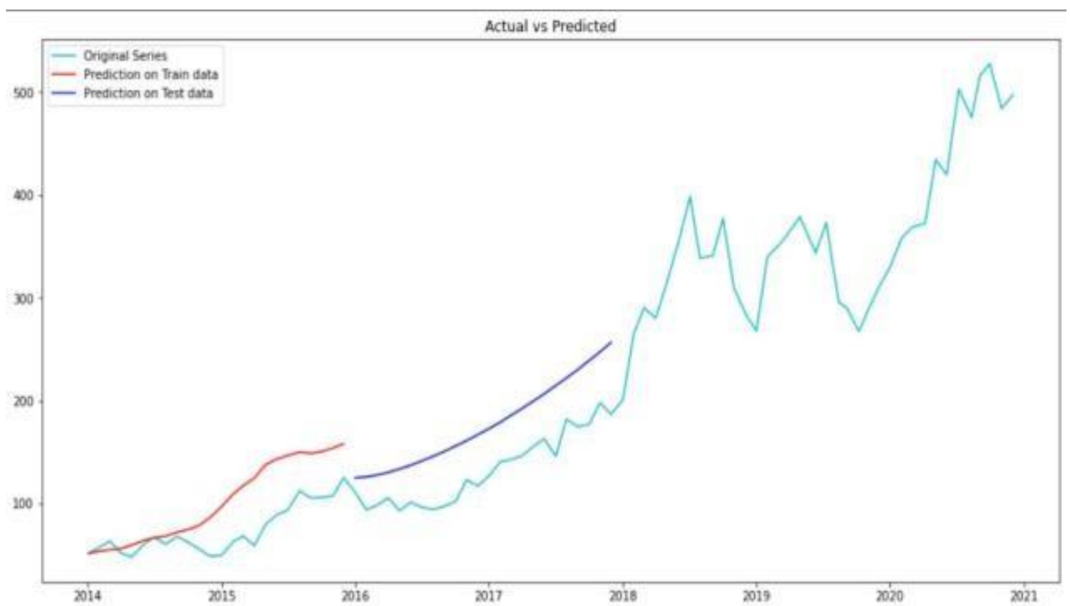

Fig17: Stock Price Prediction of Netflix

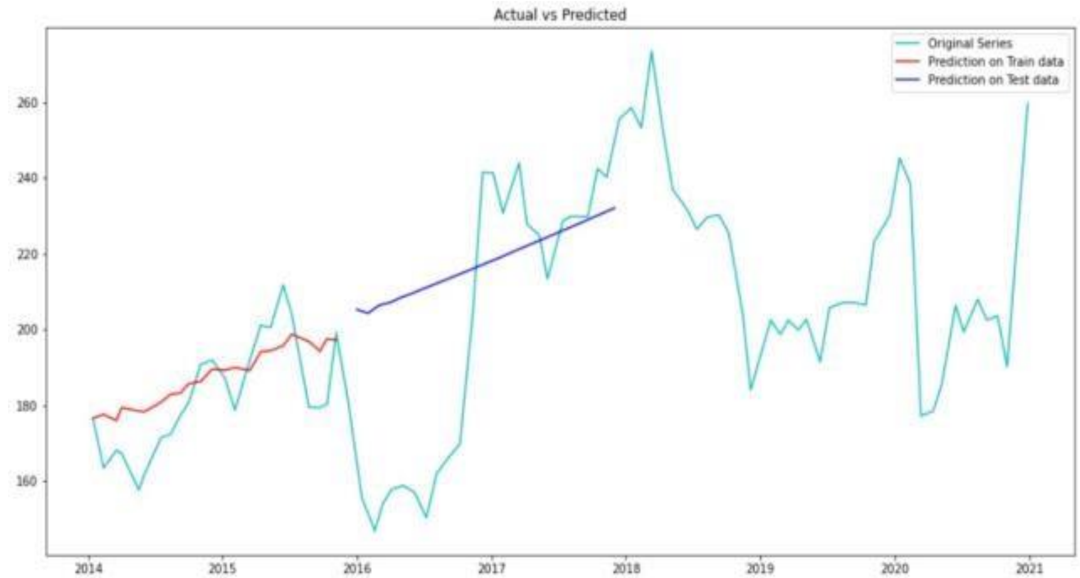

Fig18: Stock Price Prediction of Goldman Sachs

It is clear from the above graphs that all the series are non-stationary. We difference each of the series appropriate number of times till stationarity is achieved. We carry out the Augmented Dickey Fuller / Unit root test on the differenced series to be certain that they are converted to stationary series.

6.2 STANIONARITY CHECK

Unit root test/ADF Test We want to test whether each of the series is stationary under this test.

The null hypothesis is presence of unit root i.e., non- stationary.

COMPANY	TEST STATISTIC VALUE	P-VALUE	DECISION
Netflix	-4.597164	0.000131	Stationary
Goldman Sachs	-5.172841	0.000010	Stationary
Ford	-3.581487	0.006125	Stationary

Table1: Unit Root Test

As it can be seen from Table 1, the p-value corresponding to each series is less than level of significance (0.05) making accept the alternative hypothesis implying that all the differenced series are stationary. We have also checked the stationarity by ACF and PACF plots of each series.

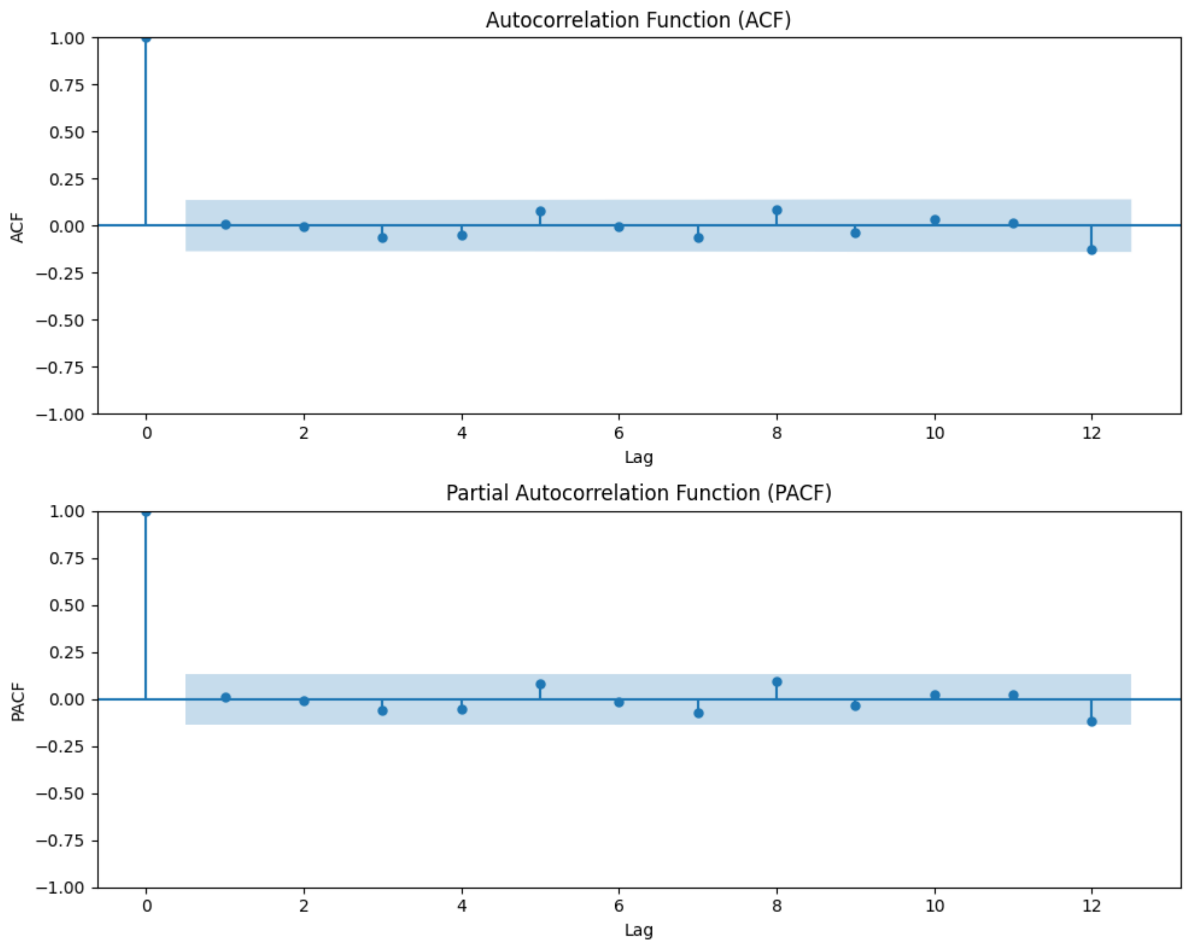

Fig19: ACF and PACF plots for Apple dataset

6.3 MODEL FITTING

We have used RMSE ass a measure for good fit. It measures the average difference between values predicted by a model and the actual values. Lesser the value of RMSE, better is the fit. The following graphs depicts the best fitted model for each state.

Hence, from the graph we infer SARIMA as the best fitted model having least RMSE value among all the other models.

CHAPTER 7 CONCLUSION AND FUTURE WORK

7.1 CONCLUSION

Stock and financial markets tend to be unpredictable and even illogical due to the unevenness of market volatility. Because of such characteristics financial or market data should possess a rather turbulent structure which often makes it hard to find reliable patterns. Modeling such structures requires methods and algorithms capable of finding hidden pattern or trend within data and predict how they will affect them in the future. One of the most common and efficient methodology is ARMA modelling of Time Series Data (TSD).

Autoregressive Moving Average (ARMA) Model converts non-stationary data to stationary data before working and is one of the most popular models to predict nearly linear time series data. ARMA model has been used extensively in the field of finance and economics as it is known to be robust, efficient and has a strong potential for short-term share market prediction. In this work we have forecasted the share prices of 4 major companies of India in 2014 for the next 6 years and as we have already observed the prediction based on our chosen ARMA model is fairly consistent.

We see a significant rise in the share price of Netflix, Apple, and Goldman Sachs in the near future, which calls for special mention. Another factor to pay attention is we do not go for a very long term prediction of such volatile data as several socio-economic factors are involved which play a major role in the rise and fall of market share prices. Though the companies that we have considered in this study i.e., Netflix, Apple, Goldman Sachs, and Ford do show a steady forecast in the upcoming few years, based on their data till 2020, but investors should not discard the influence of other external factors before making their decision.

7.2 FUTURE SCOPE

In the future, we can extend this project to incorporate more parameters and try different algorithms which are based on multivariate parameters. We can further extend this study to use the algorithms which are based on Machine Learning, Deep Learning and self-sufficient Artificial Intelligence.

Linear Regression, Logistic regression, Naïve Bayes Classifier, Random Forest Method, Decision Tree, SVM (Support Vector Machine), KNN (K-nearest neighbour), Artificial Neural Networks and various other algorithms can be used to get much more better results on the prediction of stock market that we did in this project.

This project was based purely on the univariate parameter (summarize only one variable at a time) where the non-stationary data was made stationary by applying the tests. We can study about the Seasonal Autoregressive Integrated Moving-Average with Exogenous Regressors (SARIMAX), Vector Autoregression (VAR), Vector Autoregression Moving-Average (VARMA), Vector Autoregression Moving-Average with Exogenous Regressors (VARMAX) and can apply them for the better and more precise prediction.

REFERENCES

Book

[1] Chris Chatfield, *Time Series Forecasting*. New York: Chapman and Hall/CRC, 2000.

Journal Article

[2] Sheikh Mohammad Idrees, M. Afshar Alam, Parul Agarwal,” A Prediction Approach for Stock Market Volatility Based on Time Series Data”, IEEE *Access*, vol. 7, pp. 17287-17298, 2019.

[3] Ashwini Saini and Anoop Sharma, “Predicting the Unpredictable: An Application of Machine Learning Algorithms in Indian Stock Market” Springer *Ann. Data. Sci.* 9, 791–799 (2022).

[4] Prof Tinni Chaudhuri, Dr. Abhijit Pandit,” Time-Series Model – Forecasting Performance in the Stock Market, Journal of Contemporary Issues in Business and Government Vol. 27, No. 2,2021

[5] Adebiyi, Ayodele & Adewumi, Aderemi & Ayo, Charles. (2014). “Stock price prediction using the ARIMA model. Proceedings”. UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, UKSim 2014. 10.1109/UKSim.2014.67.

[6] Dhyani, Bijesh & Kumar, Manish & Verma, Poonam & Jain, Abhishek. (2020). Stock Market Forecasting Technique using Arima Model. International Journal of Recent Technology and Engineering (IJRTE). 8. 2694-2697. 10.35940/ijrte.F8405.038620.

[7] B, Uma & D, Sundar & P, Alli. (2013). An Effective Time Series Analysis for Stock Trend Prediction Using ARIMA Model for Nifty Midcap-50. International Journal of Data Mining & Knowledge Management Process. 3. 65-78. 10.5121/ijdkp.2013.3106.

[8] Junior, Paulo & Pamplona, Edson & Salomon, Fernando. (2014). ARIMA: An Applied Time Series Forecasting Model for the Bovespa Stock Index. Applied Mathematics. 5. 3383-3391. 10.4236/am.2014.521315.

Online:

[9] "Time Series Models" https://towardsdatascience.com/time-series-models-d9266f8ac7b0

[10] "Apple Dataset" https://www.kaggle.com/datasets/meetnagadia/apple-stock-price-from19802021

[11] "Ford Dataset" https://www.kaggle.com/datasets/varpit94/ford-stock-data

[12] "Netflix Dataset" https://www.kaggle.com/datasets/akpmpr/updated-netflix-stock-priceall-time

[13] "Goldman Sachs Dataset" https://www.kaggle.com/datasets/varpit94/goldman-sachsstock-data-updated-till-1jul2021

[14] "LSTM - RNN" https://towardsdatascience.com/tutorial-on-lstm-a-computational-perspective-f3417442c2cd#:~:text=LSTM%20equations,-The%20figure%20below&text=The%20LSTM%20has%20an%20input,the%20next%20time%20step%20LSTM.