Zagdu Singh Charitable Trust's (Regd.)
THAKUR COLLEGE OF
ENGINEERING & TECHNOLOGY
Autonomous College Affiliated to University of Mumbai
Approved by All India Council for Technical Education(AICTE) and Government of Maharashtra(GoM)
Website : www.tcetmumbai.in
Conferred Autonomous Status by University Grants Commission (UGC) for 10 years w.e.f. A.Y 2019-20
Amongst Top 200 Colleges in the Country, Ranked 193rd in NIRF India Ranking 2019 in Engineering College category
• ISO 9001:2015 Certified • Programmes Accredited by National Board of Accreditation (NBA), New Delhi
• Institute Accredited by National Assessment and Accreditation Council (NAAC), Bangalore
Estd. in 2001

A.Y 2020-21
Institutional Internship

Track: Data Science using R

Prediction of Stock Market Price Using R Programming Language

Team Member Details:

Group 8

Shreyas Rathod SE CIVIL B 13
Anup Singh SE CIVIL B 27
Anshu Bhagat SE COMP A 14
Vikas Dubey SE COMP A 43
Shivam Dubey SE COMP A 40

Dr. Anand Khandare
Guide

INDEX

FIGURE INDEX

ABSTRACT:

Now every stock market has a stock exchange where the stocks of the company are listed. The stock is nothing but a piece of the company AKA share(s). The price of the shares keeps fluctuating and is dependent upon the value of the company. An investor or trader must know the right time to buy or sell shares to maximize his gains. The very uncertain nature of these fluctuations is what caused analysts to devise methods to predict the stock prices. We have collected our data through the function of package quantmod in r, then we have applied our Arima model to predict stock price and then we will be using the R programming language for forecasting.

INTRODUCTION:

Data Science is a popular subject nowadays. Everyone is all about data. What it can do and how it can help. Many times, data is represented as numbers and these numbers can represent many different things. These numbers could be the amount of sales, inventory, consumers, and last but definitely not least — *cash.* This brings us to financial data or more specifically the stock market. Stocks, commodities, securities, and such are all very similar when it comes to trading. We buy, we sell, we hold. All this in order to make a profit. The question is: How can Data Science help us when it comes to making these trades on the stock market.

MOTIVATION:

The highly publicized Enron Scandal, that eventually led to the bankruptcy of Enron Corporation is the very example of how stock prices change with public's perception. After achieving a lifetime high of stock price in mid-2000, the price of the shares plummeted to less than 1$ by the end of 2001. Hence it has been quite difficult to predict stock market prices although many theories have been devised. We will use the ARIMA model to analyse historical stock data. The results will be visualized using R.

PROBLEM STATEMENT:

To predict the stock market price of any company using the Arima model and forecasting method using R programming language.

IMPLEMENTATION:

- **TECHNOLOGIES USED**
  - R-Studio
- **ARCHITECTURE AND BLOCK DIAGRAM**

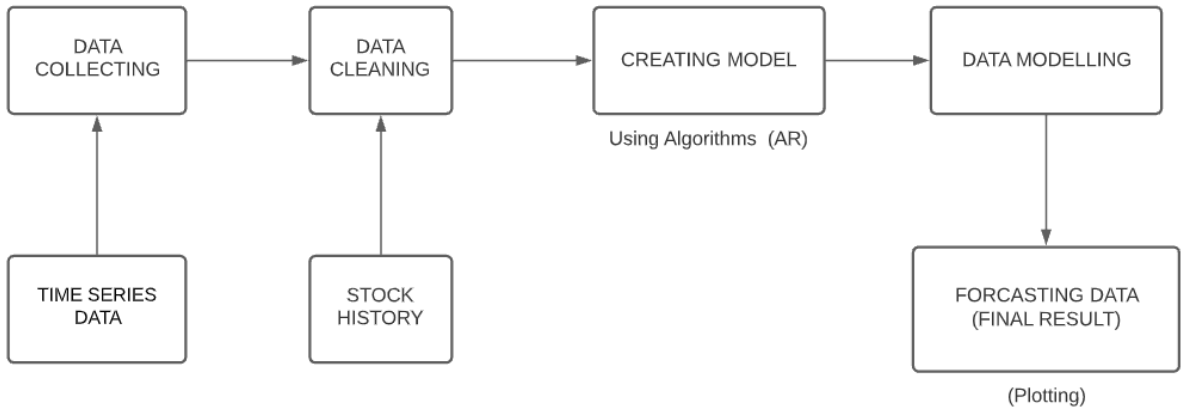


FIGURE 1 : BLOCK DIAGRAM

DATA SELECTION:

The Dataset we are using is dynamically getting from yahoo.finanical.com and it is time series data and we can get data of any time so are taking data from 01-06-2018 to 15-06-2021.

Following Dataset was used :

	AAPL.Open	AAPL.High	AAPL.Low	AAPL.Close	AAPL.Volume	AAPL.Adjusted
2018-06-18	46.9700	47.3050	46.8000	47.1850	73939600	45.62196
2018-06-19	46.2850	46.5825	45.8625	46.4225	134314000	44.88473
2018-06-20	46.5875	46.8000	46.4325	46.6250	82514800	45.08052
2018-06-21	46.8125	47.0875	46.2350	46.3650	102847600	44.82914
2018-06-22	46.5300	46.5375	46.1750	46.2300	108801600	44.69861
2018-06-25	45.8500	46.2300	45.1825	45.5425	126652400	44.03388
2018-06-26	45.7475	46.6325	45.6350	46.1075	98276800	44.58016
2018-06-27	46.3075	46.8200	46.0075	46.0400	101141200	44.51490
2018-06-28	46.0250	46.5525	45.9500	46.3750	69460800	44.83880
2018-06-29	46.5725	46.7975	45.7275	46.2775	90950800	44.74454
2018-07-02	45.9550	46.8250	45.8550	46.7950	70925200	45.24490
2018-07-03	46.9475	46.9875	45.8850	45.9800	55819200	44.45689
2018-07-05	46.3150	46.6025	46.0700	46.3500	66416800	44.81463
2018-07-06	46.3550	47.1075	46.3000	46.9925	69940800	45.43585
2018-07-09	47.3750	47.6700	47.3250	47.6450	79026400	46.06673
2018-07-10	47.6775	47.8200	47.5450	47.5875	63756400	46.01114
2018-07-11	47.1250	47.4450	46.9025	46.9700	75326000	45.41409
2018-07-12	47.3825	47.8525	47.3275	47.7575	72164400	46.17551
2018-07-13	47.7700	47.9600	47.7250	47.8325	50055600	46.24802
2018-07-16	47.8800	48.1625	47.6050	47.7275	60172400	46.14651
2018-07-17	47.4375	47.9675	47.3000	47.8625	62138000	46.27703
2018-07-18	47.9450	47.9500	47.4825	47.6000	65573600	46.02322
2018-07-19	47.4225	48.1375	47.4225	47.9700	81147200	46.38097
2018-07-20	47.9450	48.1075	47.5425	47.8600	82704800	46.27461
2018-07-23	47.6700	47.9900	47.3900	47.9025	63957600	46.31570

FIGURE 2 : DATASET FROM YAHOO FINANCE

➢ PACKAGES USED FOR DATA SELECTION:

- ♦ Quntmod(Quantitative Financial Modelling Framework)
- ♦ TimeSeries

DATA PRE-PROCESSING:

we have used only those attributes which were relevant to the cases generated. for example (Adjusted Price) After this we have calculated the Auto correlation, Partial Autocorrelation and adf test between the attributes and then we move further with modelling.

DATA VISUALIZATION:

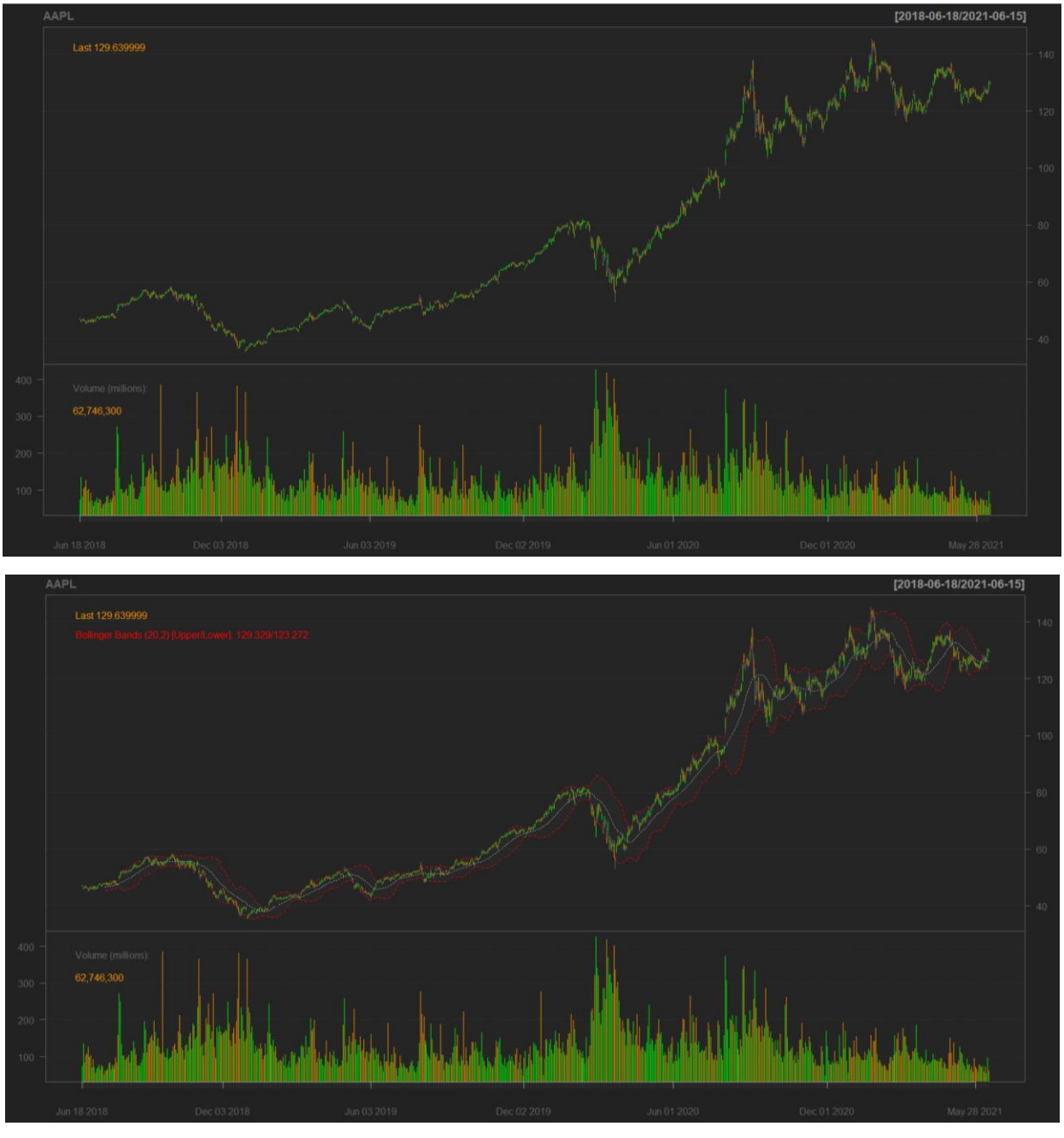


FIGURE 3: DATA VISUALIZATION OF TIMESERIES DATA

- **PACKAGE USED :**

  Quantmod

  TimeSeries

  Tseries

MODEL USED:

ARIMA Model :

ARIMA stands for auto-regressive integrated moving average and is specified by these three order parameters: (p, d, q).

$$Y_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + .. + \beta_0 Y_0 + \epsilon_t$$

$$Y_{t-1} = \beta_1 Y_{t-2} + \beta_2 Y_{t-3} + .. + \beta_0 Y_0 + \epsilon_{t-1}$$

Auto Regression AR(p) Auto regression technique estimates the future values based on the previous values. The function of an autoregressive model is denoted by AR(p), where p represents the order of the model. AR(0), the simplest process, involves no dependence between terms, preceding or current. For a first order autoregressive model AR(1), the preceding term and a percentage of error contribute to the output. AR(2) model takes into account 2 preceding values and noise to predict the output.

Moving Average MA(q) A moving average is a technique to model datasets that vary according to single factor. It finds the future trends based on the previous values that do not follow a definitive pattern. The two commonly used moving average techniques are exponential moving average (EMA) and the simple moving average (SMA). A moving average (MA(q)) component represents the error of the model as a combination of previous error terms.

The **Integrated I(d)** part of the model is used to stabilize the time series data by differencing. Differencing a series involves simply subtracting its current and previous values d times.

The order of an ARIMA model is generally represented as ARIMA(p,d,q),

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

To make our predictions accurate we have to predict the values of p, d and q. Now there are 2 ways to do this. Either using co-relations or using the auto.arima() function available in the forecast package. In the above code we could have also used the auto.arima() function to estimate the values of p,d,q.

IMPLEMENTING MODEL:

- First step is to install the Quant Mod package install. packages (quantmod') Next, we need to select a secure CRAN Mirror CRAN Mirror allows us to select a download site close to us for better bandwidth/latency.

- Add the QuantMod library library(“quantmod”).
- To retrieve data set from QuantMod get Symbols (“AAPL”) AAPL For retrieving data of a particular period Get Symbols (“AAPL”, from=2016-01-01”, to=”2017-01-01”) We can obtain the first several rows of a data set using the head function head (AAPL)
- Plot Visualization using charts chart Series (MSFT) chart Series (MSFT, subset=”2017-01: 2017-06”) By default it shows the OHLC data and the volume data For detailed analysis we can use add MACD and add B Bands Moving Average Convergence Divergence (MACD) is a centred oscillator that fluctuates above and below the zero line. It helps us analyse the strength, weakness or direction of momentum of a security’s move. BBands allows us to compare the volatility and price levels of a stock over a period of time.
- View and analyse results to get the short-term predictions, correlations can be found out once we plot the results.
- Plotting the stock values open price, close price, high price, low price, etc.
- Forecasting using Arima model: We use the forecast package to predict the stock price over the next 2 years. Prediction method=ARIMA Model h=24 (Predict 2 years into the future) level =95 (95% confidence level) installs. packages (‘forecast’) library (forecast) aapl. f=forecast (aapl. stl, method="arima", h=24, level=95) plot (aapl. f, ylab="Stock Price", xlab="Year", sub="Forecast from October 2018 to October 2020").

PERFORMANCE EVALUATION:

Improving the accuracy of model with the help of true data. Since the model uses prebuilt function named accuracy(preds, MSFT_ret_test) which works on the below mentioned formula.

$$\mathrm{RMSD} = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \hat{x}_i)^2}{N}}$$

RESULT INTERPRETATION:

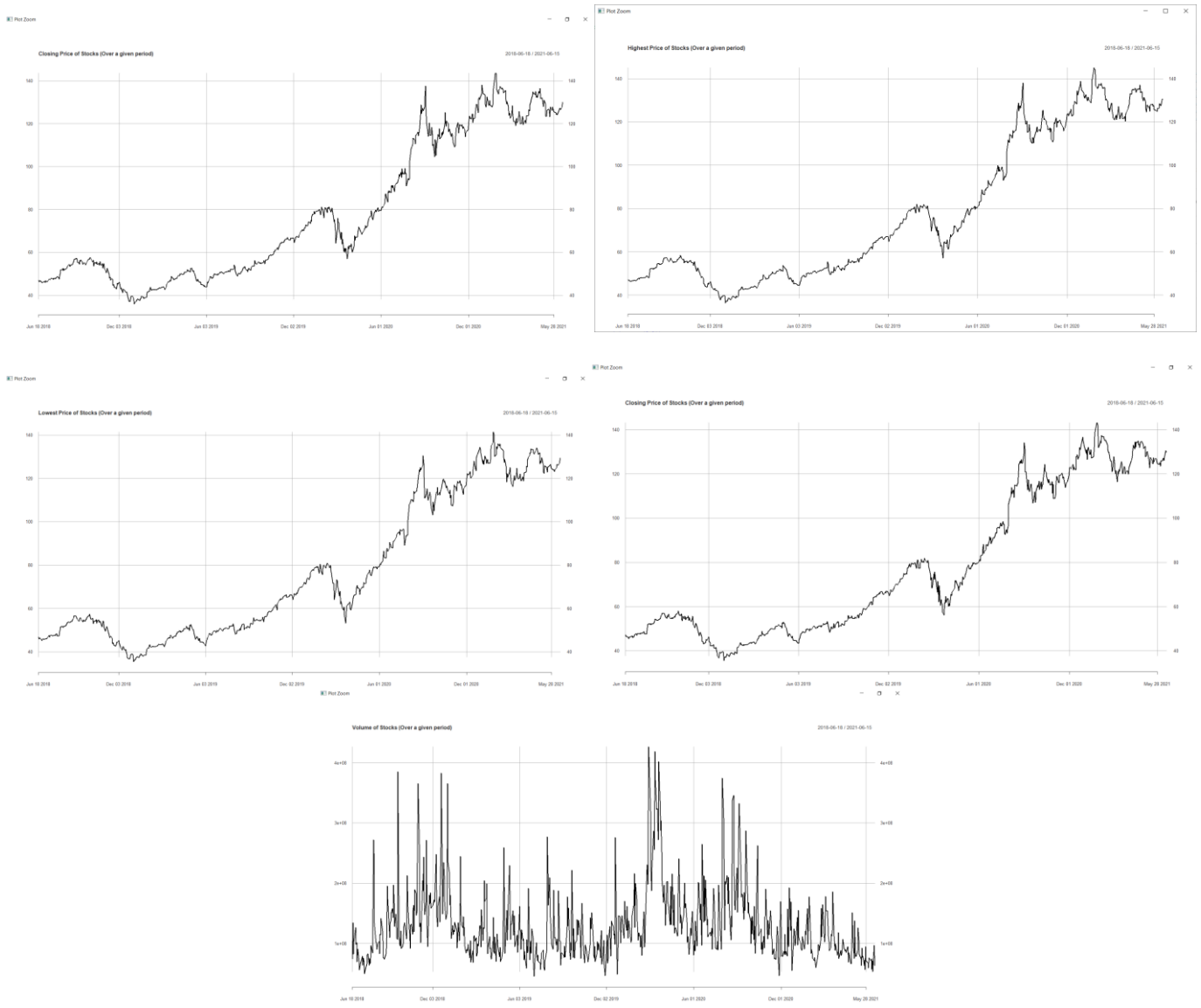
An intrinsic shortcoming of the ARIMA models, which is evident from the plot below, is the assumption of the mean reversal of the series. What this means is that after some time future, the forecasts would tend to the mean of the time series’ s historical values thus making it a poor model for long term predictions.

CODE:

```
################## Stock Market Price Prediction using R #################

##Importing Required Packages
library(quantmod)
library(tseries)
library(timeSeries)
library(forecast, quietly = T)

##Importing Dataset from Finance Websites...(Default yahoo)
getSymbols('AAPL', from = '2018-06-16', to = '2021-06-16')
View(AAPL)
#class(AAPL)
chartSeries(AAPL, type = 'auto')
addBBands()

##Assigning columns of dataset
Open_prices = AAPL[,1]
High_prices = AAPL[,2]
Low_prices = AAPL[,3]
Close_prices = AAPL[, 4]
Volume_prices = AAPL[,5]
Adjusted_prices = AAPL[,6]
#View(_prices)

par(mfrow = c(1,1))
plot(Open_prices, main = 'Closing Price of Stocks (Over a given period)')
plot(High_prices, main = 'Highest Price of Stocks (Over a given period)')
plot(Low_prices, main = 'Lowest Price of Stocks (Over a given period)')
plot(Close_prices, main = 'Closing Price of Stocks (Over a given period)')
plot(Volume_prices, main = 'Volume of Stocks (Over a given period)')
plot(Adjusted_prices, main = 'Adjusted Price of Stocks (Over a given period)')

Predic_Price = Adjusted_prices
#class(Predic_Price)

######## Finding the Linear Relation between observations ########

par(mfrow = c(1,2))
Acf(Predic_Price, main = 'ACF for differenced Series')
Pacf(Predic_Price, main = 'PACF for differenced Series ', col = '#cc0000')
Auto_cf = Acf(Predic_Price, plot = FALSE)
Auto_cf
PAuto_cf = Pacf(Predic_Price, plot = FALSE)
PAuto_cf

print(adf.test(Predic_Price))

#################### Prediction of Return ############################

return_AAPL <- 100*diff(log(Predic_Price))
AAPL_return_train <- return_AAPL[1:(0.9*length(return_AAPL))]
AAPL_return_test <- return_AAPL[(0.9*length(return_AAPL)+1):length(return_AAPL)]
auto.arima(AAPL_return_train, seasonal = FALSE)
fit <- Arima(AAPL_return_train, order = c(1,0,0))
#fit
preds <- predict(fit, n.ahead = (length(return_AAPL) - (0.9*length(return_AAPL))))$pred
preds

################## Forecasting Predicted Result ##################

test_forecast <- forecast(fit,h = 15)
test_forecast
par(mfrow = c(1,1))
plot(test_forecast, main = "Arima forecast for Apple Stock")

accuracy(preds, AAPL_return_test)
```

RESULT AND DESCRIPTION:

Following are the result of after calling plot () function:



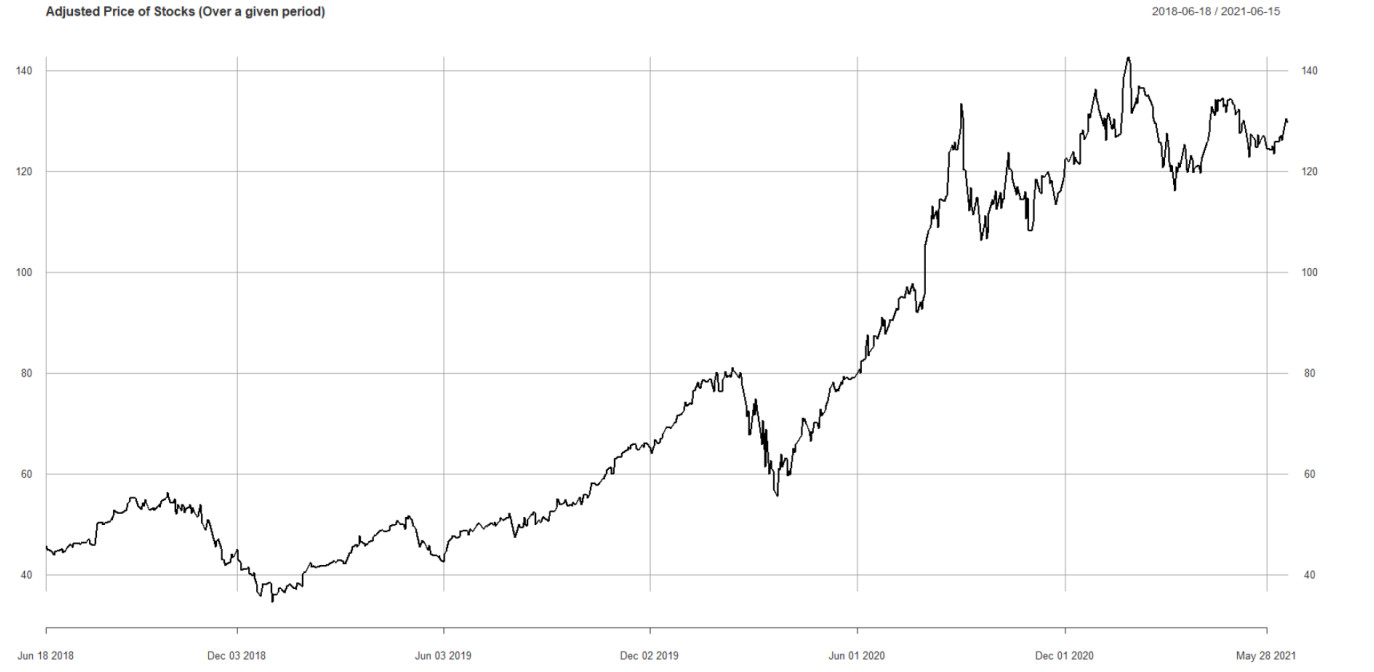


FIGURE 4: PLOTS OF DIFFERENT COLUMNS OF OUR TIMESERIES DATA

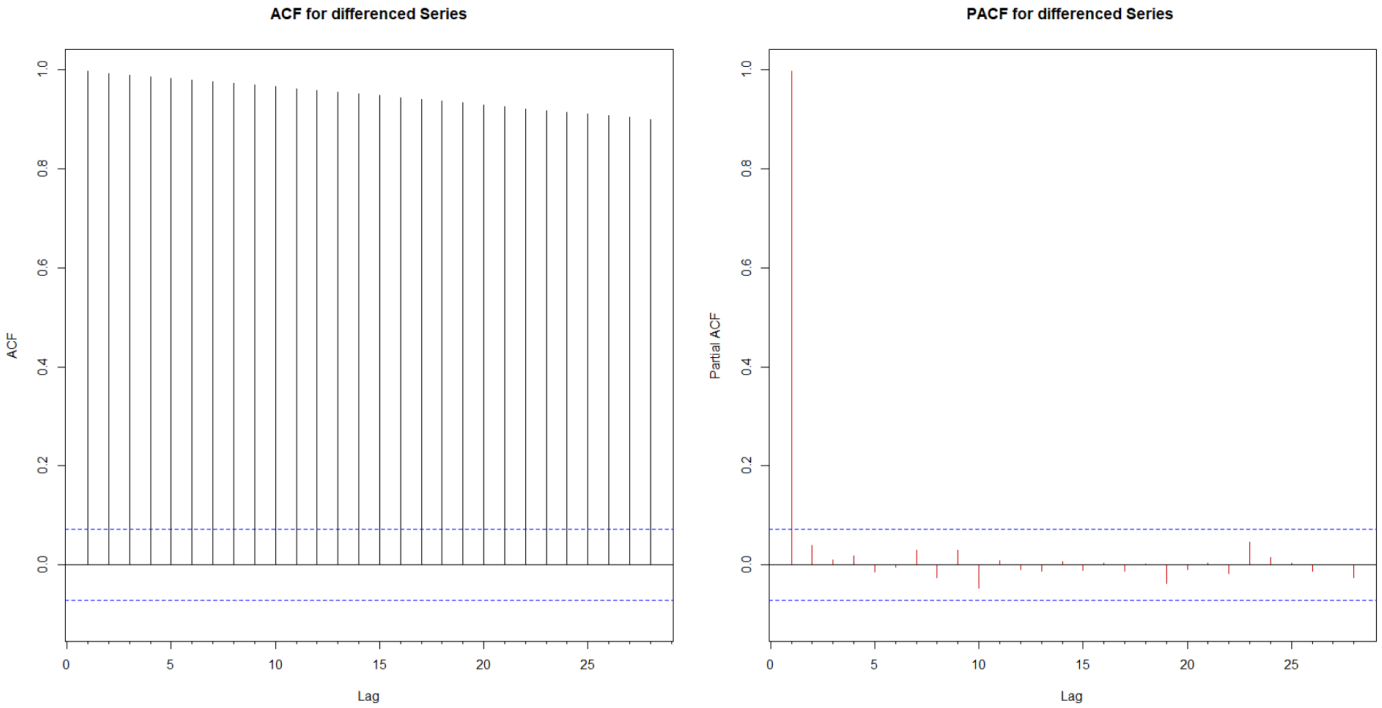


FIGURE 5: PLOT OF ACF AND PACF FOR TIMESERIES

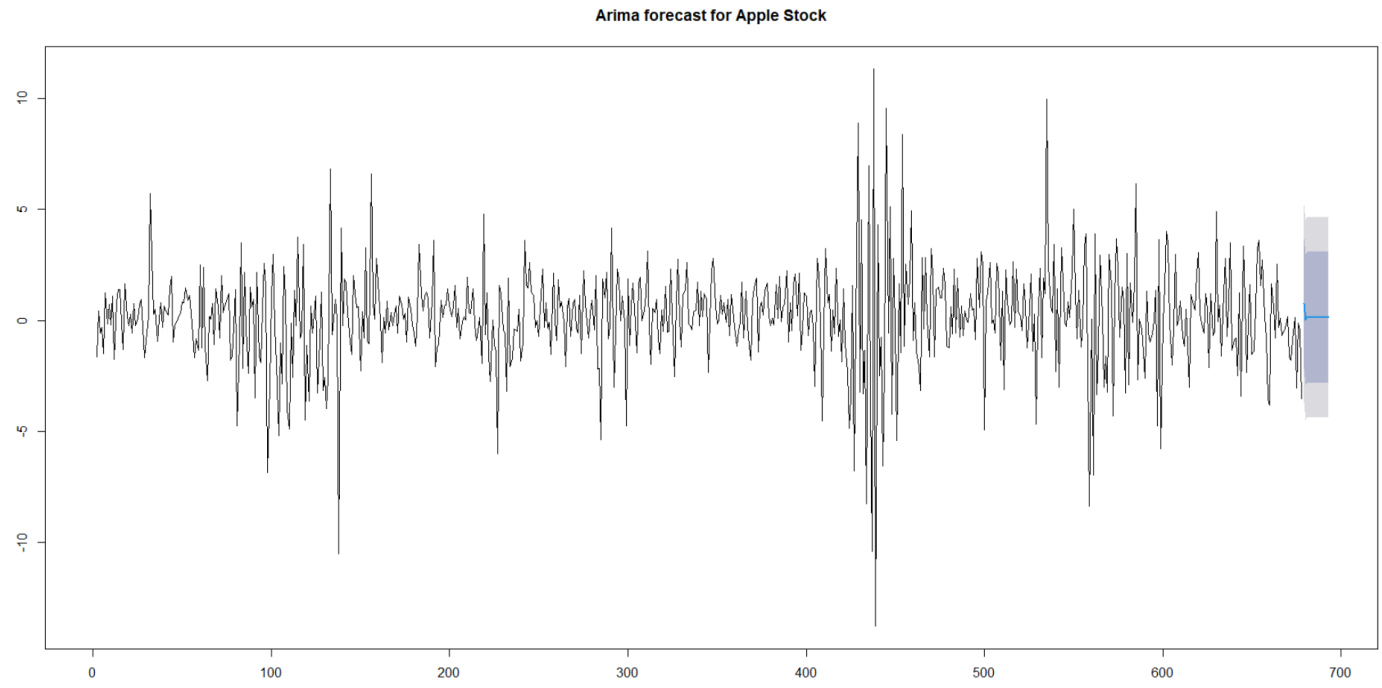


FIGURE 6: FORECASTING USING ARIMA MODEL

The deeply shaded region provides us the 80% confidence interval and the lightly shaded region provides the 95% confidence interval. The basic interpretation of a 95% confidence interval of the model tells us that the forecasted values will have a maximum deviation of +/- 2 as shown in the plot above, thus giving us a fair estimation of the values of the future stock indices.

CONCLUSION:

In this project we have predicted the share market trends by considering a company's historical time series data. The ARIMA Model played a key role in predicting the short-term trends of a stock market. The investors can do a thorough analysis of the share market by visualizing the graphical plots. There are various technical ways in which the plots can be visualized. The divergence with price signal is generated very rarely but is the most effective technique for precise analysis. This can help the investors make a profitable decision on whether to buy/sell/hold a stock.

FUTURE SCOPE:

Predicting stock market returns is a challenging task due to consistently changing stock values which are dependent on multiple parameters which form complex patterns. The historical dataset available on company's website consists of only few features like high, low, open, close, adjacent close value of stock prices, volume of shares traded etc., which are not sufficient enough. To obtain higher accuracy in the predicted price value new variables have been created using the existing variables. ANN is used for predicting the next day closing price of the stock and for a comparative analysis, RF is also implemented. The comparative analysis based on RMSE, MAPE and MBE values clearly indicate that ANN gives better prediction of stock prices as compared to RF. Results show that the best values obtained by ANN model gives RMSE (0.42), MAPE (0.77) and MBE (0.013). For future work, deep learning models could be developed which consider financial news articles along with financial parameters such as a closing price, traded volume, profit and loss statements etc., for possibly better results.

REFERENCES:

[1] https://shakil-stat-bsmrstu.gitbook.io/stockproject/stock-market-prediction-with-r

[2] Banerjee D. Forecasting of Indian stock market using time-series ARIMA model, 2nd IEEE International Conference on Business and Information Management (ICBIM), 2014, 131-135.

[3] https://www.w3schools.com/r/default.asp

[4] https://otexts.com/fpp2/arima.html

GROUP MEMBERS:

1) **Shreyas Rathod** _____
2) **Anup Singh** _____
3) **Anshu Bhagat** _____
4) **Vikas Dubey** _____
5) **Shivam Dubey** _____

SUPERVISOR:

Dr. Anand Khandare _____