Finding the best sector of Nifty to invest in And predicting its Future Price

NAME- SOURAJ CHAKRABORTY

CONTACT DETAILS- officialsouraj.bhu.ac.in@gmail.com

LINKDIN ID- MY LINKDIN ID

COURSE PURSUING- M.SC STATISTICS

TOPIC	Pg. No.
1) Executive Summary of the Project	3
2) Introduction	4
3) Datasets	5-9
a) Details of Dataset	9
4) Methodology	9-15
a) Descriptive Analysis	10-12
b) Time Series Analysis	12-15
5) Analysis	15-25
a) Descriptive Analysis	15-18
b) Time Series Analysis	19-26
• Moving Average Method	19-20
• Decomposition of the data	21-23
• Mathematical Curve Fitting	23-25
• Exponential Smoothing	25-26
6) Result	26-29
a) Accuracy	26-27
b) Forecast	28-29
7) Conclusion from analysis	30
8) Reference	31

Executive Summary of the Project:-

In this project, I studied and analyzed Nifty50 and 5 of its sectoral indices in order to find out which sector is the best to invest in the long run. So, I did a descriptive analysis of the returns . After finding the best sector, I did a detailed time series analysis of that sector along with Nifty 50.
Finally, I used Time series forecasting method and forecasted where the indices might go in the future.
I made an assumption that in the next 10 years there will be no big economic collapse (like dot-com bubble in late ‘90s, Housing market crash ’08) that can cause an unexpected and havoc change in the stock market.

INTRODUCTION

A time series is a collection of observations made sequentially through time. Examples occur in a variety of fields, ranging from economics to engineering. The methods of analyzing time series constitute an important area of statistics.

Nifty 50:- NIFTY 50 is a benchmark Indian stock market index that represents the weighted average of 50 of the largest Indian companies listed on the National Stock Exchange. It is one of the two main stock indices used in India, the other being the BSE SENSEX. The NIFTY 50 Index gives a weightage of 39.47% to financial services, 15.31% to Energy, 13.01% to IT, 12.38% to consumer goods, 6.11% to Automobiles and 0% to the agricultural sector.

Sectoral Indices:-The sectoral indices represent specific sectors and give benchmarking data of those sectors in the market. For the purpose of sectoral indices, various sectors identified are energy, healthcare, automobile, consumer products, technology & communications, and financial. Let's understand this with an example – Bank NIFTY in NSE's sectoral index represents the overall performance of the Indian banking sector. Sectoral indices are reviewed on a semi-annual basis ending January and July.

- **NIFTY Auto Index: -** This index reflects the behaviors and performance of automobile sector such as manufacturers of cars & motorcycles, heavy vehicles, auto ancillaries, tyres etc. It comprises of maximum 15 stocks.
- **NIFTY Bank Index: -**Comprised of maximum 12 banking stocks, this index is designed to indicate the large and liquid banks.
- **NIFTY IT Index:-**Comprising the stocks of maximum 20 IT companies, The index is designed to reflect the behavior of companies engaged into activities such as IT infrastructure, IT education and software training, networking infrastructure, software development, hardware, IT support and maintenance etc.
- **NIFTY Metal Index: -**It reflects the behavior and performance of the Metals sector including mining with maximum 15 stocks considered for index construction.
- **NIFTY Pharma Index: -** This indicates the behavior and performance pharmaceutical companies. The index comprises of maximum of 10 stocks.
- **NIFTY Energy Index: -** this sector is designed to reflect the behavior and performance of a diversified portfolio of companies representing the commodities segment which includes sectors such as Petroleum, Gas and Power etc.

DATASET
Average return

Year	*Nifty Bank*	*Nifty IT*	*Nifty Pharma*	*Nifty Auto*	*Nifty Metal*	*Nifty Energy*	Nifty 50
April'11-Mar'12	-10.51	-3.01	10.03	20.76	-21.69	-23.23	-8.72
April'12-Mar'13	22.23	18.63	27.05	4.5	-25.96	5.75	13
April'13-Mar'14	2.34	53.74	23.66	26.64	13.73	4.73	12.92
April'14-Mar'15	42.64	30.94	48.86	39.58	-3.95	0.25	22.18
April'15-Mar'16	-8.41	1.77	-6.34	1.23	-14.84	0.87	-4.05
April'16-Mar'17	33.12	-11.19	-10.38	23.72	45.39	45.91	18.53
April'17-Mar'18	14.19	40.65	-10.48	13.36	26.91	12.29	15.43
April'18-Mar'19	16.58	19.44	3.77	-28.17	-18.17	19.82	9.39
April'19-Mar'20	-27.65	-15.55	-0.8	-29.33	-39.79	-20.27	-16.07
April'20-Mar'21	52.23	81.91	44.41	63.37	160.66	37.07	48.39
April'21-Mar'22	10.09	23.21	-0.05	14.91	30.51	58.58	16.89

Monthly price and change% from Apr’11-May’22

	NIFTY 50		**NIFTY-IT**	
Date	**Close**	**Change%**	**Close**	**Change %**
May-11	**5560.15**	**-0.032**	**6,718.35**	**-6.01%**
Jun-11	**5647.4**	**0.0157**	**6,538.50**	**-2.68%**
Jul-11	**5482**	**-0.029**	**6,624.70**	**1.32%**
Aug-11	**5001**	**-0.087**	**6,335.10**	**-4.37%**
Sep-11	**4943.25**	**-0.011**	**5,451.25**	**-13.95%**
Oct-11	**5326.6**	**0.0776**	**5,678.90**	**4.18%**
Nov-11	**4832.05**	**-0.092**	**6,278.70**	**10.56%**
Dec-11	**4624.3**	**-0.043**	**5,893.25**	**-6.14%**
Jan-12	**5199.25**	**0.1243**	**6,139.00**	**4.17%**
Feb-12	**5385.2**	**0.0358**	**6,193.85**	**0.89%**
Mar-12	**5295.55**	**-0.016**	**6,606.85**	**6.67%**

Apr-12	5248.15	-0.009	6,516.00	-1.38%
May-12	4924.25	-0.061	6,085.40	-6.61%
Jun-12	5278.9	0.072	6,008.80	-1.26%
Jul-12	5229	-0.009	6,144.60	2.26%
Aug-12	5258.5	0.0056	5,695.25	-7.31%
Sep-12	5703.3	0.0846	6,072.35	6.62%
Oct-12	5619.7	-0.014	6,313.80	3.98%
Nov-12	5879.85	0.0463	6,087.85	-3.58%
Dec-12	5905.1	0.0043	6,263.25	2.88%
Jan-13	6034.75	0.022	6,024.95	-3.80%
Feb-13	5693.05	-0.056	6,778.00	12.50%
Mar-13	5682.55	-0.001	7,106.65	4.85%
Apr-13	5930.2	0.0436	7,219.05	1.58%
May-13	5985.95	0.0094	6,047.70	-16.23%
Jun-13	5842.2	-0.024	6,472.05	7.02%
Jul-13	5742	-0.017	6,634.15	2.50%
Aug-13	5471.8	-0.047	7,787.35	17.38%
Sep-13	5735.3	0.0482	8,382.40	7.64%
Oct-13	6299.15	0.0983	8,167.80	-2.56%
Nov-13	6176.1	-0.019	8,852.80	8.39%
Dec-13	6304	0.0207	8,820.75	-0.36%
Jan-14	6089.5	-0.034	9,517.85	7.90%
Feb-14	6276.95	0.0308	9,957.45	4.62%
Mar-14	6704.2	0.0681	10,338.55	3.83%
Apr-14	6696.4	-0.001	9,298.00	-10.06%
May-14	7229.95	0.0797	9,227.95	-0.75%
Jun-14	7611.35	0.0528	8,970.30	-2.79%
Jul-14	7721.3	0.0144	9,912.30	10.50%
Aug-14	7954.35	0.0302	10,304.70	3.96%
Sep-14	7964.8	0.0013	10,679.65	3.64%
Oct-14	8322.2	0.0449	11,302.70	5.83%
Nov-14	8588.25	0.032	11,341.05	0.34%
Dec-14	8282.7	-0.035	11,898.05	4.91%
Jan-15	8808.9	0.0635	11,216.30	-5.73%
Feb-15	8901.85	0.0106	11,824.75	5.42%
Mar-15	8491	0.0462	12,659.80	7.06%
Apr-15	8181.5	-0.036	12,083.00	-4.56%
May-15	8433.65	0.0308	11,001.05	-8.95%

Jun-15	8368.5	-0.007	11,575.05	5.22%
Jul-15	8532.85	0.0196	11,037.40	-4.64%
Aug-15	7971.3	-0.065	11,594.15	5.04%
Sep-15	7948.9	-0.002	11,605.70	0.10%
Oct-15	8065.8	0.0147	12,032.10	3.67%
Nov-15	7935.25	-0.016	11,486.90	-4.53%
Dec-15	7946.35	0.0014	11,206.00	-2.45%
Jan-16	7563.55	-0.048	11,212.55	0.06%
Feb-16	6987.05	-0.076	11,236.15	0.21%
Mar-16	7738.4	0.1075	10,278.40	-8.52%
Apr-16	7849.8	0.0144	11,309.30	10.03%
May-16	8160.1	0.0395	11,196.25	-1.00%
Jun-16	8287.75	0.0156	11,395.85	1.78%
Jul-16	8638.5	0.0423	11,120.15	-2.42%
Aug-16	8786.2	0.0171	10,913.30	-1.86%
Sep-16	8611.15	-0.019	10,546.10	-3.36%
Oct-16	8625.7	0.0017	10,292.30	-2.41%
Nov-16	8224.5	-0.046	10,082.70	-2.04%
Dec-16	8185.8	-0.004	10,087.70	0.05%
Jan-17	8561.3	0.0459	10,399.25	3.09%
Feb-17	8879.6	0.0372	9,848.50	-5.30%
Mar-17	9173.75	0.0331	10,680.95	8.45%
Apr-17	9304.05	0.0142	10,703.25	0.21%
May-17	9621.25	0.0341	9,943.70	-7.10%
Jun-17	9520.9	-0.010	10,549.10	6.09%
Jul-17	10077.1	0.0584	10,155.05	-3.74%
Aug-17	9917.9	-0.015	10,755.75	5.92%
Sep-17	9788.6	-0.013	10,558.25	-1.84%
Oct-17	10335.3	0.0559	10,475.35	-0.79%
Nov-17	10226.55	-0.010	10,837.90	3.46%
Dec-17	10530.7	0.0297	11,115.35	2.56%
Jan-18	11027.7	0.0472	11,665.75	4.95%
Feb-18	10492.85	-0.048	12,986.40	11.32%
Mar-18	10113.7	-0.036	12,809.00	-1.37%
Apr-18	10739.35	0.0619	12,511.55	-2.32%
May-18	10736.15	-0.000	13,986.25	11.79%
Jun-18	10714.3	-0.002	13,666.00	-2.29%
Jul-18	11356.5	0.0599	13,989.50	2.37%

Aug-18	11680.5	0.0285	14,587.80	4.28%
Sep-18	10930.45	-0.064	15,811.40	8.39%
Oct-18	10386.6	-0.049	15,838.05	0.17%
Nov-18	10876.75	0.0472	14,940.10	-5.67%
Dec-18	10862.55	-0.001	14,638.05	-2.02%
Jan-19	10830.95	-0.002	14,440.30	-1.35%
Feb-19	10792.5	-0.003	15,499.30	7.33%
Mar-19	11623.9	0.077	15,732.00	1.50%
Apr-19	11748.15	0.0107	15,628.20	-0.66%
May-19	11922.8	0.0149	16,705.40	6.89%
Jun-19	11788.85	-0.011	16,160.65	-3.26%
Jul-19	11118	-0.056	15,936.45	-1.39%
Aug-19	11023.25	-0.008	15,620.20	-1.98%
Sep-19	11474.45	0.0409	16,010.40	2.50%
Oct-19	11877.45	0.0351	15,540.15	-2.94%
Nov-19	12056.05	0.015	15,559.40	0.12%
Dec-19	12168.45	0.0093	14,998.05	-3.61%
Jan-20	11962.1	-0.017	15,652.40	4.36%
Feb-20	11201.75	-0.063	16,144.15	3.14%
Mar-20	8597.75	-0.232	15,212.95	-5.77%
Apr-20	9859.9	0.1468	12,763.65	-16.10%
May-20	9580.3	-0.028	14,108.40	10.54%
Jun-20	10302.1	0.0753	14,010.50	-0.69%
Jul-20	11073.45	0.0749	14,754.30	5.31%
Aug-20	11387.5	0.0284	18,071.85	22.49%
Sep-20	11247.55	-0.012	17,928.85	-0.79%
Oct-20	11642.4	0.0351	19,951.35	11.28%
Nov-20	12968.95	0.1139	20,916.85	4.84%
Dec-20	13981.75	0.0781	21,764.90	4.05%
Jan-21	13634.6	-0.024	24,251.35	11.42%
Feb-21	14529.15	0.0656	24,645.75	1.63%
Mar-21	14690.7	0.011	24,301.45	-1.40%
Apr-21	14631.1	-0.004	25,855.00	6.39%
May-21	15582.8	0.065	25,664.45	-0.74%
Jun-21	15721.5	0.0089	27,115.05	5.65%
Jul-21	15763.05	0.0026	29,168.00	7.57%
Aug-21	17132.2	0.0869	30,480.05	4.50%
Sep-21	17618.15	0.0284	34,570.20	13.42%

Oct-21	**17671.65**	**0.003**	**35,028.00**	**1.32%**
Nov-21	**16983.2**	**-0.039**	**34,408.75**	**-1.77%**
Dec-21	**17354.05**	**0.0218**	**35,043.75**	**1.85%**
Jan-22	**17339.85**	**-0.000**	**38,701.05**	**10.44%**
Feb-22	**16793.9**	**-0.031**	**34,824.55**	**-10.02%**
Mar-22	**17464.75**	**0.0399**	**33,847.85**	**-2.80%**

DATASET Details

Description of columns in the file:

- **Date** — Month of trade
- **Close** — Refers to the prior day's final price when the market officially closes for the day.
- **%change-**Refers to the % change in the price at the end of the month. A negative value indicates that prices went down from where it started.
- Negative value in the average return table indicates that the prices of the index went down from where it started that year

- **METHODOLOGY**

Δ Descriptive Analysis

Diagrams like graphs, charts, maps, pictures etc. are attractive and effective means for presentation of statistical data. It is more effective than tabular representation, being easily intelligible to a layman. Indeed, diagrams are almost essential whenever it is required to convey any statistical information to the general public. Diagrams are readily capable of revealing some features of the exhibited data.

Line Diagram

This diagram is meant for representing chronological data. In fact, it exhibits the relationship of the variable (e.g. sales of coffee of a company, price of a product) may be specified for individual points of time or for different period of time In constructing a line diagram, two axis of co-ordinates are taken, the horizontal one for time and the vertical one for variable. The scale for each axis is then selected and the data are plotted as different points on the plane, the plotting of variable values being done against points of time or mid-points of the time interval (for time period). The successive points are now

joined by straight line segments and the chart so obtained is called a line diagram for the given data. Two or mutually related time series data having same unit of measurement can be represented using the same axis of co-ordinates, by drawing a number of line diagrams, one for each series. These different line diagrams are mutually distinguished by using distinct pattern of lines such as broken lines, dotted lines or multiple coloured lines. The resulting diagram is known as a Multiple Line Diagram. It is used for comparing two mutually related time series data e.g. if we want to compare the literacy rates for a number countries last 15 years, say, we may draw multiple line diagram.

Bar Diagram

Another mode of diagrammatic representation of data is the use of bar diagrams. These have more general applicability than line diagrams in the sense that they may be used for series varying either over time or over space. In this method bars of equal width are taken for the different items of the series, drawn over base line. The length or height of a bar representing the value of the variable concerned. It is preferable to take the bars horizontally for data varying over space and vertically in the case of a series varying over time. We can compare the different items of the series by visualizing bar diagram.

Multiple Bar Diagram

In a multiple bars diagram two or more sets of inter-related data are represented (multiple bar diagram faciliates comparison between more than one phenomena). The technique of making a simple bar chart is used to draw this diagram but the difference is that we use different shades, colors, or dots to distinguish between different phenomena. *We use to draw multiple bar charts if the total of different phenomena is meaningless.*

Mean

It is the most common measure of location (A measure of location is a measure of the center of a batch of a numbers).It is defined as –

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \text{ i.e. } \overline{x} = \frac{Sum\ of\ the\ obsevations}{Total\ No.\ of\ observations}.$$

Variance

It is the most common measure of dispersion (A measure of dispersion gives a numerical indication of the “scatteredness” of a batch of numbers). It is defined as –

$$Var(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2.$$

Δ Method of Time Series Analysis

Simple Moving Average

A moving average is a calculation to analyze the data points by creating a series of averages of different subsets of the full dataset. The simple moving average of period k of a time series gives us a new series of arithmetic means, each of k successive observations of the time series. We start with the first k observations. At the next stage, we leave the first and include the (k+1)st observation. This process is repeated until we arrive at the last k observations. Each of these means is centred against the time which is the mid-point of the time interval included in the calculation of the moving average. Thus when k, the period of the moving average, is odd, the moving average values correspond to tabulated time values for which the time series is given. When the period is even, the moving average fails midway between two tabulated values. In this case, we calculate a subsequent two-item moving average to make the resulting moving average values correspond to the tabulated time periods.Moving averages are an important analytical tool used to identify current trends and the potential for a change in an established trend.

Mean absolute deviation (MAD)

Mean absolute deviation (MAD) of a data set is the average distance between each data value and the mean. Mean absolute deviation is a way to describe variation in a data set. Mean absolute deviation helps us get a sense of how "spread out" the values in a data set are.

Mean absolute percentage error (MAPE)

The mean absolute percentage error (MAPE) is the mean or average of the absolute percentage errors of forecasts. Error is defined as actual or observed value minus the forecasted value. Percentage errors are summed without regard to sign to compute MAPE.

Mean signed difference

The mean signed difference, deviation, or error (MSD) is sample statistic that summarizes how well a set of estimates match the quantities. that they are supposed to estimate.

Thus (100-MAPE)% is a measure of **accuracy** of the fitted model.

- **X11 Decomposition**

X11 method is one of the ways to decompose quarterly and monthly data.In X11 method, trend-cycle estimates are available for all observations including the end points, and the seasonal component is allowed to vary slowly over time. X11 also has some sophisticated methods for handling trading day variation, holiday effects and the effects of known predictors. It handles both additive and multiplicative decomposition. The process is entirely automatic and tends to be highly robust to outliers and level shifts in the time series.

Methods of Time Series Analysis

Mathematical Curve Fitting

It is an essential part of the concept of trend that the movement over fairly long periods is smooth. This is perhaps the best and most rational method of determining the trend. In this case, a suitable trend equation is selected and then the constants involved in the equation are estimated on the basis of the data in hand. After derived the estimated trend equation we can predict the future trend values.

Polynomial Trend Line

Here we assume that our suitable trend equation is a polynomial in time element 't'. So here we consider a mathematical model:-

Yt = a0 + a1t +a2t2 + a3t3 + ….. + aptp + et

where ai's, i = 1(1)p are constants and et denotes a random error with E(et) = 0 and V(et) = σ2

Now we can estimate the constants ai's by method of least squares and fit the polynomial model. In this method the constants are determined by minimizing,

$$S=\Sigma(yt-a0-a1t-a2t2-\cdots-aptp)2t$$

The normal equations are,

$$\frac{\partial S}{\partial a_j} = 0, \forall\, j = 0(1)p$$

$$\Rightarrow \sum_t t^j y_t = a_0 \sum_t t^j + a_1 \sum_t t^{j+1} + a_2 \sum_t t^{j+2} + \cdots + a_p \sum_t t^{j+p}, j = 0(1)p$$

By solving the normal equations we can obtain the estimates of the constants a0, a1, a2,…, ap, and fit a trend equation on the time element 't'.

Growth Curves

The family of curves (polynomials) described above represents a simple and very useful type but a curve of

$$\mathbf{Y_t = a_0 + a_1 t} \text{ or } \mathbf{Y_t = a_0 + a_1 t + a_2 t^2}$$

etc. may not be a satisfactory description of the trend of some time series for the period shown or for the prediction purpose also. Perhaps of even greater general utility, in the analysis of time series, are curves of a semi-logarithmic (exponential) type.

Exponential Curve

The simplest exponential curve may be written as

$$\mathbf{Y_t = ab^t} \text{ , where } a > 0 \text{ , } b > 0$$

Now taking logarithm both side,

$$\mathbf{\log Y_t = \log a + t \times (\log b)} \text{ , which is a straight line in t}$$

So, here we assume a mathematical model

$$\mathbf{\log Y_t = \log a + t \times (\log b) + e_t,}$$

where e_t denotes a random error with

$$\mathbf{E(e_t) = 0} \text{ and } \mathbf{V(e_t) = \sigma^2}$$

Now we can apply a least squares method to the logarithm of the original data (Y_t) to estimate a and b and fit an exponential trend equation.

Exponential Smoothing

Forecasts produced using exponential smoothing methods are weighted averages of past observations, with the weights decaying exponentially as the observations get older. In other words, the more recent the observation the higher the associated weight. This framework generates reliable forecasts quickly and for a wide range of time series, which is a great advantage and of major importance to applications in industry.

Estimation using ETS(.,.,.) Model

The ETS(·,·,··,·,·) model cstands for (Error, Trend, Seasonal). This label can also be thought of as ExponenTial Smoothing. In R, the ets() function estimates the model parameters and returns information about the fitted model. The R code below shows the most important arguments that this function can take, and their default values. If

only the time series is specified, and all other arguments are left at their default values, then an appropriate model will be selected automatically.

ets(y, model="ZZZ", damped=NULL, alpha=NULL, beta=NULL, gamma=NULL, phi=NULL, lambda=NULL, biasadj=FALSE, additive.only=FALSE, restrict=TRUE, allow.multiplicative.trend=FALSE)

y

The time series to be forecast.

model

A three-letter code indicating the model to be estimated using the ETS classification and notation. The possible inputs are “N” for none, “A” for additive, “M” for multiplicative, or “Z” for automatic selection. If any of the inputs is left as “Z”, then this component is selected according to the information criterion. The default value of ZZZ ensures that all components are selected using the information criterion.

damped

If damped=TRUE, then a damped trend will be used (either A or M).
If damped=FALSE, then a non-damped trend will used. If damped=NULL (the default), then either a damped or a non-damped trend will be selected, depending on which model has the smallest value for the information criterion.

alpha, beta, gamma, phi

The values of the smoothing parameters can be specified using these arguments. If they are set to NULL (the default setting for each of them), the parameters are estimated.

lambda

Box-Cox transformation parameter. It will be ignored if lambda=NULL (the default value). Otherwise, the time series will be transformed before the model is estimated. When lambda is not NULL, additive.only is set to TRUE.

biasadj

If TRUE and lambda is not NULL, then the back-transformed fitted values and forecasts will be bias-adjusted.

additive.only

Only models with additive components will be considered if additive.only=TRUE. Otherwise, all models will be considered.

restrict

If restrict=TRUE (the default), the models that cause numerical difficulties are not considered in model selection.

allow.multiplicative.trend

Multiplicative trend models are also available. Set this argument to TRUE to allow these models to be considered.

- **Analysis:-**
- **Descriptive Analysis**
  - **Mean**

The average return of all the sectors is calculated and tabulated

	Index						
Year	***Nifty Bank***	***Nifty IT***	***Nifty Pharma***	***Nifty Auto***	***Nifty Metal***	***Nifty Energy***	**Nifty 50**
April'11-Mar'12	-10.51	-3.01	10.03	20.76	-21.69	-23.23	-8.72
April'12-Mar'13	22.23	18.63	27.05	4.5	-25.96	5.75	13
April'13-Mar'14	2.34	53.74	23.66	26.64	13.73	4.73	12.92
April'14-Mar'15	42.64	30.94	48.86	39.58	-3.95	0.25	22.18
April'15-Mar'16	-8.41	1.77	-6.34	1.23	-14.84	0.87	-4.05
April'16-Mar'17	33.12	-11.19	-10.38	23.72	45.39	45.91	18.53
April'17-Mar'18	14.19	40.65	-10.48	13.36	26.91	12.29	15.43
April'18-Mar'19	16.58	19.44	3.77	-28.17	-18.17	19.82	9.39
April'19-Mar'20	-27.65	-15.55	-0.8	-29.33	-39.79	-20.27	-16.07
April'20-Mar'21	52.23	81.91	44.41	63.37	160.66	37.07	48.39
April'21-Mar'22	10.09	23.21	-0.05	14.91	30.51	58.58	16.89
Mean	13.35	21.8672727	11.79363636	13.68818182	13.89090909	12.88818182	11.6263636

Conclusion

On an average 11-13 %change can be seen in every sector, but It which shows a higher %change of ~21%. Hence, IT sector outperformed others in the long run.

- **Variance And Standard Deviation**

The variance and s.d. of the sectors and Nifty are –

	Nifty Bank	***Nifty IT***	***Nifty Pharma***	***Nifty Auto***	***Nifty Metal***	***Nifty Energy***	**Nifty 50**
Std. Deviation	23.88733807	29.30852132	21.19972654	27.02030304	55.57560318	25.76239306	17.28626928
Sample Variance	570.60492	858.9894218	449.4284055	730.0967764	3088.647669	663.7008964	298.8151055

Conclusion

The variance value is very high showing the risk involved, with Nifty showing the lowest variability and IT the second highest. Thus, one can say that though the IT sector gave a good return, it is also a very risky sector to bet on. To analyse the unexpectedly high variability of Metal sector, we move to the diagrammatic representation of the data.

⍰ Studying and Comparing Data Using Their Graphical Representation:-

For studying and comparing the returns of Nifty50 and the different indices Line Diagrams and Bar Diagrams are made.

1. Bar Diagram

To get a visual representation of the average return, we compared the mean %change of the indices and Nifty50.

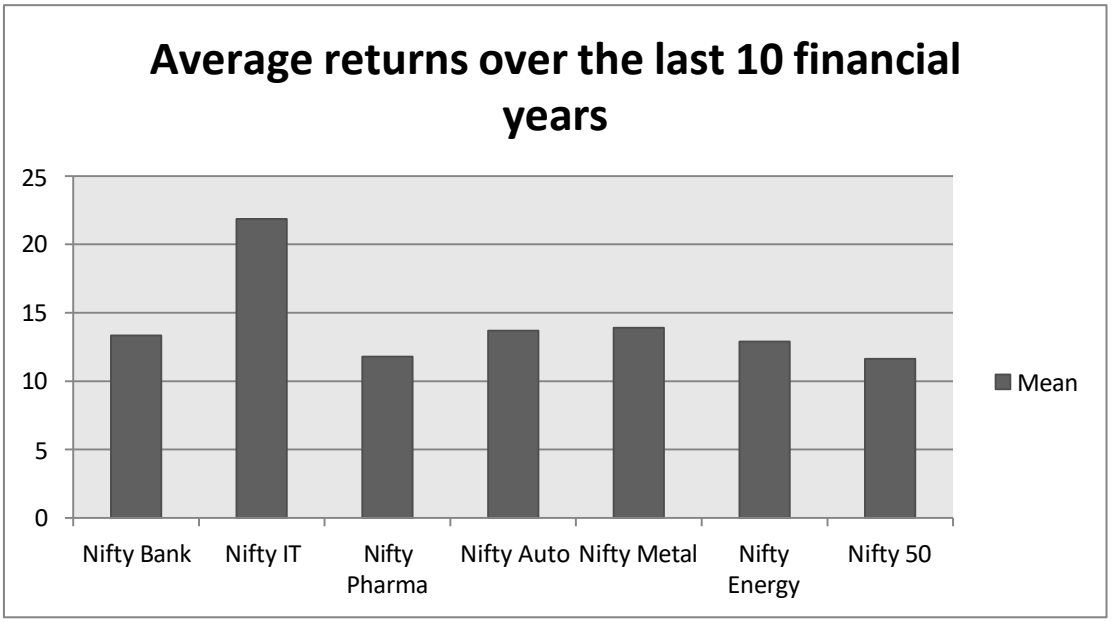


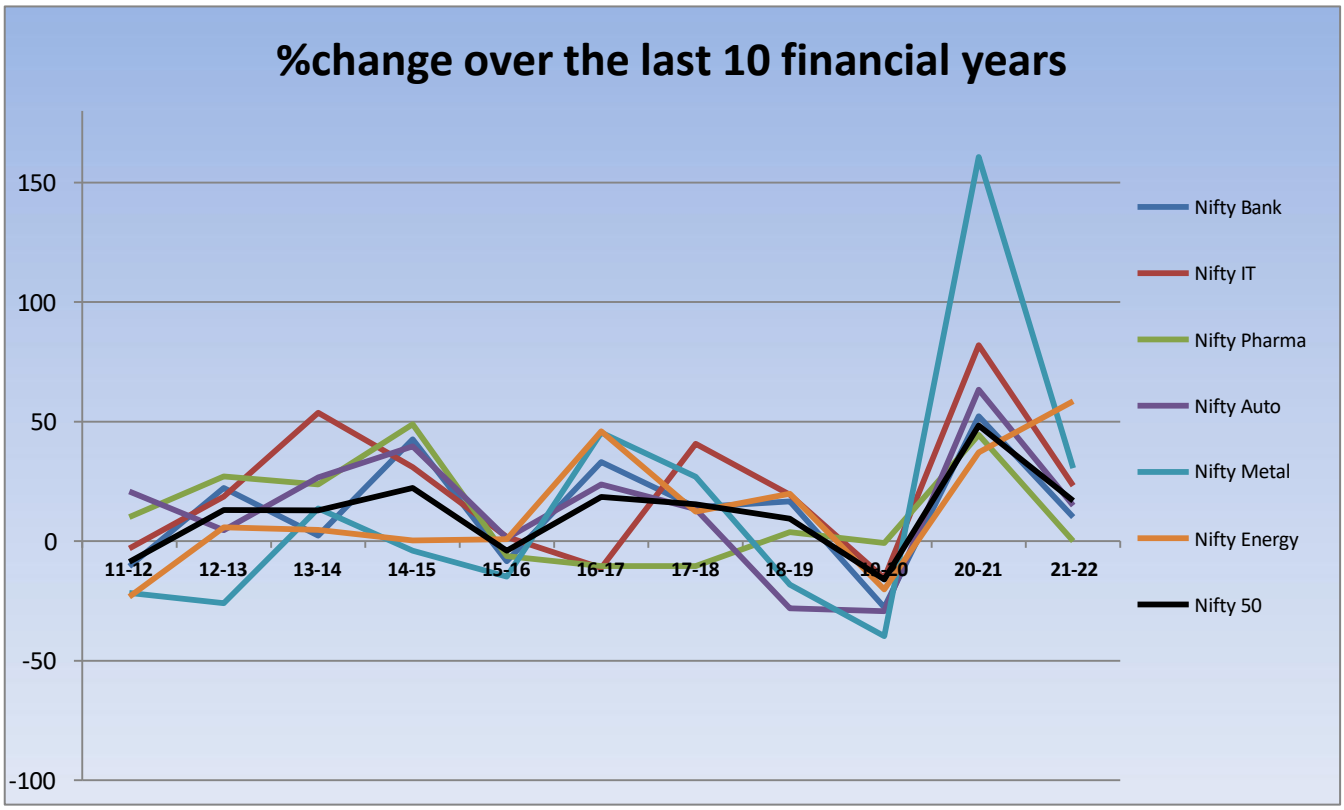
FIG.1.1

Conclusion:-

Clearly, over the last 10 financial years the average return from IT sector was highest followed by the Metal sector. Also all the sectors but nifty pharma outperformed Nifty 50. So it can be concluded that on long run, it is better to stick with IT sector.

2. Line diagram

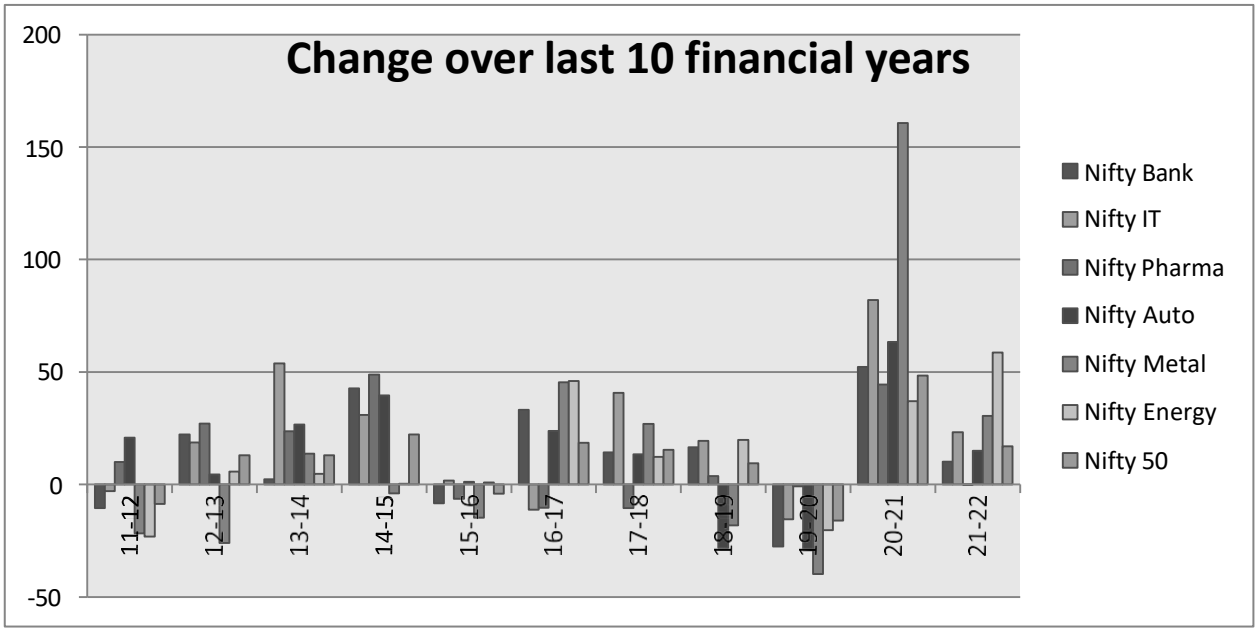
To get a visual representation of the returns over the past decade, we draw a line diagram of the %change .



Conclusion-

All sectors showed a pretty same pattern in the change % with IT sector giving +ve returns when others start to fall and vice-versa. But in 2020 all sectors fell with pharma showing a slight rise. This is probably due to the lockdown effect when market anticipated that economy will fall but medical sectors will rise due to the covid situation. Another surprising observation was the change% of metal sector. It saw a rise of about 150% during APR'20-MAR'21, outperforming the IT sector. This was due to the resuming of business post-lockdown and the shortage of supply from China , who has been cutting production due to carbon emission target.

3. Multiple Bar Diagram



Conclusion

Normally on a yearly (financial year) basis, there are two years of +ve %change(~50%) followed by a year of –ve %change. There was an astonishing +ve change in the year April'20-March'21, possibly due to the lockdown when a lot of people started investing creating a high demand – low supply imbalance, causing the prices to rise exponentially.

So, it is clear from the descriptive analysis of the data that investing in IT sector is better than investing in other sector. So, lets do a time series analysis of the IT sector and the main sector, i.e., NIFTY 50 and try to forecast where the prices may go in the near future, assuming that all the other conditions remain more or less same(like no big economic breakdown or unexpected events like the spread of corona virus,etc.)

❖ TIME SERIES ANALYSIS

Now ,we will use Time Series Analysis to analyse the data on and fit suitable curves to it.

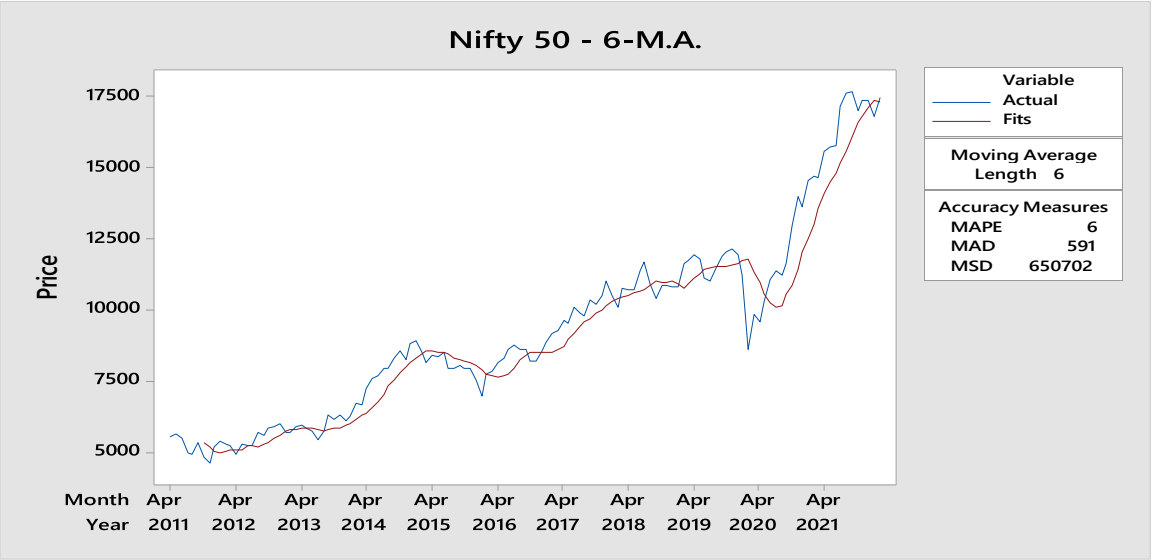
□ Analysis Trend by Simple Moving Average Method:-

At first we will use Simple Moving Average Method for analysing the trend of the annual rice production of India. We obtain simple moving average for 6-months (2nd quarter) and 12-months(4th quarter).

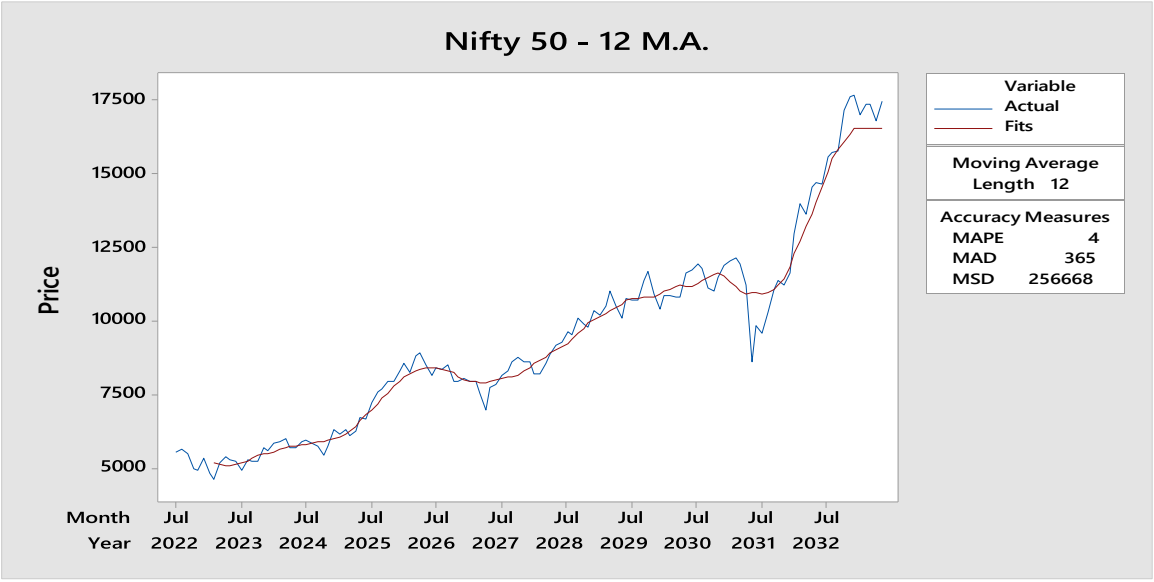
1) Plotting 6M.A. , 12M.A. and the actual data of :

a) NIFTY 50

(i) 6 months Moving Average



(ii) 12 months Moving Average

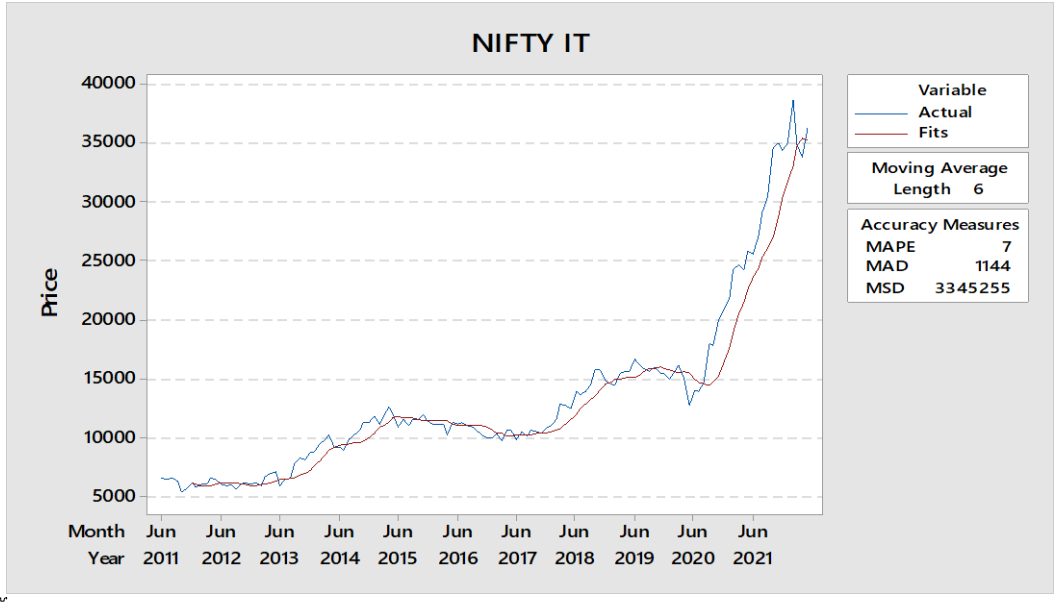


Conclusion

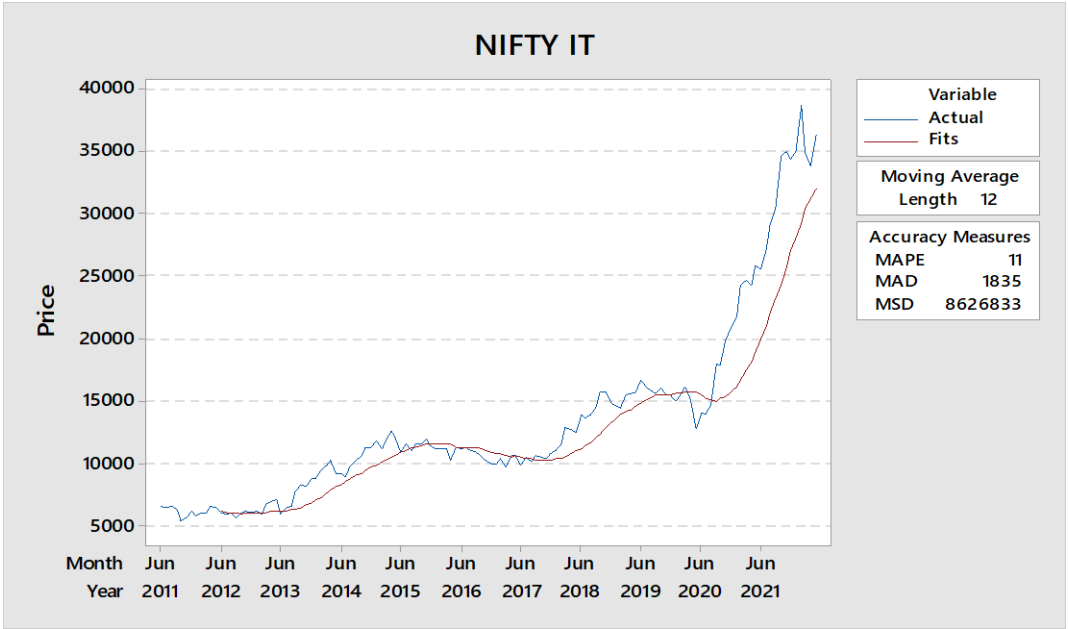
The 6month M.A. has a little lower mape. But the MAD of 6 M.A. is much lower than that of 12M.A. So the 6 months moving average better captures the trend and smoothes out seasonality better.Thus, looking at the 6M.A. we can conclude that there will be no trend in the near future as the trend line flattens but after some period, the uptrend will resume and prices will continue to rise.

2) NIFTY IT

a) 6 months Moving-Average



b) 12 Months Moving Average



Conclusion

Here also both MAPE and MAD of 6-M.A. is lower than that of 12M.A. So the 6 months moving average is better.Hence looking at the 6 months moving average, we can conclude that there will be uptrend, with a few stationary trend

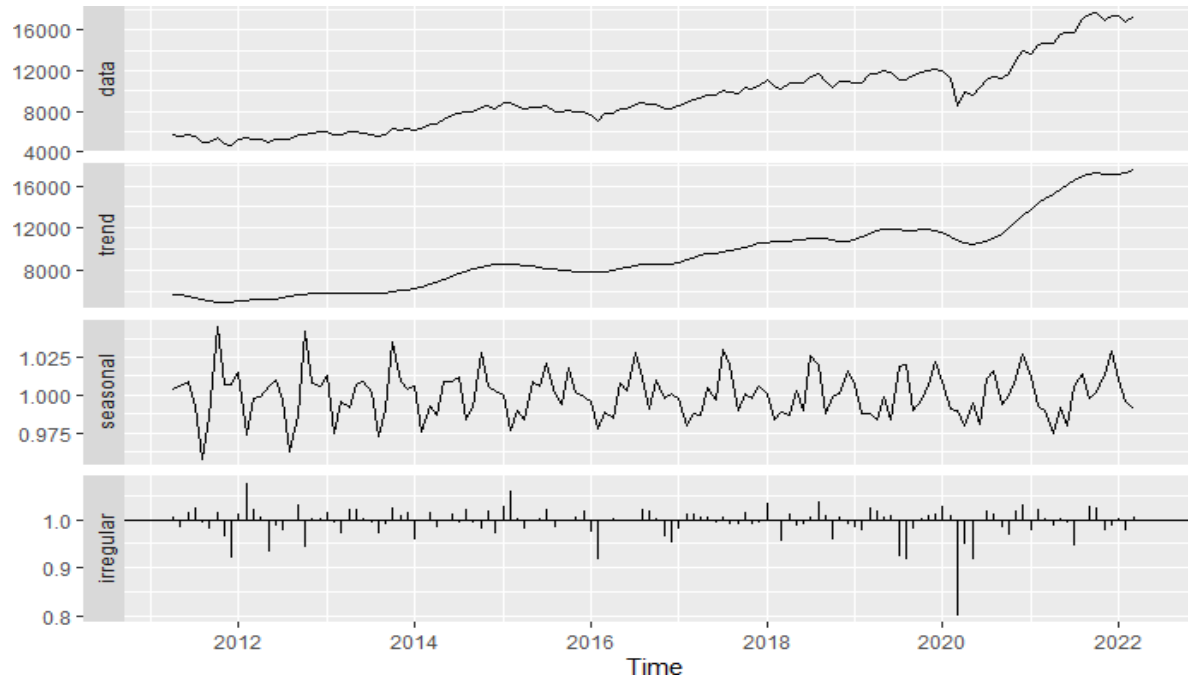
DECOMPOSITION OF TIME SERIES

a) NIFTY 50

i) Code

```
library(fpp2)
library(ggplot2)
library(seasonal)
library(seas)
library(readxl)
Nifty_50_data <- read_excel("Nifty 50 data.xlsx", col_types = c("skip", "numeric", "skip",  "skip", "date"))
Nifty_50_data$DATE=as.Date(Nifty_50_data$DATE,format="%y-%m-%d")
N<-ts(Nifty_50_data$Price,start=c(2011,4),end=c(2022,4),frequency = 12)
d<-seas(N,x11="")
autoplot(d)
```

ii) Result



Conclusion

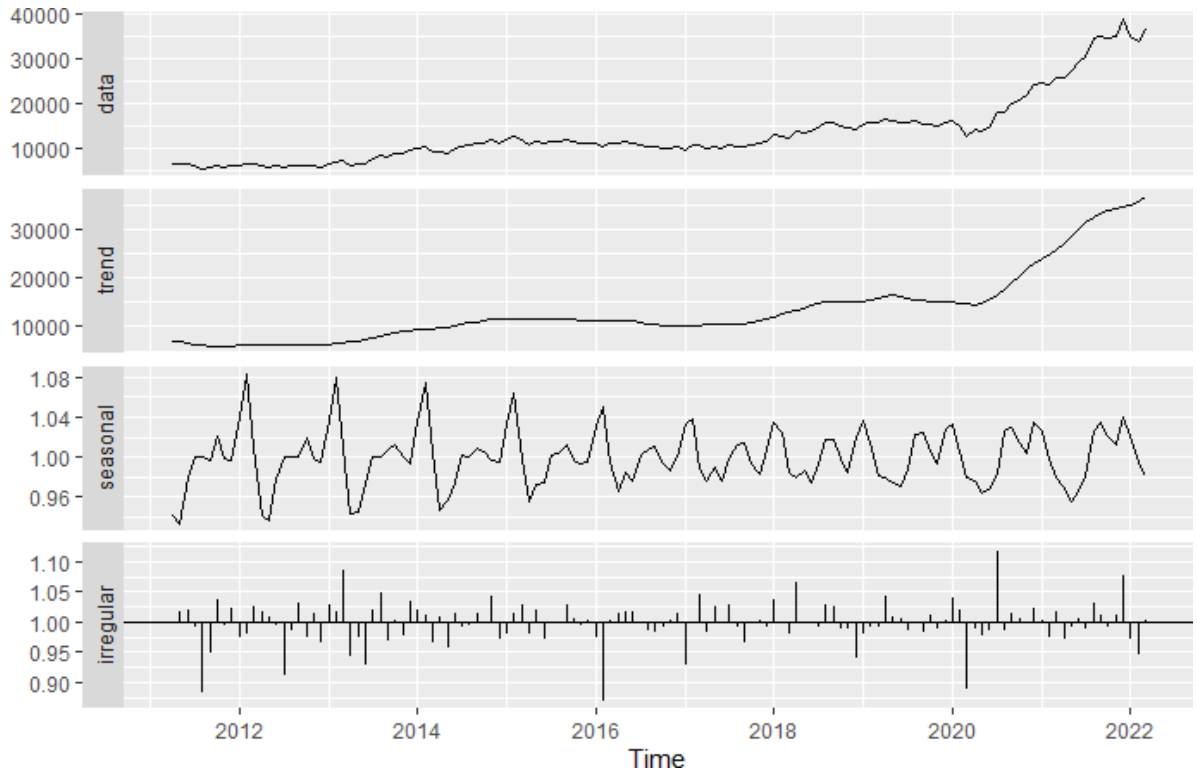
The seasonal component as well as the trend component of the data is not constant . So multiplicative model will be appropriate in this case.

b) NIFTY IT

ii) Code

```
library(fpp2)
library(ggplot2)
library(seasonal)
library(seas)
d<-seas(N,x11="")
autoplot(d)
```

iii) Result



iv) Conclusion

The seasonal component as well as the multiplicative component of the data is not constant. So, here also, a multiplicative model will be appropriate.

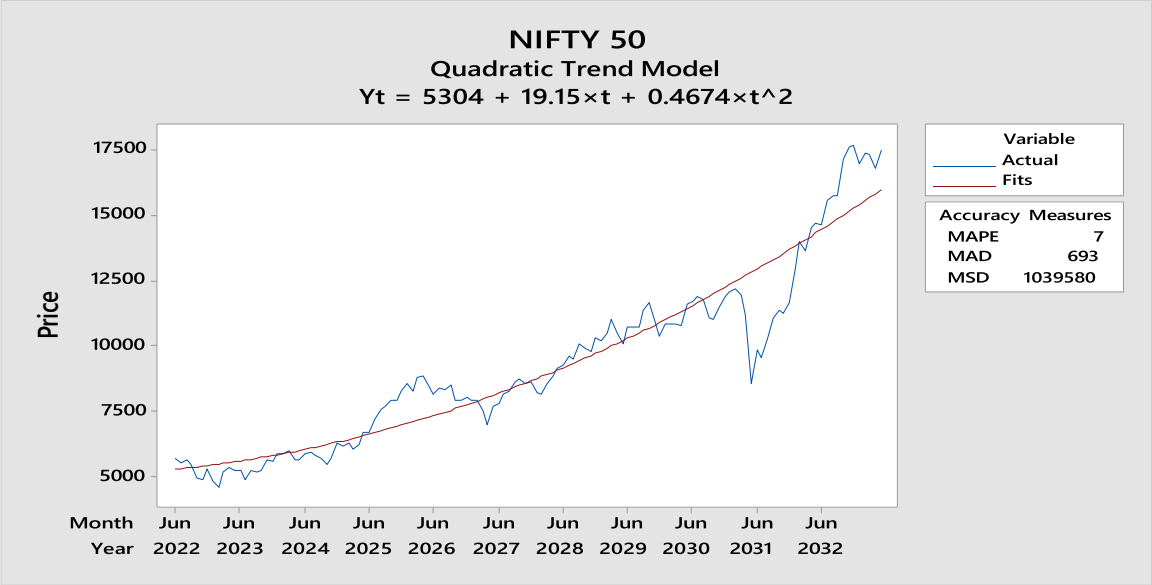
Mathematical Curve Fitting

FITTING APPROPRIATE MATHEMATICAL CURVE

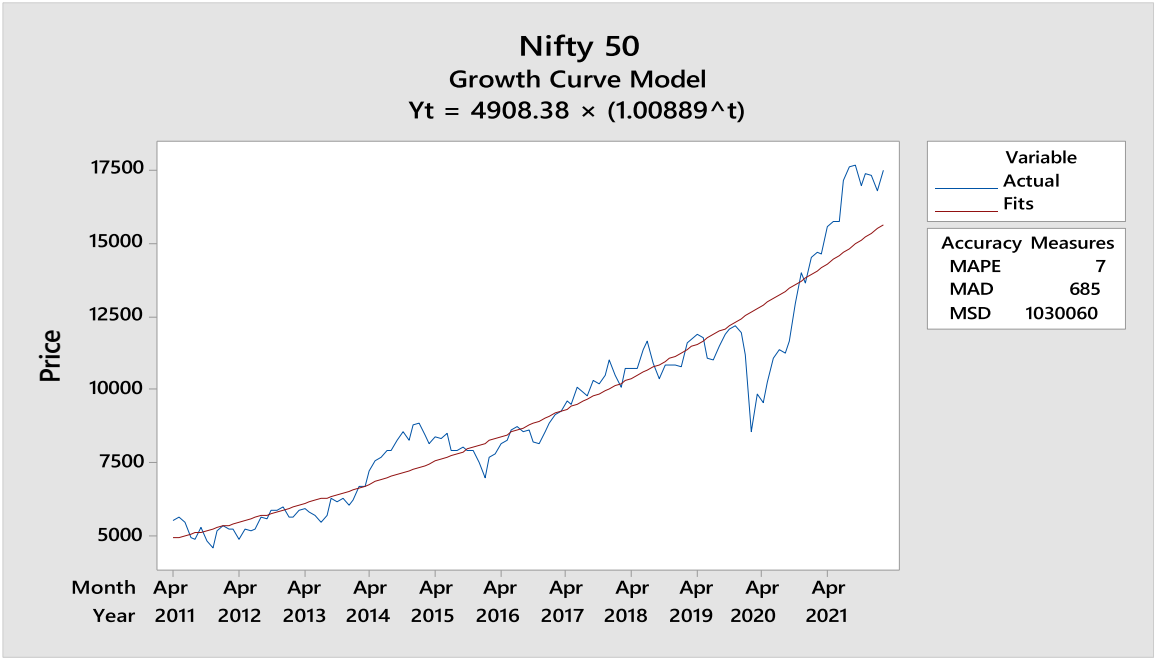
Now we fit an appropriate mathematical curve to the data.

a) NIFTY 50

i) Fitting Quadratic Curve



ii) Fitting Exponential curve

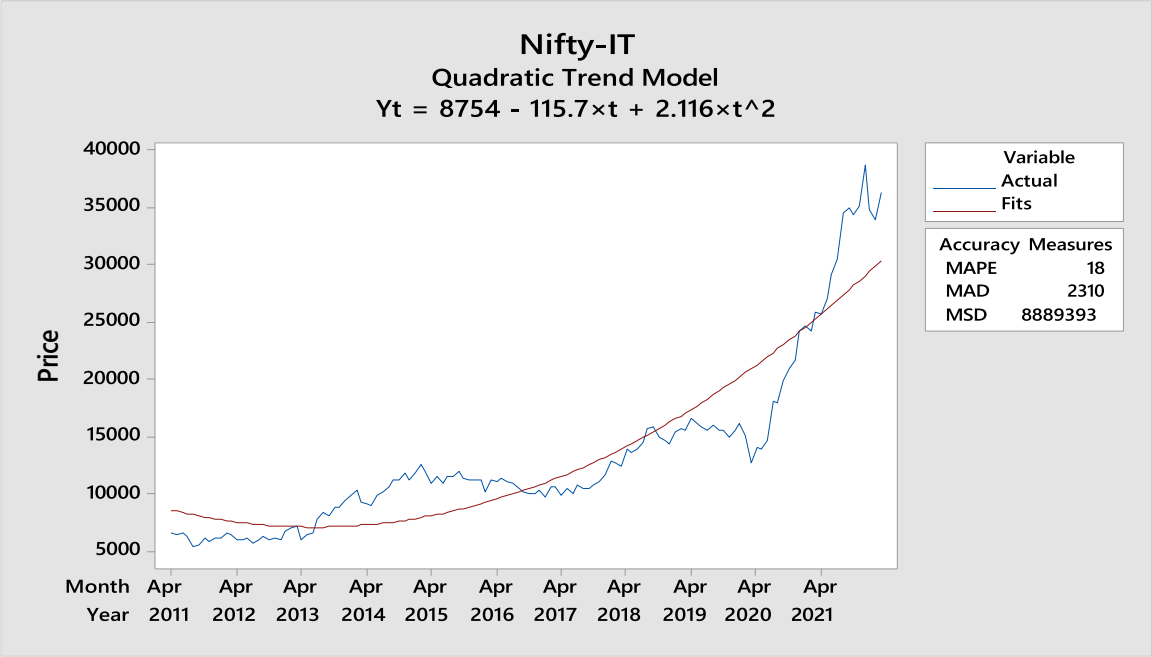


Conclusion

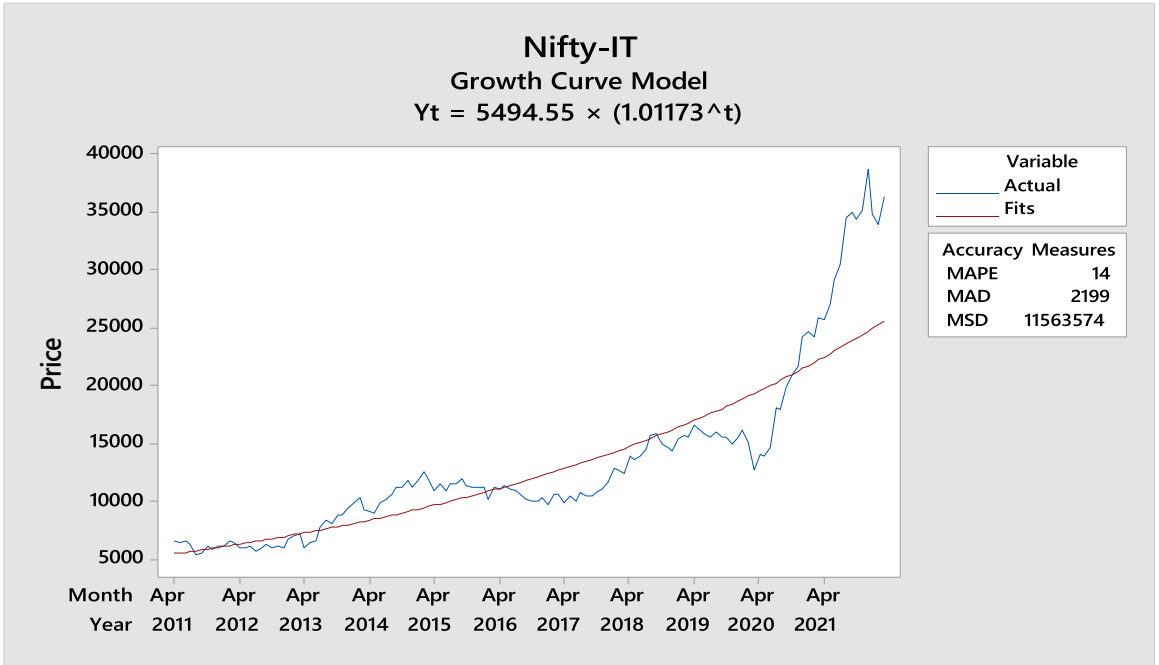
The MAD of the exponential curve is less than that of the quadratic model. So the exponential model is a better fit to the data. Accuracy of the exponential model is **(100-7)% = 93%.**

b) NIFTY IT

i) Fitting Quadratic Model



ii) Fitting Exponential Model



Conclusion

The MSD and MAD of the exponential curve is less than that of the quadratic model. So the exponential model is a better fit to the data. Now accuracy of this model is **(100-14)%=86%.**

Exponential smoothing

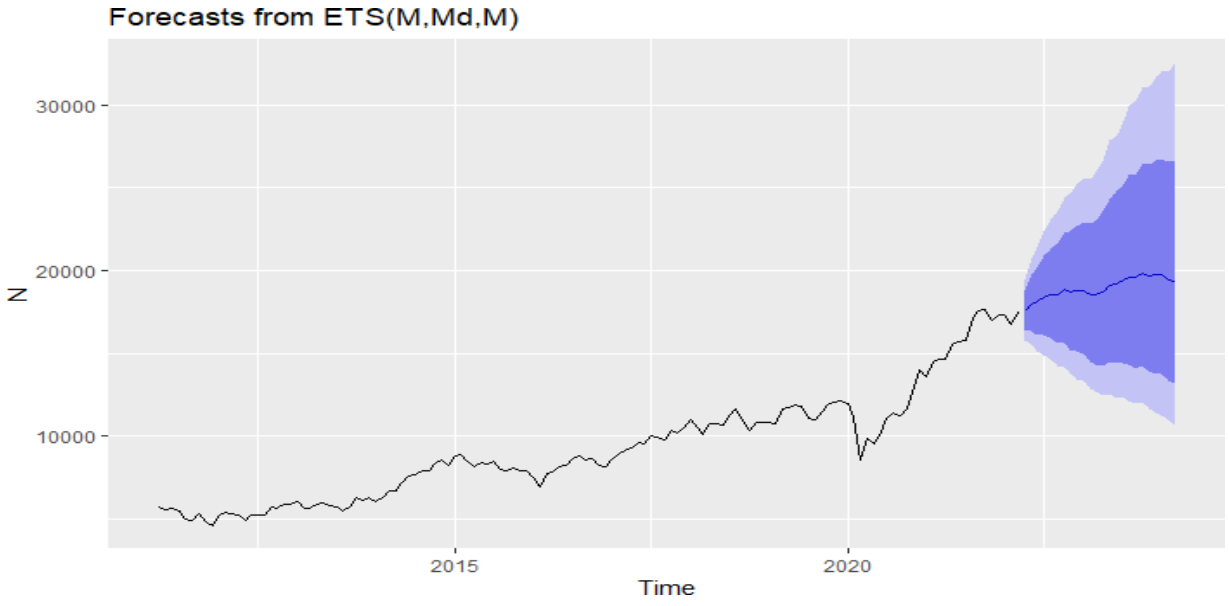
NIFTY 50

Code-

```
x<-ets(N,model = "MMM",damped = NULL,alpha = NULL,gamma = NULL,phi = NULL,additive.only = FALSE,lambda = NULL,biasadj = NULL,lower = c(rep(1e-04,3),0.8),upper = c(rep(0.9999,3),0.98),opt.crit = "mse",nmse=3,bounds = "admissible",ic=c("aicc","aic","bic"),restrict = TRUE, allow.multiplicative.trend = TRUE, use.initial.values = FALSE)

plot(forecast(x))
```

Graph



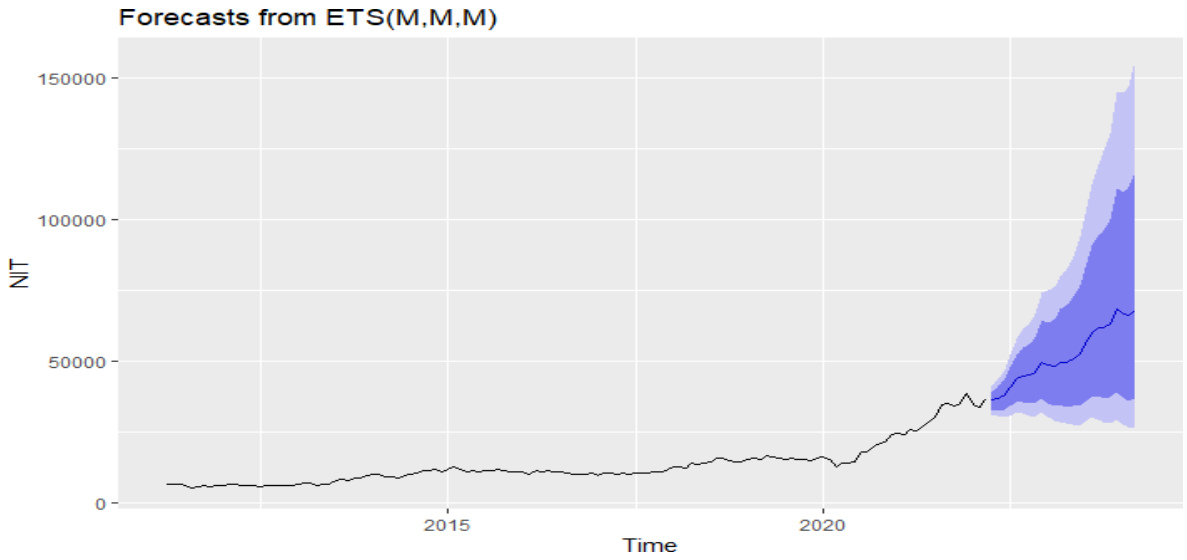
NIFTY –IT

Code

```
x<-ets(NIT,model = "MMM",damped = NULL,alpha = NULL,gamma = NULL,phi = NULL,additive.only = FALSE,lambda = NULL,biasadj = NULL,lower = c(rep(1e-04,3),0.8),upper = c(rep(0.9999,3),0.98),opt.crit = "mse",nmse=3,bounds = "admissible",ic=c("aicc","aic","bic"),restrict = TRUE,allow.multiplicative.trend = TRUE,use.initial.values = FALSE)

plot(forecast(x))
```

Graph



Result

Accuracy

1) Nifty 50

Code

```
accuracy(forecast(x))
```

Result

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	73.45459	468.9964	329.7859	0.6092453	3.690413	0.2288411	0.05626107

Conclusion

The accuracy of the data is **(100-3.690413)%= 96.309587%**. Comparatively, accuracy of exponential model was **93%**. So, this is a better a model than the exponential curve model. Hence, we use this to predict the price for next 10 years.

2) Nifty-IT

Code

accuracy(forecast(x))

Result

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	25.85718	825.5553	619.4471	-0.1748251	4.966736	0.1966669	0.01791305

Conclusion

The accuracy of the model is (100-4.966736)% = 95.033264%.Comparatively the accuracy of the exponential model was 86%. So , it is more accurate than the Growth curve model. Hence, it is better to use this to predict the price for next 10 years.

Forecasted Data

i. Nifty 50

	Point.For	Lo.80	Hi.80	Lo.95	Hi.95
Apr-22	17581.07	16371.37	18778.67	15747.65	19387.13
May-22	18000.21	16332.85	19671.85	15489.2	20535.89
Jun-22	18134.99	16131.17	20180.09	15117.93	21269.73
Jul-22	18401.98	16015.07	20802.29	14896.23	22142.62
Aug-22	18586.96	15893.34	21364.17	14653.85	22933.44
Sep-22	18580.33	15638.33	21614.17	14355.02	23550.43
Oct-22	18860.52	15632.52	22170.27	14236.36	24330.45
Nov-22	18730.7	15305.59	22285.2	13841.55	24451.42
Dec-22	18833.06	15239.81	22741.68	13587.23	24961.63
Jan-23	18834	14997.87	22969	13296.95	25410.35
Feb-23	18586	14629.5	22867.46	12973.57	25558.98
Mar-23	18520.15	14356.44	22962.08	12665.14	25884.74
Apr-23	18691.09	14366.62	23352.48	12516.52	26745.66
May-23	19102.28	14454.6	24165.8	12532.69	27336.94
Jun-23	19211.7	14369.62	24591.65	12399.1	28064.42
Jul-23	19461.5	14296.95	25105.79	12291.57	29033.82
Aug-23	19624.8	14335.74	25336.47	12074.24	29510.57
Sep-23	19586.48	14140.84	25640.1	12110.82	29699.91
Oct-23	19851.04	14173.14	26146.06	11917.01	30527.59
Nov-23	19684.74	13935.5	26354.52	11604.99	30643.94
Dec-23	19763.43	13826.52	26607.16	11401.99	31185.01
Jan-24	19736.4	13682.76	26725.35	11229.83	31653.1
Feb-24	19449.73	13370.05	26459.54	10911.27	31989.56
Mar-24	19354.95	13130.34	26632.61	10635.32	32083.63

ii. Nifty-IT

	Point.For	Lo.80	Hi.80	Lo.95	Hi.95
Apr-22	36221.48	32984.45	39487.77	31388.37	41146.76
May-22	37015.33	32739.15	41548.77	30696.73	44034.33
Jun-22	38402.46	33042.57	44108.06	30505.92	47546.29
Jul-22	40798.31	34208.97	48108.98	31436.92	52069.5
Aug-22	43814.63	36093.86	52499.15	32547.82	57909.32
Sep-22	44979.88	36098.12	54985.8	32353.17	60858.23
Oct-22	45090.33	35355.65	56025.49	31326.18	62960.49
Nov-22	46138.27	35469.16	58874.88	30932.75	66809.52
Dec-22	49752.53	37542.17	64626.17	32429.65	74238.91
Jan-23	48547.96	35808.66	64291.39	30776.79	74436.64
Feb-23	48124.02	34541.07	64963.8	29520.97	76227.31
Mar-23	49591.18	34960.18	68133.79	29288.63	81087.02
Apr-23	49744.35	34366.4	69661.87	28437.61	84192.63
May-23	50834.58	34365.21	72600.9	28374.68	88474.37
Jun-23	52739.58	34605.81	76579.99	28749.05	92265.16
Jul-23	56029.89	36135.13	82536.96	29209.2	102266.2
Aug-23	60172.32	37918.93	90411.39	30697.82	112824
Sep-23	61772.6	38127.44	94206.02	30300.54	118672.7
Oct-23	61924.29	37423.33	96425.31	29484.86	120720.7
Nov-23	63363.46	37505.75	100722.7	29023.78	127812
Dec-23	68327.06	39297.43	109929.6	30220.27	142270.7
Jan-24	66672.79	37948.57	109330.9	28614.62	142729
Feb-24	66090.57	36476.75	110971.8	27387.52	144291.3
Mar-24	68105.48	37025.79	116406.3	27188.96	156422.4

Conclusion of the project-

In the 1st section, that is, in the descriptive analysis of the data, we can see that the return from IT sector is most compared to any other sector. Although metal sector had an unexpected surge, it was one off event and it is better to invest on IT on long run. It also gave much higher return compared to the main sector, NIFTY-50.

Now, by Time Series analysis, we can see the price showed an upward trend in the long run with some sidewise trend.

By fitting different models and smoothing, it can be seen that the exponential smoothing model is the most accurate one for both NIFTY50 and NIFTY-IT. So this was used to forecast the price for the next 10 years.

This study and analysis represents the returns one can generally expect from the different sectors. It also gives an idea about the range in which the stock-price of the IT sector and NIFTY50 will vary in the near future.

Software Used

3) Microsoft Excel
4) Minitab
5) R
6) Microsoft Word

Reference

- The Analysis of Time Series An Introduction with R – Chris Chatfield & Haipeng Xing
- Fundamental of Statistics (Volume One & Two) – A.M. Goon, M.K. Gupta, B. Dasgupta
- Introduction to Statistics – Prasanta Kumar Giri, Jiban Banerjee

Appendix

- Nifty 50 calcutlations - https://drive.google.com/file/d/1alvGXjAApToWjaNlmZKkZo4J6KqwUsBK/view?usp=sharing
- Nifty IT calculations – https://drive.google.com/file/d/1lUgEkjp8iqXRkVbRYDUn1XMq-3SaDPa1/view?usp=sharing