# Stock Market Analysis using Data Visualization

**Bharat Gupta, Shefali Bhardwaj, K. Govinda, Rajkumar. R**

***Abstract*: Stock market is an important part of the economy of any country. People generally use stocks as an investment and try to get a profit out of them. Investing in the stock market is a huge risk-return. One bad decision can lead to huge losses while one good decision can result in good profit. To make sure that our investment decisions are not bad, one must choose the right stock. In this project we have tried to do the same. We have incorporated the functions to view data like closing prices, highs, lows etc. Using this data one can know the general trend of some stock but how to know whether the stock is risky or not. In this paper, we implemented two algorithms, namely Bootstrap method and Monte Carlo simulation for risk analysis.***

***Keywords: Correlation, Monte Carlo, Probability, GBM, Normal.***

## I. INTRODUCTION

Humans since a long time back have used pictures to represent information and today it has become a very big, hot topic in computer science in the form of data visualization. [1]. Visualization is the graphical representation of information while providing an analysis of the data in graphical form so that the user can draw a conclusion from it. Information visualization is something related to finding relationships between numbers, datasets and functions. Numerical data analysis needs specialized knowledge to draw out information out of them. On the other hand, it is easier to understand the visual representation of data. We can easily find out relations and trends in data in this manner as we use our visual senses which is more familiar to us. Pictures generally mean more to humans and is easier and faster for us to analyze and understand.

Stock analysis is the analysis of a particular stock, stock sector or the whole national market and one of its key parts is finding out risks associated with stocks and the future prices of stocks as well. If done and utilised in a proper manner, it can be used to gain profit and avoid loss. Monte Carlo simulations is a method used to find the probability (or set of probabilities) of different events which cannot be easily determined due to the interference of several random factors. It is used to find out risk and uncertainty of risk.

## II. LITERATURE REVIEW

**Free Cash Flow Analysis**: It is a way to look at the flow of cash to find which cash is available for use and the status of the other cash also. It can used to determine how much cash can be taken out from a company without disrupting its operations. It is a very useful method for risk analysis, however a major drawback of this is that it does not attempt to predict the future and so we do not consider possible situations and depend only on previous data. Value at Risk (VaR) [15] helps us find the risk in financial assets. It estimates how much a set of investments might lose, given normal market conditions, in a set time period such as a day. VaR is typically used by firms and regulators in the financial industry to gauge the amount of assets needed to cover possible losses. It is an improvement over FCFF as it is a quantitative measure compared to FCFF but it retains the same drawback i.e. it does not consider possible situations of the future.

Geometric Brownian motion (GBM) [2] is a process where the logarithm of a time changing value follows Brownian motion with drift. This is an innovation in stock market analysis as it is the first predictive model in stocks. However, it is not very reliable. Monte Carlo simulation [2] can be used to find probability and simulate uncertain situations. However, we can only analyze the maximum risk and confidence level from using plain Monte Carlo simulation.

Afterwards, an even better method was found which used Monte Carlo simulation as a basis called geometric Brownian motion. It beats all the previous methods in the aspect that it uses techniques of Monte Carlo simulation to predict risk, can be used to build a predict model, which though not always correct but is more reliable compared to the previously used method's.

## III. PROPOSED METHOD

In this paper, we predict the risk by using the Bootstrap method and Monte Carlo Simulation. Firstly we Store stock prices in database and Retrieve data from database. Then Plot the closing prices of a stock. Calculate the moving averages for given days and daily returns. We Correlation between different stocks and predict Risk analysis by Bootstrap method and Monte Carlo simulation.

We have used Python as a base for our project with MySQL as the database. We have connected these two using a Python library called SQLAlchemy. We are getting the requested stock data from Yahoo Finance through Pandas and storing this data into our database. Then we use the data in the database to illustrate this data in different ways and finally we show risk analysis using two different algorithms, Bootstrap and Monte Carlo simulation.

### A. Moving Average

A moving average (MA) is a measure of stock prices which tends to remove unnecessary information such as a sudden unreasonable spike in stock price for no reason. For example, a 10-day moving average would be average of first 10 stock prices.

Revised Manuscript Received on March 14, 2020.
**Bharat Gupta,** SCOPE, VIT, Vellore, India
**Shefali Bhardwaj,** SENSE, VIT, Vellore, India
**K. Govinda,** SCOPE, VIT, Vellore, India
**Rajkumar.R,** SCOPE, VIT, Vellore, India

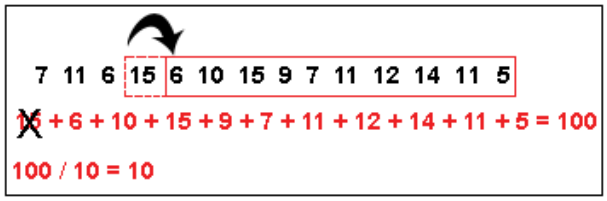


**Fig1. Calculating the Moving Averages**

*Return = ((Today's price – Yesterday's Price)/Yesterday's Price) *100.*
*#When return is a positive number, a profit is made.*
*#When return is a negative number, a loss is made.*

### B. Correlation of Stocks

Correlation [13] is the degree to which two variables move with respect to each other. It is used to compare relation between stocks. Correlation is calculated as correlation coefficient also known as Pearson's coefficient, which lies between -1 to 1.

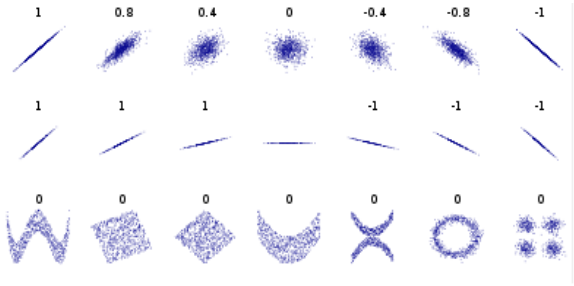


Fig2. Correlation of Stocks

### C. Monte Carlo Simulation

Monte Carlo simulation [2] can be used to find the maximum risk associated with a stock provided the conditions and its impact. It is used to find uncertainty in our prediction models. The main feature of Monte Carlo simulation is that given the range of variations of parameters that affect the quantity, we can determine the most likely outcomes.

In Monte Carlo simulation, a value/range of values is selected from a randomizing function which lies between the given ranges. Using the randomized values, we calculate our expected result. The result is recorded and we again start by randomizing the values and calculating the result. This is repeated a huge number of times, so that most cases can be taken into account for more accurate prediction of our required values. The result of the simulation is large number of plots based on several different input values and they all represent several scenarios.

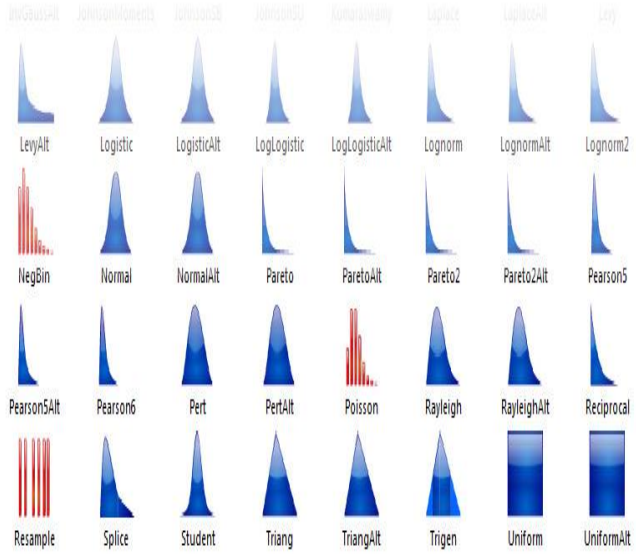


**Fig3.Probability distribution graphs using Monte Carlo Simulation**

### D. Bootstrap

Bootstrapping [6]is used to find standard error estimates and confidence levels for measures like mean. It may also be used to test hypotheses using sample tests. This method is mostly used as another method for parametric estimates when other methods' assumptions are unreliable.

In Bootstrap method, samples are generated by resampling the observed information many times. Whenever the distribution of the data is not known, Bootstrap method is considered as the best method. The main advantage of Bootstrap method is that it does not make any assumptions, instead it generated the data from the past stock data. The fact that sample distribution is poor for true distribution, is Bootstrap method's main disadvantage.

To find the risk percentage, we calculate the quantile of the Daily returns. Suppose we have to find the number from a array of data where 30% values are below and 70% above

6 9 11 12 12 13 19 21 22 31 32 35 38 45 44 52 55 68 79 80 81 88 90 91 92 100 112 113 114 120 121 132 145 146 149 150 155 174 184 192

Bootstraps follows the given procedure

- Order the given data
- Find number of observations n
- We make q=0.3 as we want a number with 30% values below it
- Our $i^{th}$ observation=q(n+1)=0.3*(40+1)=12.3

Answer: The ith observation is 12.3 which upon rounding off is 12. The $12^{th}$ number in set is35, i.e. 35 is the number with 30% values below it.

## IV. RESULTS

**For first option:** Store stock prices
Input: AAPL (stock quote)
Output: Stock stored in Database

**For second option:** Retrieve stock data
Input: AAPL.

**Table1. Stock data**

```
Enter Stock quote: aapl
         Date    High     Low    Open   Close      Volume Adj Close
0  2017-10-16     160  157.65   157.9  159.88  2.41215e+07   157.499
1  2017-10-17  160.87  159.23  159.78  160.47  1.89973e+07    158.08
2  2017-10-18  160.71   159.6  160.42  159.76  1.63742e+07    157.38
3  2017-10-19  157.08  155.02  156.75  155.98  4.25842e+07   153.657
4  2017-10-20  157.75  155.96  156.61  156.25  2.39741e+07   153.923
5  2017-10-23  157.69   155.5  156.89  156.17  2.19843e+07   153.844
6  2017-10-24  157.42   156.2  156.29   157.1  1.77572e+07    154.76
7  2017-10-25  157.55  155.27  156.91  156.41  2.12071e+07    154.08
8  2017-10-26  157.83  156.78  157.23  157.41  1.70005e+07   155.065
9  2017-10-27   163.6   158.7  159.29  163.05  4.44542e+07   160.621
10 2017-10-30  168.07  163.72  163.89  166.72  4.47008e+07   164.237
11 2017-10-31  169.65  166.94   167.9  169.04  3.60468e+07   166.522
12 2017-11-01  169.94  165.61  169.87  166.89  3.36378e+07   164.404
13 2017-11-02   168.5  165.28   166.6  168.11  4.13934e+07   165.606
14 2017-11-03  174.26  171.12     174   172.5  5.93986e+07   169.931
15 2017-11-06  174.99  171.72  172.37  174.25  3.50263e+07   171.655
16 2017-11-07  175.25   173.6  173.91  174.81  2.43615e+07   172.206
17 2017-11-08  176.24  174.33  174.66  176.24  2.44095e+07   173.615
18 2017-11-09   176.1  173.14  175.11  175.88  2.94826e+07    173.26
19 2017-11-10  175.38  174.27  175.11  174.67  2.51455e+07   172.687
20 2017-11-13   174.5   173.4   173.5  173.97  1.69821e+07   171.995
21 2017-11-14  173.48  171.18  173.04  171.34  2.47825e+07   169.395
22 2017-11-15  170.32  168.38  169.97  169.08  2.91581e+07    167.16
```

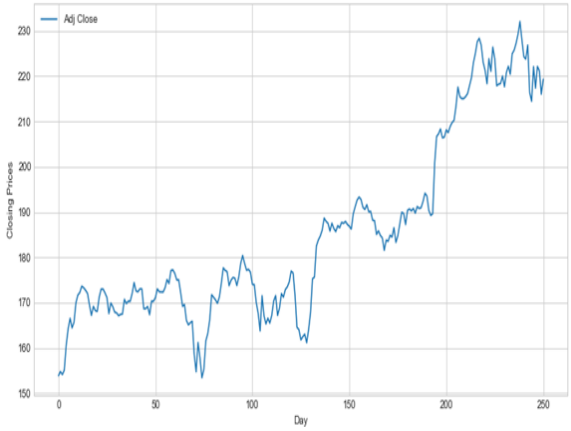


**Fig4. Closing Price of Stock**

The Figure4 shows the Adjusted Closing prices of the stock for last one year. An Adjusted Closing price [17] is a stock's end cost on some random day of exchange that has been corrected to incorporate any circulations and corporate activities that happened whenever before the following day's open. For instance, an organization choose to part the organization's stock two-for-one. Along these lines, the organization's shares outstanding increment by a various of two, while its offer cost is isolated by two. In the event that a stock shut at $200 the day preceding its stock split, the end cost is changed in accordance with $100, or $200 isolated by 2, per offer to demonstrate the impact of the corporate activity.

**For fourth option:** Find moving averages for given days
Input: Stock Quote and No. of days

```
Enter Stock quote: AAPL
Enter the number of days for which you want to calculate the moving average: 20
```


**Fig5. Moving averages for given days.**

A moving average (MA) is a broadly utilized marker in technical stock analysis that allows us to filter sudden random spikes in stock data which distort the trend in stocks. So, the Graph in blue is the graph for prices and the Graph in orange is the graph for moving average, which shows that the general trend is "Increase in prices".

**For fifth option:** Daily returns
Input: AAPL

```
Enter Stock quote: AAPL
```

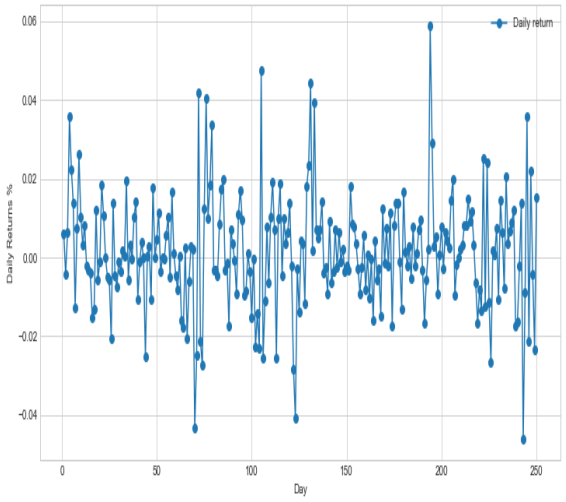


**Fig6. Daily returns**

A Return or a Financial Return is simply known as the money made or lost by investing in a stock over certain period of time. We can find return by simply finding the ratio of profit/loss and its last price. Y-axis here shows the return% of Apple stock for one year.

**For Sixth Option:** Correlation between different stocks
Input: AAPL FB TSLA

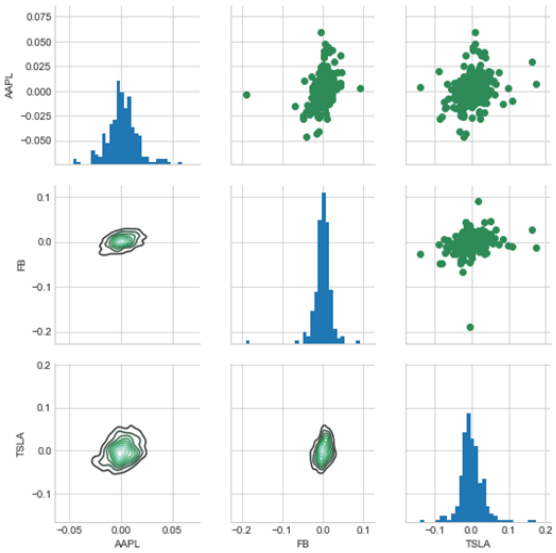


**Fig7. Correlation between different stocks**

As we know that denser and more linear the points are, more is the Pearson's correlation coefficient, we can tell that Apple and Facebook are most correlated. In contrast, Apple and Tesla are least correlated. Whereas, Tesla and Facebook are moderately correlated.

**For seventh option:** Risk analysis by Bootstrap method
Input: AAPL and Confidence percentage

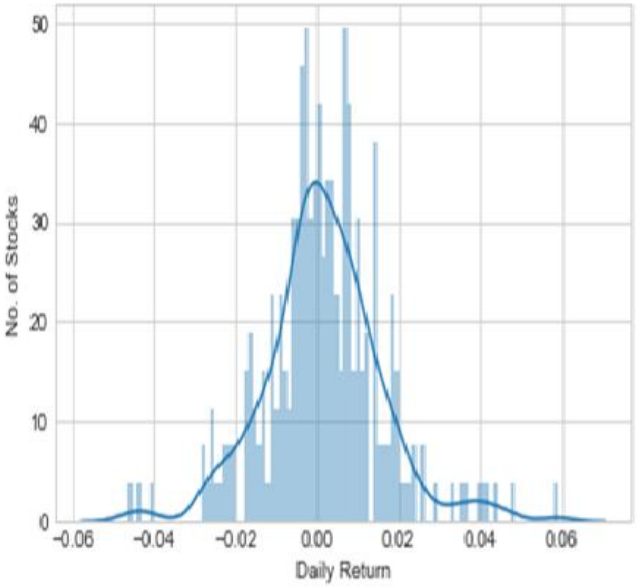


**Fig8. Risk analysis by Bootstrap method**

Or simply we can say that out of all the prices we have, what is the minimum return, above which 95%(say) of the prices reside.


**Fig9. Risk analysis by Monte Carlo simulation.**

This was just a demonstration of how the graphs will look like for 100 runs. But for better results we have to calculate the average for at least a thousand possible values.

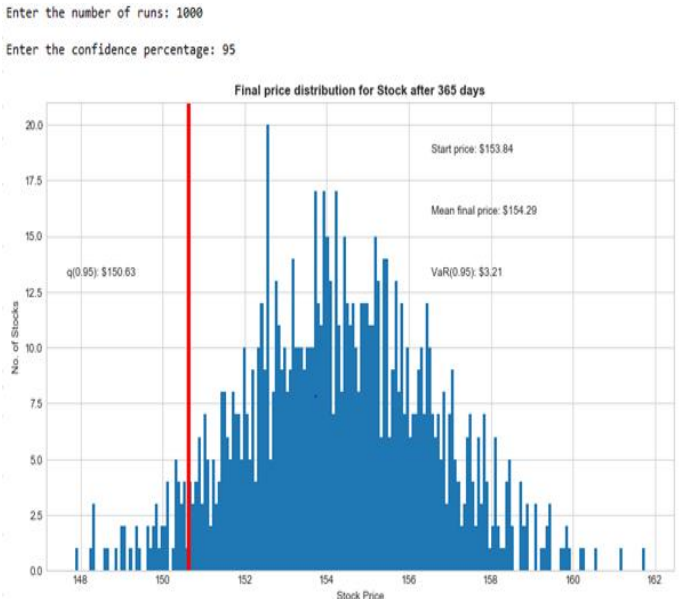


**Fig10. Price distribution of the stock**

From this graph we can say that for stock Apple, if the start price is $153.84 then after one year the price should be $154.29. And with 95% confidence we can say that the loss would not be more than $3.21 i.e. your final price should not go below $150.63.

If we compare Bootstrap algorithm and Monte Carlo Simulation, the Bootstrap algorithm is purely based on the past stock prices whereas Monte Carlo simulation works on both the average of the past prices and random value generation over huge number of runs and then

taking the mean of those numbers.

**Table2. Comparative Analysis**

| METHOD | | CONFIDENCE | VALUE AT RISK |
|---|---|---|---|
| **Stock Apple (AAPL)** | Bootstrap Method | 95% | $3.47 |
| | Monte Carlo Simulation | 95% | $3.36 |
| **Stock Microsoft (MSFT)** | Bootstrap Method | 95% | $1.78 |
| | Monte Carlo Simulation | 95% | $1.54 |
| **Stock Tesla (TSLA)** | Bootstrap Method | 95% | $16.27 |
| | Monte Carlo Simulation | 95% | $20.60 |

## V. CONCLUSION

Stock analysis is one of the most important things in the finance sector. Different graphical representations and mathematical calculation can avoid bad investments and forecast future prices. We saw graphs for Moving Average and Daily Return. We also saw correlation plots that show us how two different stocks are dependent on each other. For risk analysis we saw two different algorithms that are Bootstrap algorithm and Monte Carlo simulation. Bootstrap method predicts the risk using past data which is not a practical approach. Monte Carlo simulation creates its own possibilities for hundreds of times following a mathematical formula and then gives the final price after averaging the prices of each test case. Hence, we can conclude that Monte Carlo Simulation is better than Bootstrap method.

## REFERENCES


1. https://www.cs.uic.edu/~kzhao/Papers/00course_Data_visualization.pdf
2. https://stackoverflow.com/questions/16729574/how-to-get-a-value-from-a-cell-of-a-datafram.
3. Govinda.K, Sahaj Singh Maini, "Stock Market Prediction using Data Mining Techniques", *in the Proceedings of* International Conference on Intelligent Sustainable System, Dec'2017.
4. K.Govinda, Shruthi Hiremath, "Rainfall Prediction Using Artificial Neural Network", International Journal of Applied Engineering Research,Vol:9,No:23, pp: 21231-21241.
5. Krishna, A. V., Ramasubbareddy, S., & Govinda, K. (2020). A Unified Platform for Crisis Mapping Using Web Enabled Crowdsourcing Powered by Knowledge Management. In International Conference on Intelligent Computing and Smart Communication 2019 (pp. 195-205). Springer, Singapore..
6. Saraswathi, R. V., Nalluri, S., Ramasubbareddy, S., Govinda, K., & Swetha, E. (2020). Brilliant Corp Yield Prediction Utilizing Internet of Things. In Data Engineering and Communication Technology (pp. 893-902). Springer, Singapore.
7. Kumar, I. P., Gopal, V. H., Ramasubbareddy, S., Nalluri, S., & Govinda, K. (2020). Dominant Color Palette Extraction by KMeans Clustering Algorithm and Reconstruction of Image. In Data Engineering and Communication Technology (pp. 921- 929). Springer, Singapore.
8. Kumar, I. P., Sambangi, S., Somukoa, R., Nalluri, S., & Govinda, K. (2020). Server Security in Cloud Computing Using BlockChaining Technique. In Data Engineering and Communication Technology (pp. 913-920). Springer, Singapore.
9. Kalyani, D., Ramasubbareddy, S., Govinda, K., & Kumar, V. (2020). Location-Based Proactive Handoff Mechanism in Mobile Ad Hoc Network. In International Conference on Intelligent Computing and Smart Communication 2019 (pp. 85-94). Springer, Singapore.
10. Bhukya, K. A., Ramasubbareddy, S., Govinda, K., & Srinivas, T. A. S. (2020). Adaptive Mechanism for Smart Street Lighting System. In Smart Intelligent Computing and Applications (pp. 69-76). Springer, Singapore.
11. Srinivas, T. A. S., Somula, R., & Govinda, K. (2020). Privacy and Security in Aadhaar. In Smart Intelligent Computing and Applications (pp. 405-410). Springer, Singapore
12. Rajasekaran Rajkumar, Mridul Bhasin, K.Govinda, Jolly Masih, M.Sruthi: *IoT-Based Health Services Framework for Endless Ailment Administration at Remote Areas*. Incorporating the Internet of Things in Healthcare Applications and Wearable Devices, 10/2019: pages 184-198; IGI-global., ISBN: 2327-9354.
13. Rajkumar Rajasekaran, Govinda k, Ashrith Reddy, Uday Sai Reddy, Yashwanth Reddy: *Visual Analysis of Temperature Time Series and Rainfall Using Big Data*. DOI:10.36872/LEPI/V50I3/201023.
14. Rajasekaran Rajkumar, K.Govinda, Anushka Jindal, Rushil Mehtani, Jolly Masih, Maddineni Charan Sai: *Visualization Effect of Changing Climatic Trends on Natural Calamities:Hurricanes*. DOI:10.36872/LEPI/V50I3/201022.


## AUTHORS PROFILE

**Bharat Gupta,** I am 20 years old. I've been born and brought up in New Delhi, India. I live in a joint family which includes 5 members – My father, mother, sister, grandmother and me. Currently, I am pursuing Computer Science Engineering from VIT Vellore. My hobbies include skating, swimming, chess, lawn tennis, badminton and table tennis. I have been an international level chess player..

**Shefali,** is a student of 3rd year BTech Biomedical at VIT Vellore. She is an 19 year old from Delhi. She is the recipient of Scholar's Gold Certificate for seven consecutive years during her school days and also won a Silver Medal in French Olympiad. Shefali has completed her schooling from DPS RK Puram and cleared her CBSE with a percentile of 93.2%. As a person she is always willing to learn new things in life.

**Dr. K.Govinda.** Associate Professor in VIT from 2005 to till date, received the degree in computer science and engineering from Nagarjuna University in 1998. M.Tech from JNTU, Hyderabad. Life member in CSI, IEEE and ACM. Published 100 + papers and participated in 50+ conferences. Interest area include Database, Data Mining, Cloud Computing, IOT and Big Data.

**Dr. Rajkumar Rajasekaran,** graduated in Electrical and Electronics Engineering from Madras University in 2000 and received Master's degree in Computer Science & Engineering from VIT University, Vellore. He has completed his Ph.D. from Vellore Institute of Technology University, Vellore in Computer Science and Engineering. Her research area includes Data Analytics, Data Mining, Big Data, Machine learning and Data Visualization. He has been teaching from last 17 years with emphasis on Data Science.