International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-6, March 2020

Prediction of Bitcoin using Recurrent Neural Network

Pratik Mehta, E. Sasikala

Abstract*: For the past couple of years, Machine learning and trading helped by artificial intelligence has drawn growing interest. Here, the approach is used to test the hypothesis that the inefficiency of cryptocurrency industry can be exploited in order to produce anomalous revenue. For the duration between Nov. 2015 and Apr. 2018, daily data for 1, 681 crypto currencies were analyzed. Simple trade techniques supported by state-of -the-art machine learning algorithms are seen to outperform the traditional benchmarks. The results obtained imply that non-trivial, but fundamentally simple, algorithmic processes will help to predict the short-term future of the cryptocurrency market. The popularity of cryptocurrencies had skyrocketed in 2017 due to several consecutive months of super-exponential growth of market capitalization. There are over 1,500 currently recorded cryptocurrencies actively trading today with the cryptocurrencies sitting on more than $300 billion [2], and a total market capitalization of over $800 billion in January 2018. According to a recent survey, between 2.9 and 5.8 million privates as well as institutional investors are in the numerous investment networks and access to markets has become easier over time. In a number of online markets, major crypto currencies can be purchased using fiat currency, and then used in order to purchase less known crypto currencies. The average trading amount is globally exceeding $15bn. About 170 money market funds had been invested in cryptocurrencies since 2017, and Bitcoin futures are launched in order to satisfy the Bitcoin trading and hedging demand for the market. The main objective of the work is to predict the Bitcoin prices, one of the most popular and widely used cryptocurrency which is a source of attraction for many investors as a source of profit or investment. But the market for the cryptocurrencies been volatile since the day it was first introduced. So, the approach towards the survey is to use LSTM RNN and use the available dataset and train the model to give the highest possible accuracy and to provide a real-time price of the Bitcoin for the following days.

Keywords: Bitcoin, Cryptocurrency, Context layer, Time stamp, Blockchain, Recurrent Neural Networks

I. INTRODUCTION

The Bitcoin (BTC) [1] is a decentralized, digital currency system that has no central governing body at all till the present date. Then, it passes and re performed through a peer-to-peer Internet-connected device network.

Manuscript received on February 10, 2020.
Revised Manuscript received on February 20, 2020.
Manuscript published on March 30, 2020.
* Correspondence Author
Pratik Mehta*, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur
Dr. E. Sasikala, Associate Professor in Computer Science and Engineering at SRM Institute of Science and Technology, Kattankulathur.

Time series estimation isn't a new concept in today's date. Extensive research had been carried out on evaluating existing financial markets, such as the stock market. Bitcoin is an interesting contrast to this, as it is still in its early transitory period. A time series estimation problem in such a sector. As a consequence, the sector is seeing very high volatility and this provides an opportunity in terms of analysis.

It is very easy to obtain Bitcoins by mining or in return for goods, services or other currencies.

Furthermore, Bitcoin is the world's largest, and mostly used cryptocurrency with the acceptance steadily growing over time. Because of its open nature, Bitcoin also provides another concept in comparison to traditional financial markets, as it is in its transitional stage. No other financial markets have this degree of revelation.

The research by Detzel et al. [11] that is the first review of technical analysis applied to Bitcoin, according to the authors assertion. In this work the analysts examine trading strategies to forecast Bitcoin returns based on moving averages of Bitcoin prices across various horizons. The first studies on bitcoin price projection begin to appear in the last years.

Prediction: The valuation of the Bitcoin varies just like any other stock. There are many algorithms that are used for price forecasts on stock market data. The parameters which affect Bitcoin, however, are different. Therefore, predicting the value of Bitcoin is important so that proper investment decisions can be made. Unlike the stock market, Bitcoin's price is not based on economic activities or government authorities interfering. Therefore, to forecast the value that, it is important to use machine learning [9][10] technologies to predict Bitcoin's price.

II. RELATED WORKS

A. Financial Data Analysis

Several strategies are listed in the literature, including one that is also known as a "map" that predicts future prices [3]. According to it, stock market markets do not obey random runs, that is–a series of trends accompany the price moves. These changes in demand can be used to predict the future price [4]. There are some other empirically constructed patterns that can be used to predict future values, such as heads-and-shoulders, double-top-and bottoms.

B. Time Series Data Analysis

Classical approaches are fairly popular in the form of future price forecasts. Autoregressive integrated moving average models (ARIMA) are a popular choice for short term forecasting. This functions very well when over time the data shows regular or constant trend with the least outliers possible.

The ARIMA approach only functions well when the data shows "stationarity," meaning the sequence stays nearly constant. But in the real time case, where the data fluctuates significantly, this is not always realistic and it is highly volatile. Ediger and Akar used the ARIMA seasonal model to predict the future demand for fuel oil in Turkey over a span of some years [5]. This scenario, however, is not guaranteed to work on unseasonal or nonlinear results. Random forest model is quite useful for increasing the speed of computing due to its ability to tackle nonlinearities in the data to solve the real time prediction problems on big data.

Three models based on a Bayesian Neural Network [15], linear regression, and support vector regression [16] were properly investigated and showed that the Bayesian Neural Network model outperformed the other two predictive models.

C. Bitcoin Price Prediction Analysis

Using Neural Networking Systems [8]. The connection between the performance of Bitcoin and the next day's price change of Bitcoin using an Artificial Neural Network Ensemble solution called the Selective Neural Network Ensemble Dependent Genetic Algorithm was discussed, then the neural network using Multi-Layered Perceptron was constructed. The organization was used to predict the price of Bitcoin's next day course through a sequence of nearly 200 blockchain apps over a 2-year cycle to better understand Bitcoin's practicality and utility of real-world applications. With a series of almost 200 blockchain apps over a period of 2 years, the firm was used to predict the next day path of Bitcoin's price to better understand the practicality and utility of it in real world applications. The program was used to predict the next day's market path for Bitcoin with a selection of about 200 blockchain apps over a 2-year span to better understand the practicality and efficacy of real-world applications. An ensemble-based trading methodology was compared over a span of 50 days with a previous day trend following a back-test trading strategy. The former trading strategy produced about 85 percent returns, outperforming the previous day trend following a special trading strategy that yielded about 38 percent returns and a trading strategy that follows the one-time, best MLP (Multilayer Perceptron) model in the ensemble that yielded around about 53 percent. [9] Provides multi-layered perceptron for estimating bitcoin level. Jang and Lee's work [12] which predicts the bitcoin price using Bayesian Neural Network and blockchain information. There are multiple online platforms nowadays, such as CoinTracking, BitcoinCharts, Bitcoinity, and BitcoinWisdom, that enable traders to use many technological analytics tools to identify trends and market feelings that are useful for entering a trade. Several Regression models which was based on Linear Regressions, Random Forests [13], Gradient Boosting Technology [14], and Basic Neural Networks. Kim et al. [17] and Li et al [18] found the Bitcoin price volatility to be expected by social evidence.

D. Inference from the survey

An implementation of random forest model to forecast the closing price of cryptocurrencies Bitcoin one day ahead is found not to be stable on the Bitcoin market. Because the Bitcoin market is open to exploitation, bitcoin price may be a severe drop or a severe spike. This is basically a very large problem for predicting the time series. Other attempts to use machine learning to predict other than Bitcoin blockchain values come from channels other than academic ones. None of the above analysis however addressed CNN-based prediction models and different combinations of deep learning models. Most of these analyzes focused on a limited number of currencies, which did not provide direct comparisons to their results. To grasp these severe ups and severe downs accurately, the model needs to provide knowledge that can predict manipulations.

III. LONG SHORT-TERM MEMORY(LSTM) RNN

Bitcoin Prediction Model: In forecasting the average cryptocurrency price for 1,681 currencies present to date and the efficiency of three versions is checked thoroughly. Three models are shown, two of them were based on gradient-boosting decision trees, while the other is based on recurrent neural networks with Long Short-Term Memory (LSTM) Fig. 1.

In all cases, we construct investment portfolios based on the forecasts and we compare their return on investment results. From the outcome of the return it is very well observed that all of the three models will perform better than an 'easy moving average' model of baseline where the price of a currency is calculated as the average price over the previous days. The approach based on alternating neural networks with long short-term memory consistently provides the best return on investment.

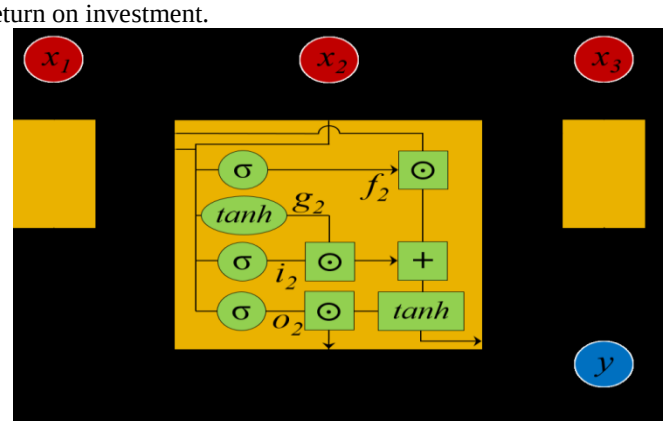

Fig. 1. LSTM model architecture

The Bitcoin Prediction Model consists of various steps:

A. Dataset Collections

Initially started with knowledge of the Bitcoin market which was obtainable publicly on Kaggle two. The Dataset contains historical knowledge of Bitcoin from Dec 1st, 2014 to Jan 8th, 2018 separated into intervals of one minute. This set of points consists of 1,574,274 minutes. The information includes the details of the difference value, the closing value, the best value, the low always value, and the degree and weighted value of each time stamp. The Dataset was labeled as "True" if the value went up at the top of the one-minute timestamp and was labeled as "False" if it remained or attenuated at a constant position.

B. Data Processing

Scraping the data gives m samples a 2D tensor by n parameters. Transform time series is used to transform this into a collection of window data with window size w=50, yielding a 3D shape tensor (m-w) observations by n features per w day window size.

For e.g., for each of 0-49 days the first data point m=0 had a 2D tensor of m-features. The data is then normalized. Finally, this was split into the input and output data by extracting the last day and rendering the output data there.

C. Predicting Polarity using RNN (Recurrent Neural Network)

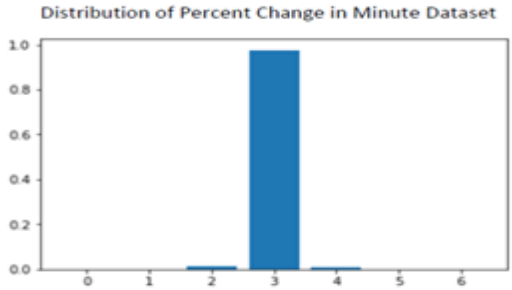

Fig. 2. Distribution of percent change in minute dataset

The Fig. 2. graph above shows that nearly all the 1.5 million minutes fall within the "third bin" which represented price. The obvious end goal of creating a neural network focused on crypto-currency is to forecast real-time price volatility. A start with an extremely temporally resolved dataset was made with this goal in mind. When information was provided on a minute by minute or second by second time scale, an even better job could be achieved in forecasting inflation and also keeping ahead in terms of the market. There are also issues with easily resolved results, however. Looking at the minute chart, there is an expectation that on a minute time scale there would be no improvement, or that it was very low and quiet, if there was change.

D. Deep Learning

A Bidirectional 3-layer RNN is used to forecast Bitcoin's closing price, provided a variety of previous-day results. The Code for scraping crypto currency data is included, as well as the LSTM model

E. Algorithm

Machine Learning

Machine Learning is a system which can learn from an example or a model by self-improvement, without any direct coding by the developer. The breakthrough came with the idea that a machine would singularly benefit from the data provided in order to achieve highest possible accurate results. Machine learning combines data with statistical tools in order to predict an output.

Random Forest

A Random forest model is equipped in this project to forecast the closing price on next day. Random forests at a higher level are arrays of decision trees that are used for classification and regression activities. When developed, each decision tree classifies an unknown point by casting a vote, and with the most votes, the random forest declares the marking or interest.[6]

Random Forest model is very useful for increasing the speed of computing due to its ability to tackle the nonlinearities in the data to solve the real time prediction problems on big data.

It is a type of ensemble model that combines the predictions of many different "reasonably good" models to generate a overall forecast that estimates the true hypothesis better than that. Ensemble models are highly successful as machine learning devices, because many single models escape the chance-dependent pitfalls.

Gradient descent techniques, for instance, can get trapped in local minima, but the combination of many models increases the chance of reaching the global minimum. An average of each forecast will lead to a prediction that fits the underlying truth more closely, even if none of the models in the ensemble yield the correct hypothesis. In spark Random Forest use a separate subsample of data to train each tree. Instead of directly replicating data, memory is saved using a Tree Point structure that stores in each sub-sample the number of replicas of each instance.[7]

In Map-Reduce's view represented in Fig. 3., random forests can also be outlined step by step like this:

1. From a CSV file the data is properly read and trained.
2. For n number of Trees, n number of mappers are called
3. A 90% subset is created of the training data with the replacements
4. The data is taken when the mapper is running
5. Each mapper begins constructing tree after receiving data and generates prediction for test dataset.
6. The basic test data and the label are passed as key and the value of it is passed to Reducer
7. The Reducer counts the majority of label in accordance with the key.
8. Results are written to the output file.

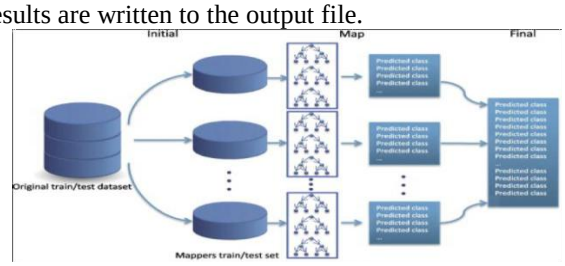

Fig. 3. Map-Reduce's view of random forests

TensorFlow Platform

The TensorFlow architecture basically works in three parts: First step is the Preprocessing of the data; Second comes the building of the model or architecture and third step is to train and estimate the accuracy of the model. TensorFlow is named because it takes data as a series of multidimensions, known as tensors. Develops a kind of flowchart operations (called a graph) that can be performed at that order. The data goes in at one end, and then it passes through this multiple-operation system and then comes out to the other end as the feedback. That's why it is known as TensorFlow, because a sequence of operations passes through the tensor that goes into it, and then the other side comes out.

F. Architecture Diagram

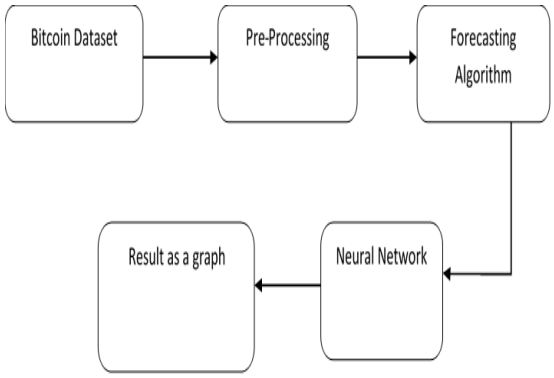

Fig. 4. Basic Diagram of the Bitcoin prediction process

Basic structural diagram for the prediction of Bitcoin process as shown in Fig. 4. The process it follows are:

i. Bitcoin Dataset is taken
ii. Preprocessing of the data present inside the dataset
iii. The forecasting algorithm is used and then apply our Neural Network in order to train the dataset
iv. Next the trained dataset and tested dataset is compared with the actual dataset to get the accuracy of the prediction.

IV. RESULTS AND DISCUSSION

A. Dataset and Programming Outcome

First a sample input dataset Fig. 5. (here as basic input we take daily dates) is taken and loaded for the Bitcoin prediction process. A small amount of data is taken from a very large dataset and is processed.

Timestamp	Open	High	Low	Close	Volume_(BTC)	Volume_(Currency)	Weighted_Price
2018-11-09 00:00:00+00:00	6419.05	6419.06	6419.05	6419.06	0.939389	6029.991228	6419.054827
2018-11-10 00:00:00+00:00	6342.06	6342.07	6342.06	6342.07	2.075609	13163.654823	6342.067698
2018-11-11 00:00:00+00:00	6365.00	6365.00	6364.99	6364.99	1.044918	6650.897185	6364.992297

Fig. 5. Sample data from the dataset

The result of random forest model Fig. 6. trained on data set from 2012 to 2018 (starting from Jan 2012 to June 2018) and tested on second quarter of 2018(Sep 2018 – Nov 2018). Model used the window size of 7 and 100 epochs done during training and testing process.

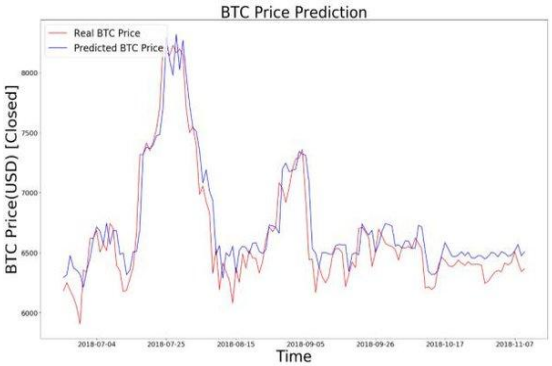

Fig. 6. Bitcoin prediction using Random Forest

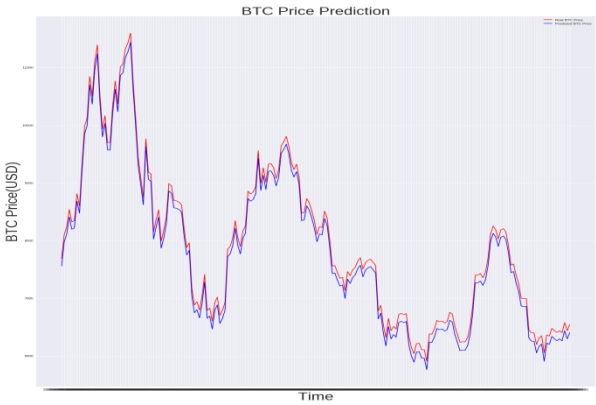

Fig. 7. Bitcoin prediction using Long Short-Term Memory RNN

Table 1: Price Prediction for Bitcoin using Random Forest and LSTM RNN

Method	Accuracy
Random Forest	84%
LSTM RNN	92%

B. Inference from the output

From the output and our testing model Random Forest (Table 1.) gave the accuracy of 84% and the Long Short-Term Memory (LSTM) RNN gave the accuracy of an outstanding 92% thus making it the best choice algorithm used for Bitcoin price prediction. Cryptocurrency exchanges are attracting attention due to recent developments in their underlying technology in today's world and buyers see this as part of the new space for various investments. With this rapid growing interest from the investment community, the cryptocurrency markets are an important asset class for both the analysts and traders alike.

V. CONCLUSION

Overall, the comparison is done for the prediction of bitcoin using LSTM RNN Fig. 7. and Random Forest algorithm using different amount and size of datasets. Taking the time and the speed also in consideration and the complexity of the database.

Depending on the statistical accuracy of the Bitcoin prices where the array of datasets is taken first. Then preprocessing is completed, and the function is extracted from it. With the neural network a clear idea of the result of the graph is obtained. It is seen that the prediction of bitcoin prices is very highly predicted by the LSTM Recurrent Neural Network (RNN) and that Random Forest algorithm predicts Bitcoin prices quite closely but less than LSTM RNN, but LSTM RNN is very effective if the dataset is really high and with a large number of volumes, days and hour. If some more information can be added making the database more complex LSTM RNN can function properly giving the best prediction over time.

This work is the first to study the BTC / USD market by means of an agent-based architecture model including more realistic prediction strategies. In future research models investigation based on application of more advanced and complex model and datasets will be done paving the way to predict the Bitcoin and other cryptocurrencies prices very easily so that cryptocurrencies can become a investment region for many people around the globe.

Moreover, approach will be made towards the security region for the cryptocurrency in order to detect the fraud and make the environment for Bitcoin and cryptocurrencies safe.

REFERENCES

1. Nakamoto, S. Bitcoin: A Peer-to-Peer Electronic Cash System. Technical Report. Available online: https://bitcoin.org/bitcoin.pdf (accessed on 12 July 2019). W.-K. Chen, *Linear Networks and Systems* (Book style). Belmont, CA: Wadsworth, 1993, pp. 123–135.
2. Amjad, M. J. & Shah, D. (2017). "Trading Bitcoin and Online Time Series Prediction". Proceedings of the Time Series Workshop at NIPS 2016, (pp. PMLR 55:1-15). B. Smith, "An approach to graphs of linear forms (Unpublished work style)," unpublished.
3. Areekul, P., Senjyu, T., Urasaki, N., & Yona, A. (2010). Next day price forecasting in deregulated market by combination of Artificial Neural Network and ARIMA time series models. 2010 5th IEEE Conference on Industrial Electronics and Applications. doi:10.1109/iciea.2010.5514828
4. Lo, A. W. & MacKinlay, A. C. (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. The Review of Financial Studies, Volume 1, 41–66.
5. Areekul, P., Senjyu, T., Urasaki, N., & Yona, A. (2010). Next day price forecasting in deregulated market by combination of Artificial Neural Network and ARIMA time series models. 2010 5th IEEE Conference on Industrial Electronics and Applications. doi:10.1109/iciea.2010.5514828
6. Leo Breiman Statistics and Leo Breiman. Random forests. pages 5–32, 2001.
7. Thomas G. Dietterich. Ensemble methods in machine learning. In Proceedings of the First International Workshop on Multiple Classifier Systems, MCS '00, pages 1–15, London, UK, UK, 2000. Springer-Verlag.
8. Lipo Wang and Edwin Sin School of Physical and Mathematical Sciences Nanyang Technological University Singapore presented ―Bitcoin Price Prediction Using Ensembles of Neural Networks‖.
9. Neha Mangla,‖Unstructured Data Analysis And Processing Using Big Data Tool-Hive And Machine Learning Algorithm-Linear Regression, International Journal of Computer Engineering & Technology (IJCET) Volume 9, Issue 2, March-April 2018, pp. 61–73, Article ID: IJCET_09_02_006 .
10. Neha Mangla,‖Machine Learning Approach for Unstructured Data Using Hive,‖ International Journal of Engineering Research‖, Volume 5,Issue 4, Pages 790-991, 2016.
11. Detzel, Andrew L. and Liu, Hong and Strauss, Jack and Zhou, Guofu and Zhu, Yingzi, Bitcoin: Learning and Predictability via Technical Analysis. Paris December 2018 Finance Meeting EUROFIDAI-AFFI. Available at SSRN: https://ssrn.com/abstract=3115846 or http://dx.doi.org/10.2139/ssrn.3115846 (2018).
12. Jang H., Lee J., An Empirical Study on Modeling and Prediction of Bitcoin Prices with Bayesian Neural Networks Based on Blockchain Information. IEEE Access, 6, 5427-5437, (2018).
13. Ho, T.K. Random decision forests. In Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, Canada, –16 August 1995; Volume 1, pp. 278–282.
14. Freund, Y.; Schapire, R.E. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. J. Comput. Syst. Sci. 1997,55,119–139.
15. Bishop, C.M. Pattern Recognition and Machine Learning; Springer: Berlin/Heidelberg, Germany, 2006.
16. Cortes, C.; Vapnik, V. Support-vector networks. Mach. Learn. 1995, 20, 273–297.
17. Kim, Y.B.; Kim, J.G.; Kim, W.; Im, J.H.; Kim, T.H.; Kang, S.J.; Kim, C.H. Predicting Fluctuations in Cryptocurrency Transactions Based on User Comments and Replies. PLoS ONE 2016, 11. [CrossRef].
18. Li, T.R.; Chamrajnagar, A.S.; Fong, X.R.; Rizik, N.R.; Fu, F. Sentiment-Based Prediction of Alternative Cryptocurrency Price Fluctuations Using Gradient Boosting Tree Model. arXiv 2018, arXiv:1805.00558.

AUTHORS PROFILE

Pratik Mehta did his primary and secondary education from Delhi Public School, Ruby Park, Kolkata. Currently doing final year of undergraduate study at SRM Institute of Science and Technology, Kattankulathur in B.Tech Computer Science and Engineering (2016-2020). Area of interest always been around Cryptocurrency and Machine Learning processes with various Neural Network Techniques.

Dr. E.Sasikala currently an Associate Professor in Computer Science and Engineering at SRM Institute of Science and Technology, Kattankulathur. She completed her B.Tech in Information Technology at K.S.R. College of Engineering, Anna University, 2006 . Did her Masters of Engineering in Computer Science and Engineering at Anna University of Technology, 2010. She further completed her Doctorate degree in Network Security at Anna University, 2017. She is Life member of the ISTE (Membership Number LM 74419), Annual Member of IET and Annual Member of Indian Science Congres