International Journal of Management and Humanities (IJMH)
ISSN: 2394-0913 (Online), Volume-8 Issue-10, June 2022

Comparative Analysis of Prediction Models for Cryptocurrency Price Prediction (May 2022)

Harshvardhan Sinha, Sangle Gaurav Keshavrao, Pulkit Verma

Abstract: ***The decentralization of cryptocurrencies has greatly reduced the level of central control over them, impacting international relations and trade. Further, wide fluctuations in cryptocurrency price indicate an urgent need for an accurate way to forecast this price. This project proposes a method to predict fluctuations in the prices of cryptocurrencies, which are increasingly used for online transactions worldwide. This project proposes a novel method to compare the efficiency of five different models predicting the cryptocurrency price by considering various factors such as market cap, volume, circulating supply, and maximum supply based on deep learning techniques such as the Linear Regression, Support vector regression (SVR), Auto Regressive Integrated Moving Average (ARIMA), Long Short-Term Memory (LSTM) and a hybrid model of ARIMA and LSTM which are effective learning models for training data. The hybrid model outperforms the LSTM and ARIMA model after comparing RMSE values. The proposed approach is implemented in Python and validated for benchmark datasets.***

Index Terms**: *Bitcoin, cryptocurrencies, predictive model, regression, ARIMA, LSTM, RMSE.***

I. INTRODUCTION

Cryptocurrency become a recent trend of mode of transaction. The recent research are going on to develop new cryptocurrencies and to make them available for users. Also, it's important to predict the price of cryptocurrency as many investors are trying to invest more in cryptocurrencies. Researchers have tried several approaches to design a model for prediction. Machine learning algorithms are playing very much effective role in this kind of problems. We are trying to compare some of the models that researcher used and using the comparative analysis on these models we are trying to modify or design a new model to get the more accuracy and effective regarding the time and space complexity.

II. MOTIVATION

Bitcoin is a successful cipher currency introduced into the financial market based on its unique protocol and Nakamoto's systematic structural specification. It was first launched in 2008 to serve as a transaction medium between participants without the need for any intermediary. Unlike existing fiat currencies with central banks, Bitcoin aims to achieve complete decentralization. Since 2017, cryptocurrencies have been gaining immense popularity. A recent survey revealed that more than 1,500 cryptocurrencies are actively traded by individual and institutional investors worldwide across different exchanges. Cryptocurrencies are new economic and financial tools with special and innovative features. However, cryptocurrencies are becoming more and more popular, and many merchants accept payments, especially online payments. Cryptocurrencies have faced periodic rise and sudden dips in specific time periods, and therefore the cryptocurrency trading community has a need for a standardized method to accurately predict the fluctuating price trends. There are salient differences between cryptocurrencies and traditional currency in terms of price fluctuation. Research on the attributes of cryptocurrencies has made steady progress but has a long way to go. This project is a comparative study of four different models predicting the fluctuations in the price of cryptocurrencies. These models are namely, Linear regression, Auto Regressive Integrated Moving Average (ARIMA), Support Vector Regression (SVR), Long Short-Term Memory (LSTM) model. Therefore, the main objective is to establish a more accurate cryptocurrency price prediction model.

III. BACKGROUND

There are some research works done on this topic. Yecheng Yao et al. [1] proposed a deep learning way to predict the price of cryptocurrencies. The paper emphasized on using Long Short-Term Memory (LSTM) and Recurrent Neural Network (RNN) for predictive analysis of cryptocurrency price. The factors that are considered are market cap, volume, circulating supply, and maximum supply. The authors also gave idea about some factors like political environment and human regulations. Guus van Heijningen [2] in his thesis worked to predict the cryptocurrency using web scraping. The data sources are scraped from web and the features are determined. Author also used some sentimental analysis on textual data. The accuracy found is 63% when Random Forest algorithm is used. Also, this model worked well on popular cryptocurrencies. Muhammad Ali Nasir et al. [3] proposed a model where a cryptocurrency returns, and volumes are forecasted using search engines. A weekly dataset from 2013 to 2017 is used and captured a dependence structure by using empirical approaches like VAR framework, a copulas approach etc. Jethin Abraham et al. [3] discussed about how tweet volumes and sentiment analysis can be used to predict the cryptocurrency price.

Manuscript received on 26 May 2022.
Revised Manuscript received on 18 June 2022.
Manuscript published on 30 June 2022.

* Correspondence Author
Harshvardhan Sinha*, Department of Computer Science, VIT University, Vellore (Tamil Nadu), India. Email: harshvardhansinha4@gmail.com
Sangle Gaurav Keshavrao, Department of Computer Science, VIT University, Vellore (Tamil Nadu), India. Email: gauravksangle@gmail.com
Pulkit Verma, Department of Computer Science, VIT University, Vellore (Tamil Nadu), India. Email: pulkitverma12@gmail.com

The main focus is given on prediction of values of bitcoins and Ethereum. The model contains a linear model to take input tweets and Google trends data. This model measures the overall interest regarding the cryptocurrency price rise or fall in terms of volume. The authors also support the idea of using linear regression model along with lagged variables to get the better results. N. I. Indera et al. [5] proposed a bitcoin prediction model using PSO-optimized parameters and moving average technical indicator. The model uses a Multilayer perceptron based Non-Linear Autoregressive with Exogeneous Inputs (NARX). Particle Swarm Optimization is used to optimize the model by optimizing hidden layers and input lags n NARX model. Moving Average is helpful to improve the prediction of the model. Aggarwal A. et al. [6] discussed about various parameters affecting bitcoin price prediction based on Root Mean Square Error (RMSE) using various deep learning models like Convolutional Neural Network (CNN), Long ShortTerm Memory (LSTM) and Gated Recurrent Unit (GRU). Phaladisailoed T. et al. [7] his thesis worked to discover the most efficient and highest accuracy model to predict Bitcoin prices from various machine learning algorithms. Using 1-minute interval trading data on the Bitcoin exchange website named bitstamp, some different regression models with scikitlearn and Keras libraries had experimented. Wimalagunaratne M. et al. [8] in his thesis construct a more holistic approach to a predictive price model for the cryptocurrency market. Key factors used are public perception, trading data, historic price data, and the interdependencies between Bitcoin and Altcoins. Machine learning techniques used are Artificial Neural Network (ANN), Support Vector Machine (SVM), Random Forest, Naïve Bayes. Ross C. Phillips and Denise Gorse [9] discussed about the financial price bubbles for a number of cryptocurrencies using a hidden Markov model. Using Social media data mining and Bubbles and epidemic detection, valuable predictive information pertaining to cryptocurrency price movements is achieved. Chongyang Bai et al. [10] in his thesis used C2P2 algorithm considering the fact that the price of a cryptocurrency c might depend not only on historical prices, sentiments, global stock indices, but also on the prices and predicted prices of other cryptocurrencies. it uses similarity metrics in conjunction with collective classification to compare multiple cryptocurrency features to jointly predict the cryptocurrency prices for all the coins considered. Brauneis A. and Mestel R. [11] stated in his paper that bitcoin is not weakly efficient but becomes more efficient in the second half of the sub sample. Th paper also states that Bitcoin is the least predictable and a heterogeneous pattern of efficiency, probably related to liquidity and size is found. It is found that the turnover ratio is a measure of liquidity to positively affect efficiency. Alessandretti, L., ElBahrawy, A., Aiello, L.M. and Baronchelli, A. [12] have tested the performance of three models in predicting daily cryptocurrency price for 1681 cryptocurrencies. Two of the models are based on the gradient boosting decision treeljbljkbs and one is based on Long ShortTerm memory. In all the cases, investment portfolios are built based on the predictions and performance is compared based on the return on investment. Mittal R. et al. [13] have predicted the cryptocurrency prices based on their open, low and high cost. Some Machine Leaning Algorithms are used to predict the cryptocurrency daily price changes. The Dataset used in this paper is relatively smaller. Multivariate Linear Regression has been used to predict the Highest and lowest prices of cryptocurrency. R is used as a platform to predict the cryptocurrency prices based on the dependent features. Various statistical measures like F Score and p value are used to test the model's accuracy and probability respectively. Khuntia S. and Pattanayak J. K. [14] have verified the emerging efficiency in the transition of bitcoin prices and states that the evidence of the dynamic efficiency adheres to the proposition of the AMH. The paper states also state that the speculators and arbitrageurs can exploit extra returns and the bitcoin price movement is either efficient or inefficient is not practically true. Wei W. C. [15] tests the market efficiency in the wide cross-section of cryptocurrencies. The liquidity of 456 different cryptocurrencies is examined and is shown that the return predictability diminishes in cryptocurrencies with high market liquidity. The paper also shows a strong relationship between the Hurst exponent and liquidity on a cross- sectional basis. It finally states that liquidity plays a significant role in market efficiency and return predictability of new cryptocurrencies.

IV. METHODOLOGY

We have taken the dataset from Kaggle which contains the dataset for bitcoin, Ethereum, Litecoin etc. The dataset includes columns like Date, Open, High, Close, Volume, Market Cap. We have represented the closing and opening price of bitcoin:

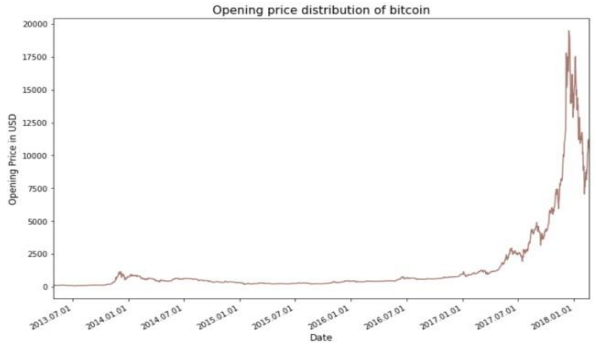


Fig. 4.1 Opening Price Distribution of Bitcoin

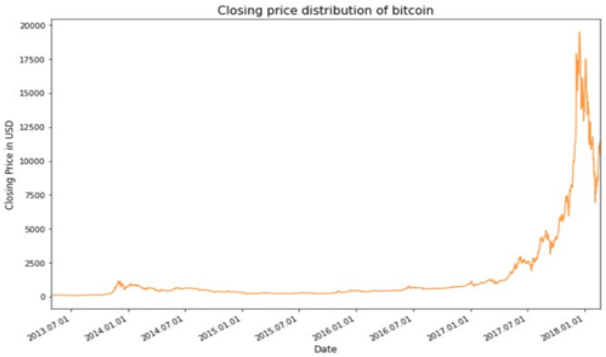
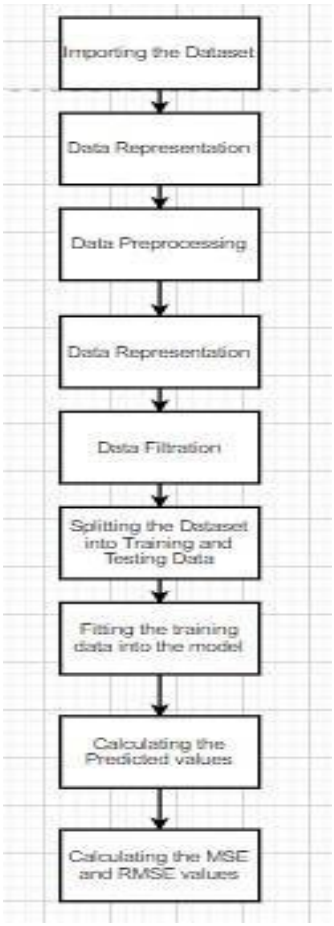
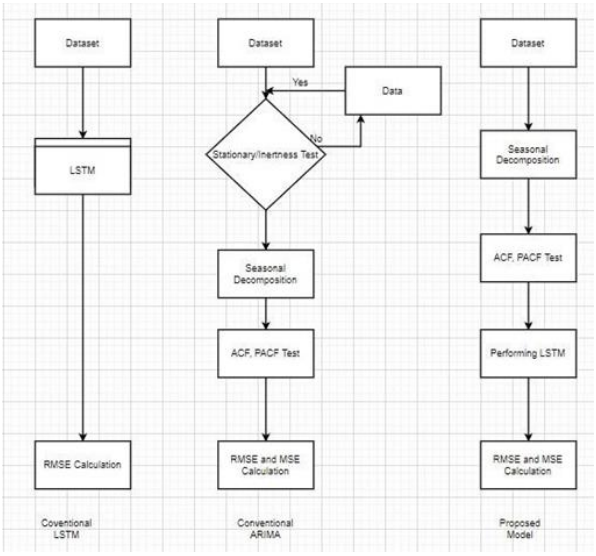


Fig. 4.2 Closing Price Distribution of Bitcoin

The basic design approach looked like the figure given below.



After looking closely to each and every model's Algorithm, we thought we could build an algorithm combining the 2 Time dependent models. So, this was our proposed Model which is diagrammatically represented below.



So, now there are 5 models through which we are predicting the cryptocurrency prices.

A. Linear Regression

It is a time independent model and thus, does not require time. In our project, we have taken Open and Close columns of the dataset to be put into the model. We have taken 80% data as training set and 20% as testing set. The accuracy of model is around 98.99%. After performing 6fold cross validation the accuracy found is 99.39%. The RMSE value found is around 180-280.

B. Support Vector Regression

It is a time independent model and thus, does not require time. In our project, we have taken Open and Close columns of the dataset to be put into the model. We have taken 80% data as training set and 20% as testing set. The accuracy of model is around 98.99%. After performing 6fold cross validation the accuracy found is 99.39%. The RMSE value found is around 200-300.

C. LSTM

They are a special kind of RNN. They have a tendency to persist information of the past. So, there is no need to start something from the scratch. Therefore, it is comparatively faster than all the other models. We have parsed the date values to match the pattern. After that we group the values by date. After separating the training and testing values we have taken a sequential regressor with 4 units, sigmoid function. There is one dense layer. We have used adam optimizer for this purpose. The batch size is 5 and we have used 100 epochs. The RMSE value is around 230300.

D. ARIMA

Auto Regressive Integrated Moving Average model is a time dependent series model. It is fitted to time series data to predict future points in the series. In our project, we use time as the input variable and train the model with the close price at a particular time. It results in predicting the closing price in future.

We have parsed the date values to match the pattern. We have sampled the dataset based on months. We have done seasonal decomposing to check the trend.

Dickey-Fuller test gives the p-value 0.9988 which show that timeseries is not stationary with 99.88% confidence. After that Box-Cox transformation is done to suppress some of the variance. The p-value becomes 0.511886. We have done seasonal differencing for differencing the data. After doing Seasonal differentiation of 12 months the p-value becomes 0.665690. We have done seasonal differencing for differencing the data. After doing Seasonal differentiation of 3 months the p-value becomes 0.017885 which shows that series become stationary. After that we have done Autocorrelation and Partial Autocorrelation factor test to identify the patterns in the data which is stationary based on mean and variance. After this we implemented ARIMA model. We have used AIC (Akaike Information Criterion) value to find the best fitted model. The best model passed the Dickey-Fuller test with p-value 0.005727. The RMSE value of this model is around 131.54.

E. Proposed Model (Hybrid model of LSTM and ARIMA)

This model is hybrid model of LSTM and ARIMA. So, the basic parsing of date values is same as of LSTM and ARIMA. We have taken 70% data as training set and 30% as testing set. We have done ACF-PACF test as we have done in ARIMA to find any seasonality in the dataset.

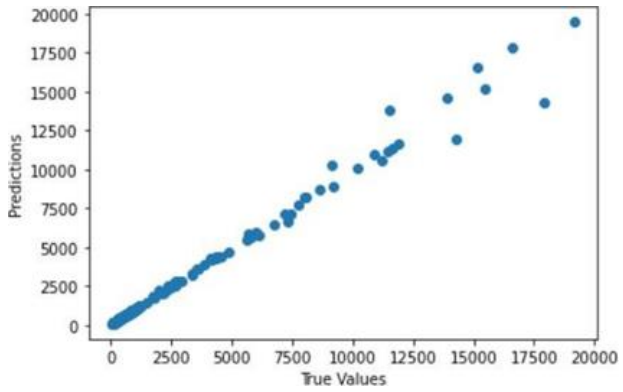
Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

After this we proceed with LSTM model with sequential model and one dense layer. The optimizer used is adam optimizer. The RMSE value of this model is around 255.55.

V. RESULT AND DISCUSSION

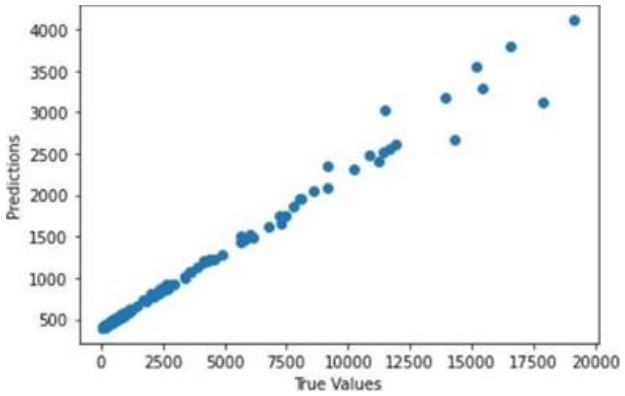
Our project deals with mainly calculating the RMSE values of all the models after fitting the same dataset to them and then comparing them. We have plotted the graphs of the predicted values with the corresponding actual values to get a vision of how much the model is accurate in predicting our dataset values. So, following are the graphs of all the 5 models.

A. Linear Regression



X-Axis → True Values
Y-Axis → Predicted Values

B. Support Vector Regression



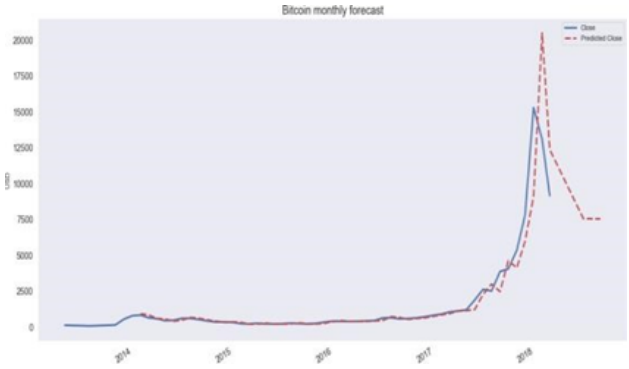
X-Axis → True Values
Y-Axis → Predicted Values

C. LSTM

X-Axis → Date
Y-Axis → Bitcoin Price

The red color line denotes the Real Bitcoin Close Value and the Blue color denotes the actual bitcoin close Value.

D. ARIMA



X-Axis → Date
Y-Axis → Bitcoin Price

The red color denotes the Bitcoin Naïve Forecast value, and the Blue color denotes the Closing Price of the Dataset.

E. Proposed Model

This is the model which we had proposed earlier and have been successful in implementing it and thereby calculating the RMSE value. We have basically incorporated the ARIMA model in the LSTM model. After doing so, the graph which we got is provided below.

X-Axis → Date
Y-Axis → Bitcoin Price

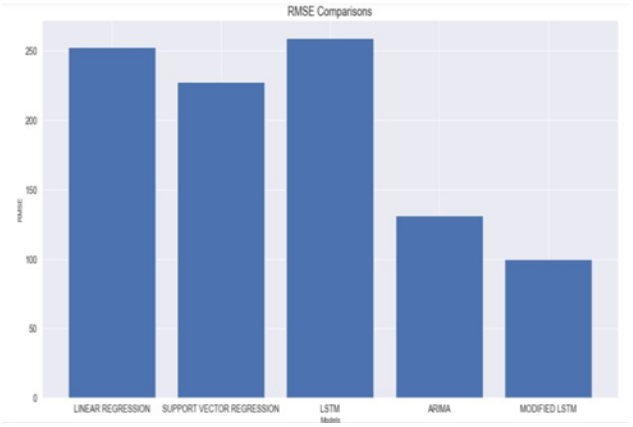
The red color denotes the predicted price, and the green color denotes the actual price of Bitcoin. We have compared these 5 models based on the RMSE Values. The following table shows the MSE and the RMSE values of the models.

Table- I: MSE vs RMSE

Model	MSE	RMSE
Linear Regression	67,096.5409	259.03
Support Vector Regressing	51,157.3924	226.18
LSTM	85,170.5856	291.84
ARIMA	17,300.1409	131.53
Proposed Model	3,741.7689	61.17

Based on the RMSE values we can observe that Linear Regression and Support Vector Regression models are not that much bad compared to the results of other models although they are time independent. The ARIMA models performed well than LSTM model. The main reason behind that is in ARIMA seasonality and trends are considered. Hence, we were able to correlate the values with some previous value. After implementing the concept of ARIMA in LSTM we have created a hybrid model which outperforms the ARIMA and LSTM model.

The plot of RMSE values of different models in as shown below:



Hence, we can say that seasonality and trend play a major role in deciding price of any cryptocurrency.

VI. CONCLUSION

Cryptocurrency in today's world is in a great demand. Big investors are betting on Cryptocurrency especially on Ethereum Datasets currently. Although there are several prediction models available today, the error rate is too high for us to trust the predicted value. So, we need a model which is accurate, precise and fast in predicting the future value. The major drawback in today's models is that they don't give the same result in all the environments. After testing the models, the conclusion can be gathered aa some models give good results with a particular dataset while bad results with the other datasets. Some models like SVM give good results but are very slow. Hence, they are not suitable for large datasets. A hybrid algorithm which will remove all the above said defects of all the present algorithms has been implemented. This algorithm uses features of ARIMA like seasonal decomposition and ACF-PACF test also it can work like a LSTM. The RMSE values comparison revealed the accuracy of model. The major thing to keep in mind while developing a new algorithm is that although, it takes a lot of time to pre-process the dataset, it should give an accurate result. Also, it be concluded that the models can perform differently based on the datasets. The models as well as training-testing sets should be chosen according to the datasets. The proposed model in this paper can perform well in cryptocurrency, stocks etc. related predictions. In future the model can be tested and modified well for further improvements related to performance and LSTM models. Hence, we can say that seasonality and trend play a major role in deciding price of any cryptocurrency.

REFERENCES

1. Yecheng Yao, Jungho Yi, Shengjun Zhai, Yuwen Lin, Taekseung Kim Guihongxuan Zhang, Leonard Yoonjae Lee, "Predictive Analysis of Cryptocurrency Price Using Deep Learning", International Journal of Engineering & Technology, Volume 7, Issue 3.27, pp. 258-264, 2018 [CrossRef]
2. Guus van Heijningen, "Making predictions in highly volatile cryptocurrency markets using web scraping", 2017-2018
3. Muhammad Ali Nasir, Toan Luu Duc Huynh, Sang Phu Nguyen and Duy Duong, "Forecasting cryptocurrency returns and volume using search engines", Financial Innovation 5, Article number 2, 2019 [CrossRef]
4. Jethin Abraham, Daniel Higdon, John Nelson and Juan Ibarra, "Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis", SMU Data Science Review, Volume 1, 2018
5. N. I. Indera, M. Yassin, A. Zabidi1 and Z. I. Rizman, "Nonlinear autoregressive with exogeneous input (NARX) bitcoin price prediction model using PSO and moving average technical indicators", Journal of Fundamental and Applied Sciences, Volume 3, Issue 3S, pp. 791-808, 2017 [CrossRef]
6. Aggarwal A., Gupta I., Garg N., and Goel A., "Deep Learning Approach to Determine the Impact of Socio-Economic Factors on Bitcoin Price Prediction", Twelfth International Conference on Contemporary Computing (IC3), 2019 [CrossRef]
7. Phaladisailoed T. and Numnonda T., "Machine Learning Models Comparison for Bitcoin Price Prediction", 10th International Conference on Information Technology and Electrical Engineering (ICITEE), 2018. [CrossRef]
8. Wimalagunaratne M., and Poravi G., "A Predictive Model for the Global Cryptocurrency Market: A Holistic Approach to Predicting Cryptocurrency Prices", 8th International Conference on Intelligent Systems, Modelling and Simulation (ISMS), 2018 [CrossRef]
9. Ross C. Phillips and Denise Gorse, "Predicting Cryptocurrency Price Bubbles Using Social Media Data and Epidemic Modelling", IEEE Symposium Series on Computational Intelligence (SSCI), 2017 [CrossRef]
10. Chongyang Bai, Tommy White, Linda Xiao, V.S. Subrahmanian and Ziheng Zhou, "C2P2: A Collective Cryptocurrency Up/Down Price Prediction Engine", IEEE International Conference on Blockchain (Blockchain), 2019
11. Brauneis A., and Mestel R., "Price discovery of cryptocurrencies: Bitcoin and beyond", Economics Letters, 165, pp. 58-61, 2018 [CrossRef]
12. Alessandretti, L., ElBahrawy, A., Aiello, L.M. and Baronchelli, A., 2018. Anticipating cryptocurrency prices using machine learning. Complexity, 2018. [CrossRef]
13. Mittal R., Arora S. and Bhatia M. P. S., "Automated cryptocurrencies prices prediction using machine learning", International Journal on Soft Computing, Volume 8, Issue 4, pp. 1758-1761, 2018
14. Khuntia S. and Pattanayak J. K., "Adaptive market hypothesis and evolving predictability of bitcoin", Economics Letters, 167, pp. 26-28, 2018 [CrossRef]
15. Wei W. C., "Liquidity and market efficiency in cryptocurrencies", Economics Letters, 168, pp. 21-24, 2018 [CrossRef]

AUTHOR PROFILE

Harshvardhan Sinha, is currently a Software Development Engineer in an MNC in Bangalore. He holds B Tech in Computer Science from VIT University, Vellore. He has been a core member as well as Vice President of IEEE MTT-S chapter during his B Tech tenure. He is a recipient of several awards in the industry as well as his college tenure. He holds immense interest in Web Development, Machine Learning and Database Management.

Sangle Gaurav Keshvrao, is currently a Software Development Engineer in an MNC. He holds B Tech in Computer Science from VIT University, Vellore. He holds immense interest in Web Development, Machine Learning and Data Science. He has played key roles in various projects in current company as well as in his college tenure.

Pulkit Verma, is currently a Software Development Engineer in an MNC in Bangalore. He holds B Tech in Computer Science from VIT University, Vellore. He holds immense interest in Data Mining, Machine Learning and Data Science. He has played important roles in various projects in current company as well as in his college tenure.