

A Deep Learning Approach to Anticipate Crypto Currency Trends Using Sentiment Analysis

Ardra.M*, Dr. Beena M.V**

*(Department of Computer Science & Engineering, Vidya Academy of Science And Technology, Thrissur
Email: ardranairm96@gmail.com)

** (Department of Computer Science & Engineering, Vidya Academy of Science And Technology, Thrissur
Email: beena.m.v@vidyaacademy.ac.in)

Abstract:

Cryptocurrencies are a growing market that are becoming more and more significant in the financial world. Recent research in big data analytics and natural language processing has resulted in the development of automated techniques for assessing sentiment in news headlines. In addition, the rising user base of social media and the availability of news give useful sentiment data for predicting cryptocurrency price fluctuations. In this research, we show how to use news sentiment data and market data to predict changes in Bitcoin and Ethereum prices. Bitcoin and Ethereum are the two most popular cryptocurrencies in terms of market capitalization. The research uses a deep learning model(Bi-Lstm) to increase the perfection of cryptocurrency price prediction by collecting a maximum amount of time series data and analysing it in connection to related headlines.

Keywords —Bitcoin, Ethereum, Cryptocurrency, Deep Learning, Bi-LSTM, Sentiment Analysis.

I. INTRODUCTION

Cryptocurrency is a type of virtual or digital money that exists in financial institutions. It cannot be replicated or falsified because of the encryption used to secure it. It also varies from conventional currencies in that it's a decentralised virtual money that may be converted via cryptographic methods and isn't issued by a central bank or other financial organisation. The other feature is that it was created using the extremely complex technology known as blockchain, which tries to store data that makes it difficult or impossible to change, hack, or cheat the system. Bitcoin has started to carve out a niche for itself, which may either help cryptocurrencies become more widely accepted or be the major reason they fail. It's difficult to predict whether or not cryptocurrencies will ever be used significantly in global markets because they are still in their infancy. Bitcoin, the most well-known cryptocurrency, was created in 2009 and was the

sole Blockchain-based cryptocurrency for over two years. However, the cryptocurrency market already has over 2 000 cryptocurrencies and 180 million active users. Bitcoin has recently received a lot of attention from the fields of economics, cryptography, and engineering because to its inherent ability to combine monetary units with encryption technology.

The core technology of the Bitcoin economic system is termed Blockchain (BC), and it's commonly recognised that the Internet of Things (IoT) ecosystem relies on this infrastructure because the backbone for ensuring enhanced security and privacy. Majority of it a distributed digital transactional ledger over the whole computer networkblockchain-based computer systems. The blockchain consists of two fundamental components; the first one may be a transaction, and also the second may be a block. The transaction represents the action triggered by the participant, and therefore the block could be a data collection that records the

transaction and additional details like the right sequence and creation timestamp. Each autonomous system (AS) within the blockchain network joins the defensive alliance employing a signalling mechanism called BloSS, which could be a multi-domain, blockchain-based, cooperative DDoS protection system. The implications of networks on competitive nature within the developing cryptocurrency market over time, as shown in terms of cryptocurrency exchange rates, are revealed by the reference.

Based on their market capitalization, we take into account the two cryptocurrencies bitcoin and ethereum for this research. We next gather news headlines about these coins from Google News and analyse their sentiments. We successfully predicted the value of cryptocurrencies for the upcoming week using these sentiment scores and market data using a bi-lstm model. Due to price volatility and dynamism, it is challenging to predict cryptocurrency prices. Numerous cryptocurrencies are used by customers all around the world. In this paper, we concentrate on the top two. The paper's goal is to develop a reliable model for Predicting Cryptocurrency Trends Using News Sentiment Analysis and Deep Learning.

The II section consists of literature survey, III section consists of the proposed system and final IV section consists of the results and section V is about conclusion and future work

II. LITERATURE SURVEY

Shaomi Rahman, Jonayed Nafis Hemel, Syed Junayed Ahmed Anta, Hossain Al Muhee, Jia Uddin [1] proposed a Sentiment Analysis Using R: An Approach to Correlate Cryptocurrency Price Fluctuations with Change in User Sentiment Using Machine Learning. Machine learning techniques were used to suggest a relationship between changes in Bitcoin's price and user sentiment. ten classification algorithms and five regression algorithms were used. The classification algorithms used are Logistic Regression, K-Nearest Neighbour, Support Vector Machine, Kernel SVM, Naive Bayes, Decision Tree Classification, Random Forrest Classification, Artificial Neural Network,

Principle Component Analysis, Linear Discriminant Analysis, and Kernel PCA. The regression models used are Multiple Linear Regression, Polynomial Regression, Support Vector Regression, Decision Tree Classification, and Random Forrest Tree Classification. Using K-fold Cross Validation, the accuracy of each algorithm was calculated. 89.65 percent is the highest accuracy rate that has been attained. The reason why the Naive Bayes algorithm performs best is because the Naive Bayes classifier is a typical illustration of how generative assumptions and The process of learning is made easier by parameter estimates.

Sean McNally, Jason Roche, Simon Caton [2] Predicting the Price of Bitcoin Using Machine Learning. This study aims to determine how accurately the direction of the price of bitcoin in US dollars can be anticipated. The Bitcoin Price Index is the source of the price information. A Bayesian-optimized recurrent neural network (RNN) and an LSTM network are used to execute the job, with varied degrees of success. The classification accuracy that the LSTM achieves with an RMSE of 8% is 52%. As a contrast to the deep learning models, the well-known ARIMA model for time series forecasting is used. The ARIMA model performs badly. The training time on the GPU outperformed the CPU implementation by 67.7 percent when both deep learning models were benchmarked on both a GPU and a CPU.

Siddhi Velankar, Sakshi Valecha, Shreya Maji [3] proposed a Bitcoin Price Prediction using Machine Learning. This paper makes an effort to provide an accurate prediction of the Bitcoin price while taking into account numerous factors that influence the Bitcoin value. Aim to understand and recognise daily market trends for bitcoins while obtaining knowledge of the best conditions surrounding the price of bitcoins throughout the initial phase of the study. The data set comprises of numerous characteristics linked to the Bitcoin price and payment network over a five-year period that were daily recorded. The sign of the daily price change will be predicted as accurately as feasible for the second part of the study utilising the information that is already available. The two models utilised in the study to tackle the Bitcoin prediction problem

are Bayesian Regression and GLM/Random Forest. These models were chosen after the learning framework had been established and the normalisation process had been completed.

Karunya Rathan, Somarouthu Venkat Sai, Tubati Sai Manikanta [4] proposed Crypto-Currency price prediction using Decision Tree and Regression techniques. The market for bitcoin is booming, and numerous studies in the economics and price prediction domains have been conducted. The suggested work uses machine learning models like Decision Tree and Linear Regression to analyse the Bitcoin dataset from 2011 to the present. Decision Tree and linear regression models are also used to estimate prices for the next five days. The suggested learning approach recommends the top algorithm to select and use for the crypto currency prediction challenge. Results of the experimental study demonstrate that linear regression performs better than the alternative with high accuracy for price prediction.

Stephen Wooley, Andrew Edmonds, Arunkumar Bagavathi†, Siddharth Krishnan [5] Extracting Cryptocurrency Price Movements From the Reddit Network Sentiment. Cryptocurrency prices are heavily impacted by public sentiment because, unlike traditional currencies, they are not backed by any government entities. It is possible to make better predictions of price movement by understanding the connection between cryptocurrency pricing and public opinion. An investigation into the price movements of Bitcoin and Ethereum using a network of 24 Reddit forums related to these cryptocurrencies. Using submissions and comments from the chosen subreddits, create a collection of 112 time series features, run Granger causality tests on these time series against changes in cryptocurrency prices, and then utilise these time series to predict changes in cryptocurrency prices using classification models. Results from these models confirm the Granger causality test findings that the direction of Bitcoin and Ethereum price movements can be predicted with 74.2 percent and 73.1 percent accuracy, respectively, using only lagged price values and lagged values from a single Reddit data derived feature.

Valdés-Aguirre [6] proposed Price Movement Prediction of Cryptocurrencies Using Sentiment Analysis and Machine Learning. The cryptocurrency market is a great topic for research due to its low entrance barrier and extensive data availability. Although there have been some prior research, the majority have only looked at the behaviour of Bitcoin. The use of popular machine learning technologies and readily available social media data is suggested in this study as a method for forecasting price changes in the Bitcoin, Ethereum, Ripple, and Litecoin cryptocurrency markets. Compare the use of market and Twitter data as input characteristics to neural networks (NN), support vector machines (SVM), and random forests (RF). Results indicate that at least one model can forecast market movements for the Bitcoin, Ethereum, Ripple, and Litecoin markets by outperforming chance in precision scores. This forecast just covers the market's direction; it excludes information on the size or length of any future market action. The most accurate and precise model for Bitcoin was an MLP, which achieved scores of 0.72 and 0.74 using both Twitter and market data. The best model for Ethereum was an MLP that used Twitter and market data to produce accuracy and precision scores of 0.44 and 0.56, which were not noticeably better than random. The best model in Ripple once more obtained accuracy and precision scores of 0.64 and 0.68 and was an MLP that solely relied on market data. Litecoin was the only cryptocurrency where the SVM model performed the best, using both Twitter and market data it obtained 0.66 accuracy and 0.8 precision scores.

Jiayun Luo [7] proposed a Bitcoin price prediction in the time of COVID-19. The goal of this study is to evaluate the efficacy of four different machine learning models on forecasting the Bitcoin return rate and price trend using data from the Bitcoin exchange, COVID-19, and Twitter from January 2020 to July 2020. Four input feature sets are used to generate the data: historical bitcoin exchange data, historical bitcoin exchange data plus COVID-19 data (recovery, confirmed, death), historical bitcoin exchange data plus Twitter, and historical bitcoin exchange data plus COVID-19 data plus Twitter. Random forest, Decision Tree, AdaBoost, and

Support Vector Machine are the four machine learning models that were applied. Finally, it was discovered that: (1) Twitter data can enhance model performance; (2) People consider information within 5 days when making investment decisions; (3) Support vector machine underperforms in predicting Bitcoin return rate or price trend; and (4) COVID-19 data underperforms in improving prediction. The latter assertion may or may not be true, but since we only have a small amount of COVID-19 data, it is difficult to say for sure.

Lokesh Vaddi, Vaishnavi Neelisetty, Bhavana Chowdary Vallabhaneni, Kolla Bhanu Prakash [8] Predicting Crypto Currency Prices Using Machine Learning and Deep Learning Techniques. This study looked into a number of methods for predicting the price of cryptocurrencies like Bitcoin. Compared the outcomes of Recurrent Neural Networks with LSTM cells, Linear Regression with Features, and Prediction using Linear Regression. The research contribution of this technique is that it used many features to train the model and predicted a numerical value of pricing rather than doing binary classification. With an accuracy of 96.2%, the Recurrent Neural Network model incorporating LSTM cells outperformed Linear Regression.

Huisu Jang and Jaewook Lee [9] proposed An Empirical Study on Modeling and Prediction of Bitcoin Prices with Bayesian Neural Networks Based on Blockchain Information. In this work, the recent extremely volatile Bitcoin prices are addressed by employing a BNN to examine the time series of Bitcoin price utilising Blockchain data in addition to macroeconomic indicators. The BNN model accurately predicted Bitcoin's fluctuations up to August 2017, which is a relatively recent period, according to empirical study. The BNN model succeeds in rather precise direction prediction, in contrast to other benchmark models that fall short. The BNN model is forecast to perform similarly with more recent data based on these experimental findings. It is expected that the growth and application of the BNN model will be useful for the analysis and prediction of the Bitcoin process as the variety of the Bitcoin process gained attention.

B. Mahir Iqbal, Muhammad Shuaib Iqbal, Fawwad Hassan Jaskani, Khurum Iqbal and Ali Hassan [10] proposed a Time-Series Prediction of Cryptocurrency Market using Machine Learning Techniques. A machine learning-based time series analysis has been used to forecast the market price and stability of Bitcoin in the cryptocurrency market. The ups and downs in the price of Bitcoin can be forecast using time-series analysis. Time series analysis using machine learning techniques such as ARIMA, FBProphet, and XG Boosting was done for this aim. Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R2 are the parameters used to evaluate these models. Finally, it was discovered that ARIMAX, with an RMSE of 322.4, is the best algorithm for predicting changes in the market price of bitcoin. While the RMSE scores for the FBProp and XGBoost algorithms are 229.5 and 369.

III. PROPOSED SYSTEM

This study aims to predict cryptocurrency prices correctly for the coming week. Due to price volatility and dynamism, it is challenging to predict bitcoin prices. Different cryptocurrencies are used by users all over the world. In this study, we concentrate on the top two, Bitcoin and Ethereum. The project's goal is to create a reliable model for predicting cryptocurrency trends using sentiment analysis of news stories and deep learning.

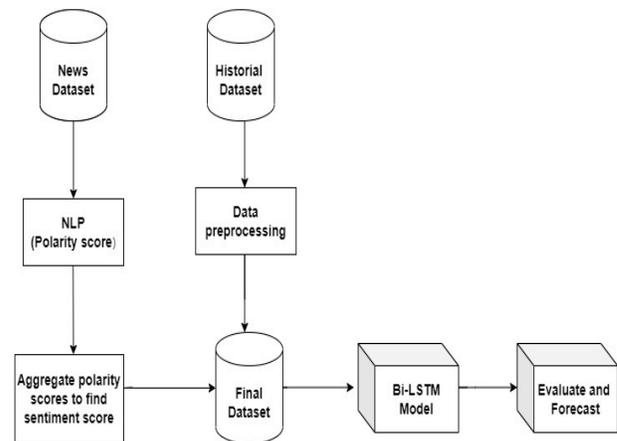


Fig 1. System Architecture

A. Data Set Creation

In this study, two data sets are utilised. The first is market information, which contains data on cryptocurrencies market values. The second is a data set of headlines with associated cryptocurrency news headlines.

(1)Market Data

During this research, Bitcoin and Ethereum were employed as cryptocurrencies. A financial website with 7 features, yahoo finance, was used to compile historical data on cryptocurrencies. dates, open prices, highs, lows, adj.close prices, and volumes. Data is gathered over the course of five years, from 2017 to 2022.

(2) News Data

Google News was used to get the news headline data. Information obtained through the web scraping technique. News headlines collected on a daily basis. By the end of the collection each year, over 7000 data,s are contained. Overall, data from 5 years were gathered.

B. Sentiment Aanalysis

Sentiment analysis was performed on the news headlines using the FinBERT sentiment analysis tool, which is integrated into the Natural Language Tool Kit. A pretrained NLP model called FinBERT analyses the tone of financial content. It is created by further honing the BERT language model for financial sentiment categorization utilising a large financial corpus and additional training in the finance industry. The application takes sentences as input and generates four different ratings utilising vocabulary, including negative, positive, neutral, and compound, which is the greatest score among the three scores. Compound scores were introduced as a third column to the news data frame, which previously only included the date column, and were sent to the sentiment analysis programme with each new headline in our data frame. Each news headlines calculated sentiment score ranges from -1 to 1. When the emotion score is close to 1, it means the news is positive; when it is close to -1, it means the news is negative; and when it is close to 0, it means the news is neutral. The sentiment scores of all news headlines published on a given day are then combined to determine the overall sentiment scores

for that day using the news headline data that has been grouped by date. The crypto market typically attracts a lot of attention from traders and investors on days when the general sentiment is favourable.

C. Correlation Analysis

Here we show the degree of correlation between our data's final set of features. The correlation matrix of Bitcoin and Ethereum is created as shown in figures 2 and 3. We can infer a high correlation between the close, open, high, low, average daily average, and volume features from the matrix.

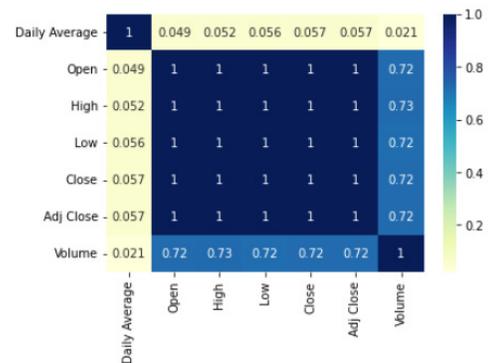


Fig 2. Correlation Matrix of Bitcoin

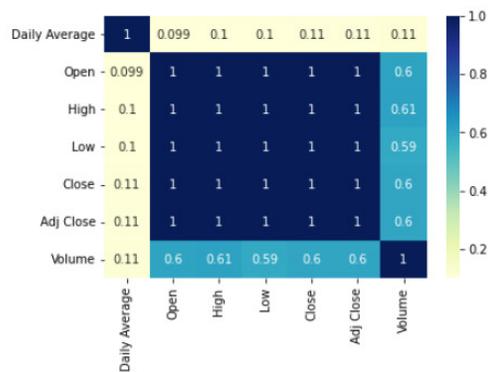


Fig 3. Correlation Matrix of Ethereum

D. Bi-LSTM(Bi-Directional Long Short Term Memory)

Bi-LSTM was created by Schuster and Paliwal to train a network utilizing past and future input data sequences. Bidirectional recurrent neural networks(RNN) are really just putting two independent RNNs together. This structure allows the networks to have both backward and forward information about the sequence at every time step. Using bidirectional will run your inputs in two ways,

one from past to future and one from future to past and what differs this approach from unidirectional is that in the LSTM that runs backwards you preserve information from the future and using the two hidden states combined you are able in any point in time to preserve information from both past and future.

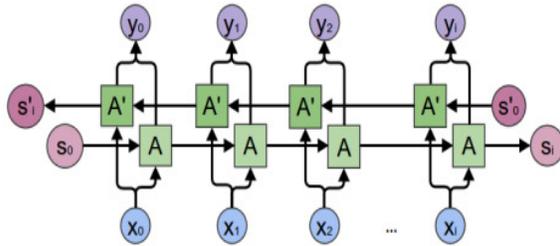


Fig 4. Bi-LSTM Architecture

IV. RESULTS

This section presents the findings from the Bi-Lstm model's forecasting of the future of Bitcoin and Ethereum as cryptocurrencies. The evaluation metrics are (1) Mean Squared Error (MSE), which is calculated by squaring the difference between your model's predictions and the actual data and averaging it across the entire dataset, and (2) Mean Absolute Error (MAE), which is the result of comparing two continuous variables. (3) Root Mean Square Error (RMSE) measures the deviation between predicted and actual values using a Euclidean distance. (4) Mean Absolute Percentage Error (MAPE), a frequently used regression performance statistic, is the average of absolute relative errors (5) Median Absolute Error (MedAE), calculates the median deviation between predictions and actual values. The result is given in Figures 5 and 6 which demonstrate how effectively the Bi-Lstm model predicts the prices of Bitcoin and Ethereum.

MSE	=	0.413791
MAE	=	0.538931
RMSE	=	0.643266
MAPE	=	0.473686
MedAE	=	0.487599

Fig.5. Bitcoin Evaluation Result

MSE	=	0.105359
MAE	=	0.256951
RMSE	=	0.324591
MAPE	=	0.22337
MedAE	=	0.212358

Fig.6. Ethereum Evaluation Result

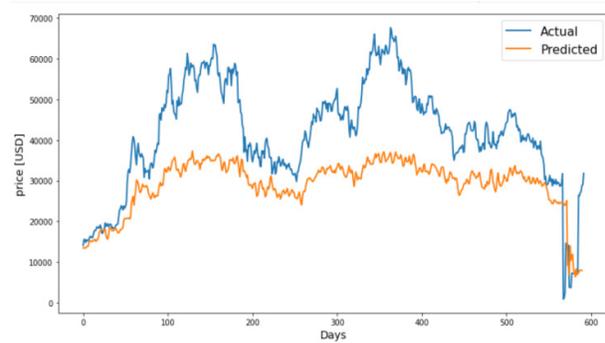


Fig .7. Predicted price Vs Original price of Bitcoin

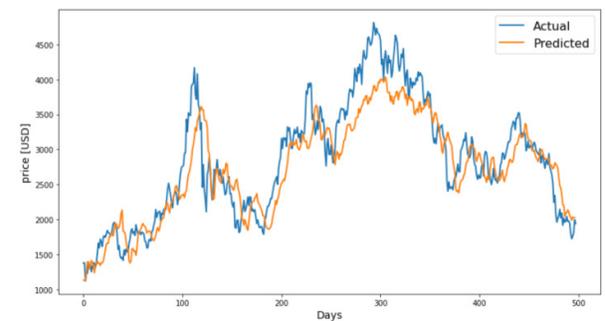


Fig.8. . Predicted price Vs Original price of Ethereum

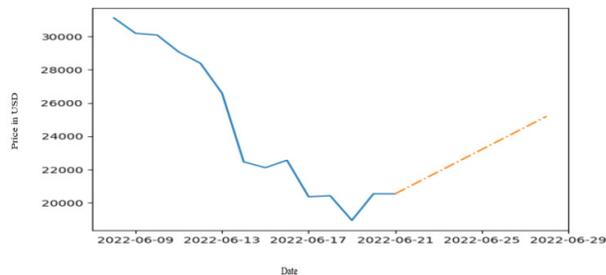


Fig.9. Forecasted Bitcoin price for the upcoming week

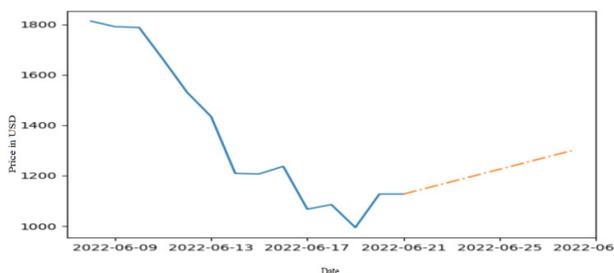


Fig.10. Forecasted Ethereum price for the upcoming week

v. CONCLUSION AND FUTURE WORK

In this paper, we demonstrated that deep learning and sentiment analysis can be used to forecast the direction of price fluctuations for the emerging cryptocurrency market. Using news headlines and market data, we assessed the effectiveness of the Bi-Lstm prediction model for Bitcoin and Ethereum. The findings imply that there is a significant correlation between news headlines and cryptocurrency prices. We'll leave future research on the various approaches to creating cryptocurrency forecast models to them. Some of these include increasing the quality of the content and the number of sources from which such content is acquired, which will improve the application of sentiment analysis for gathering social signals. By removing duplicate information and screening out content from bots or ads, quality could be improved. It is possible and probably advantageous to use content from other social networks such as Facebook or Reddit.

REFERENCES

- [1] Shaomi Rahman, Jonayed Nafis Hemel, Syed Junayed Ahmed Anta, Hossain Al Muhee, Jia Uddin, Sentiment Analysis Using R: An Approach to Correlate Cryptocurrency Price Fluctuations with Change in User Sentiment Using Machine Learnin, 2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV 2018)
- [2] Sean McNally, Jason Roche, Simon Caton, Predicting the Price of Bitcoin Using Machine Learning. 26th Euromicro International Conference on Parallel, Distributed, and Network-Based Processing. DOI 10.1109/PDP2018.2018.00060 (2018)
- [3] Siddhi Velankar, Sakshi Valecha, Shreya Maji , Bitcoin Price Prediction using Machine Learning. International Conference on Advanced Communications Technology(ICACT), ISBN 979-11-88428-01-4 ICACT2018 February 11 ~ 14, 2018(2018)
- [4] Karunya Rathan, Somarouthu Venkat Sai, Tubati Sai Manikanta , Crypto-Currency price prediction using Decision Tree and Regression techniques, Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019) ,IEEE Xplore Part Number: CFP19J32-ART; ISBN: 978-1-5386-9439-8.
- [5] Stephen Wooley, Andrew Edmonds, Arunkumar Bagavathi†, Siddharth Krishnan , Extracting Cryptocurrency Price Movements From the Reddit Network Sentiment, 2019 18th IEEE International Conference on Machine Learning and Applications (ICMLA 2019).
- [6] Franco Valencia, Alfonso Gómez-Espinosa and Benjamín Valdés-Aguirre, Price Movement Prediction of Cryptocurrencies Using Sentiment Analysis and Machine Learning, Entropy 2019, 21, 589; doi:10.3390/e21060589(2019)
- [7] Jiayun Luo , Bitcoin price prediction in the time of COVID-19, 2020 Management Science Informatization and Economic Innovation Development Conference (MSIED), 978-1-6654-1541-5/20/\$31.00 ©2020 IEEE, DOI 10.1109/MSIED52046.2020.00050(2020)
- [8] Lokesh Vaddi, Vaishnavi Neelisetty, Bhavana Chowdary Vallabhaneni, Kolla Bhanu Prakash, Predicting Crypto Currency Prices Using Machine Learning and Deep Learning Techniques, International Journal of Advanced Trends in Computer Science and Engineering ,Volume 9, No.4, July – August 2020, ISSN 2278-B3091 (2020)
- [9] Huisu Jang and Jaewook Lee , An Empirical Study on Modeling and Prediction of Bitcoin Prices with Bayesian Neural Networks Based on Blockchain Information, This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2017.2779181, IEEE Access(2020)
- [10] Market using Machine Learning Techniques, EAI Endorsed Mahir Iqbal, Muhammad Shuaib Iqbal, Fawwad Hassan Jaskani, Khurum Iqbal and Ali Hassan, Time-Series Prediction of Cryptocurrency Transactions on Creative Technologies Online First, (2021)