Predicting The Prices Of Bitcoin Using Data Analytics

Dr. M. Sharmila Begum[1], G. Jayashree[2], Z. Mahaboob Asfia[3], R. Subhasri[4] and L. Vishnu Priya[5]

[1]Associate Professor, [2,3,4,5] Final Year Students, Department of Computer Science and Engineering
Periyar Maniammai Institute of Science and Technology, Vallam, Thanjavur
[1]sharmila_se@pmu.edu.[2]jayashree.g.129@gmail.com,[3]mahaboobasfia@gmail.com,[4]subhasri1152000@gmail.com, [5]vishnupriya9515@gmail.com

ABSTRACT: The foremost aim of our paper is to predict next-day and any particular month Bitcoin prices with respect to the company as early as possible. To obtain results at the earliest we made our implementation in Apache Spark, a big data tool. We have also utilised one of the widely used machine learning libraries namely pandas for dataset manipulation, and preferred Pyspark since it is the combination of Apache Spark and Python. For investor interaction with our system we have designed a Graphical User Interface (GUI) and named it as 'PMIST' with Tkinter which is a Python's GUI. The result predicted will be seen in the form of line and bar graphs along with a message prompt where right date for doing investments are suggested. By analyzing those graphs, investors can be able to get idea about the future prices and they can take decision to either invest in future or change their investment time. Also a rewarding system is designed for the investors in which we will provide 50% offer in Swiggy when a quiz has been answered correctly. On the whole, this paper is meant for predicting next day and/or any particular month Bitcoin prices along with the rewarding system for the investors.

Keywords: Bitcoin; GUI; Apache Spark; Pyspark; Graphs; Rewarding system.

1. INTRODUCTION

Cryptocurrency is a technology in which a digital asset work as a medium of exchange wherein the ownership records are stored in a ledger existing in a form of computerized database using strong cryptographic techniques. This secures transaction records, controls the creation of additional coins, and verifies the transfer of coin ownership. It is often termed as a soft currency which doesn't available in the form of hard notes physically. Moreover cryptocurrencies are decentralized that without any third-party intervention all virtual currency users can get the services. Meanwhile, these services impact on international relations and trade, due to its high price volatility and fluctuations. There are several virtual currencies namely Bitcoin, ripple, ethereum, ethereum classic, lite coin, etc. Though many types of virtual currencies exist, Bitcoin has a greater acceptance from different bodies such as investors, researchers, traders, and policy-makers. Moreover this paper can make even new entrepreneurs to do investments in Bitcoin similar to gold by just analyzing our system results. Hence this paper aims at Bitcoin prices prediction.

It is one of the decentralised crypto currencies in which all can contribute their fiat currency and obtain corresponding Bitcoins in their own fashion rather than relying on government. Nevertheless, it has gained a lot of attention from social media and among public due to the fluctuation in Bitcoin prices. People who invested in those digital markets had suffered from severe loss in their businesses. That too for new entrepreneurs investing in a digital market is not preferable until they get familiar with cryptocurrencies. Now here comes the role of computationally quick prediction system with Data Analytics. It is a technique of analyzing the raw data and making decisions based on the information obtained from the data. So we feel Data Analytics as a fruitful technique to forecast the Bitcoin price fluctuations in a quicker manner so that we can prevent the investor to do investments at the wrong time.

1.1 BITCOIN

Bitcoin (abbreviated as BTC) is a soft currency that was emerged in January 2009. It follows the ideas of Satoshi Nakamoto which was set in a mysterious whitepaper. Actually Bitcoin is one of the cryptocurrencies where cryptocurrency is a digital asset.
Bitcoin address consists of 26-35 letters and numbers that always begin with the number 1 or 3 for representing the destination of Bitcoin payment. The currently available two address formats are 1BvBMSEstWetqTFn5Au4m4GFg7xJaNVN2 and 3MXknxVapwv6QkMoQv99MBuXZ2XpPewHn9.

1.2. DATA ANALYTICS

As the name indicates data analytics is a technology designed to analyse, process and extract from large datasets in which the traditional data processing software could never deal with. This technology is classified as operational big

data technologies and analytical big data technologies. The operational big data is all about the normal data that we generate in our daily life and this is just opposite to the analytical big data technologies. Online ticket bookings, online shopping, data from social media sites, the employee details of MNC Company are some of the examples for operational big data technologies.

Analytical big data is a complex and advanced version of big data technologies. Here the actual performance and the crucial real-time business decisions are made by analyzing the operational big Data. Its examples include space mission, weather forecast, patient's health status monitoring system and stock marketing.

1.3. PRICES PREDICTION

Due to the high volatility and price fluctuations, it is really difficult for the investors to forecast even the next day prices of the Bitcoin. Therefore it is very essential to predict the Bitcoin prices which make the investor to do investments at the right time. We also aimed to display the future Bitcoin results along with the company name. For that, we feel data analytics is the best technology which forecast the results within few minutes.

1.4. PROBLEM STATEMENTS

1. Many algorithms like Support Vector Machine (SVM), Linear Regression had taken more time to make predictions.
2. Those algorithms forecasted either next day or a month prices and not both.

1.5. PROJECT OBJECTIVES

To develop a fast computing prediction system for forecasting next day and any specific month Bitcoin price fluctuations with respect to the company by making use of the Graphical User Interface 'PMIST App' along with the rewarding module.

1.6 PROJECT SCOPE

Many have started to accept Bitcoin as a valid currency. By the year 2030, nearly half of the world will march towards the non cash transactions. The field cryptocurrency is really largest with the powerful market and it cannot be banned by any other countries. Many will start doing investments in Bitcoin and treat it similar to gold. In our project, a wide scope is there just because of our rewarding system where the investor can enjoy the Swiggy offers. And the main part of our project is apart from the investment time, investor can also view the companies list. It makes the investors to invest in right company along with the right time.

2. LITERATURE REVIEW

1. To reduce the time complexity [1] have linked their algorithm with one of the artificial intelligence technologies namely LASSO (least absolute shrinkage selection operator). In data pre-processing stage, they used data mining techniques which produced quick results with high accuracy and does not work well on high correlation.
2. Ziaul Haque Munim, et al. (2019) made predictions on the Bitcoin prices using autoregressive integrated moving average (ARIMA) and neural network auto regression (NNAR) models. For that, they employed a static forecast approach, which forecast next-day Bitcoin price with and without re-estimation of the forecast model for each step. To check the superiority of forecast results they included the Diebold Mariano test. Both NNAR and ARIMA model produced good accuracy but the re estimation approach used in this network consumes more time.
3. Mohammed Mudassir, et al. (2020) demonstrated a machine learning based classification and regression models in short and medium terms. Here they predicted the Bitcoin prices for one, seven, thirty and ninety days. This paper also used SVM, LSTM, Stacked artificial neural network and ANN. To evaluate the results they preferred regression models. The error rate was as lower as the previous system and the accuracy produced by the system was only about 65%.
4. Nicola Uras, et al. (2020) forecasted the closing prices of the cryptocurrencies like Bitcoin, Litecoin and Ethereum by making use of data on prices and volumes of prior days. Simple Linear Regression (SLR) model, Multiple Linear Regression (MLR) model, Multilayer Perceptron (MLP) and Long Short-Term Memory (LSTM) are the methodologies handled by them for prediction. Performs better in terms of time complexity but it is worse in the case of MAPE error.
5. Suhwan Ji, et al. (2019) used various deep learning approaches to predict the prices. The approaches used by them include deep neural network (DNN), long short-term memory (LSTM) model, a convolutional neural network, a deep residual network, and its combinations. They made use of Bitstamp Bitcoin market (USD) time series data as the dataset. The results demonstrated through LSTM model outperformed than other existing models and DNN model shown the best performance in price ups and downs. Also the accuracy of this model turned high. But this model takes too much time for comparison.
6. Zeba Ayaz, et al. (2020) collected the historical data of Bitcoin prices from the years 2013 to 2019 and made predictions for the year 2020. The core methodology used in this project is ARIMA model and made only the closing

price predictions for first seven days of January 2020 with ASP.NET. Though this model can able to forecast large number of time series datasets, it was unable to respond immediately.

7. Thearasak Phaladisailoed and Thanisa Numnonda (2018) used various machine learning algorithms for prediction. The step by step procedure used in this model includes Data selection, feature selection, data preparation, modeling, LSTM, RNN. However this takes more time to calculate than Huber regression.

8. Aniruddha Dutta, et al. (2019) used a set of advanced machine learning methods which involves a fixed set of exogenous and endogenous factors for prediction. Also analysed and compared various machine learning approaches using the root mean squared error (RMSE). This model outperformed well than other traditional models. But it posses limited number of features thus reduces the model performance.

9. S. Tandon, et al. (2019) given many neural network approaches in particular Recurrent Neural Network (RNN) and Long Short Term Memory (LSTM) along 10-fold cross validation. To increase the efficiency of this model new activation functions were also included. Along with RNN and LSTM, methods like Linear Regression and Random Forest applied in the same domain. Keras, Tensorflow and Scikit Learn like some packages utilised for experimental work. Enhanced accuracy was seen in terms of Mean Absolute Error (MAE). Cross validation technique make the Bitcoin dataset unstable.

10. Zheshi Chen, et al. (2020) extended the feature dimensions and evaluated different machine learning techniques for solving problems of multiple frequencies of Bitcoin prices. It includes both statistical and neural networks. For that they implemented the approaches namely Bayesian regression, generalized linear model (GLM), machine learning. Machine learning models include Random Forest, XGBoost, Quadratic Discriminant Analysis, Support Vector Machine and Long Short-term Memory tends to be more superior to statistical models. CoinMarketCap.com is the site they used to acquire the dataset. Apart from accuracy, the mean square error is as large as the standard deviation.

3. PROPOSED SYSTEM

We aimed to forecast tomorrow and a month Bitcoin prices with respect to the companies along with rewarding system and an interactive GUI as shown in Fig.1. Various sites like Coinmarketcap, Kaggle, bitinfocharts, and coindesk are available for Bitcoin which contains the real-time data about its prices. But the datasets available in those sites are more common without mentioning anything about the firm name in which the investor invested in it. So in order to demonstrate the company wise Bitcoin prices we have generated the dataset by our own. For that we took a list of seven companies and created the dataset with the necessary insights which is discussed in the dataset used session. After the system is predicted by the Apache Spark, a Graphical User Interface named 'PMIST' is created by the Tkinter in which the investor can view the results generated as line graph for month wise prediction and bar graph for next day Bitcoin prices prediction. Also GUI includes description about the project developers and our project details.

Finally, a rewarding system is designed to appreciate the investor participation with our application which asks mail from the investor and displays a simple quiz related to Bitcoin. If it is correctly answered, a mail stating 50% offer in Swiggy with the coupon number is sent. However, for wrong answers 10% offer will be given for any Swiggy orders. Thus prediction is made quickly along with the useful rewarding system.
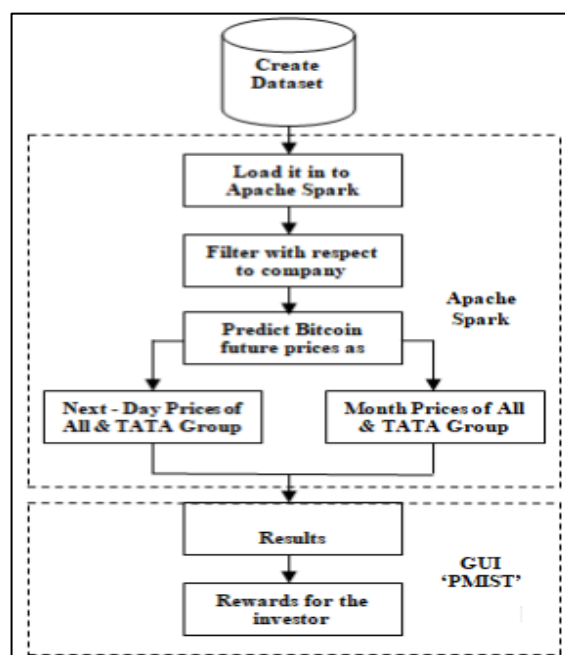

Fig. 1 Flow Diagram

4.   DATASET USED

As our prediction involves companies, we created the dataset by our own. It includes seven companies namely Indian Oil Corporation, TATA Group, Cognizant Corporation, Reliance Group, Adani Group, Hindustan Petroleum Corporation and Zoho Corporation starting from 01-01-2000 to 31-12-2020 along with the Bitcoin prices like starting price, ending price, highest price and lowest price.

5.   IMPLEMENTATION

5.1.   PRE - PROCESSING MODULE

It is the starting phase of our project. Here basic setups like Apache Spark, Pyspark, Java and Pycharm are installed. Apache Spark can run its applications using Java at background. Pycharm is the editor used to implement the spark applications whereas Pyspark is the python API which supports Apache Spark.

Before loading, a spark session is developed which acts as the entry point to create any spark applications. Then the dataset is loaded into the Apache Spark to handle and process vast kinds of data in clusters. The file always gets loaded as a Data Frame and spark reads the file in CSV format.
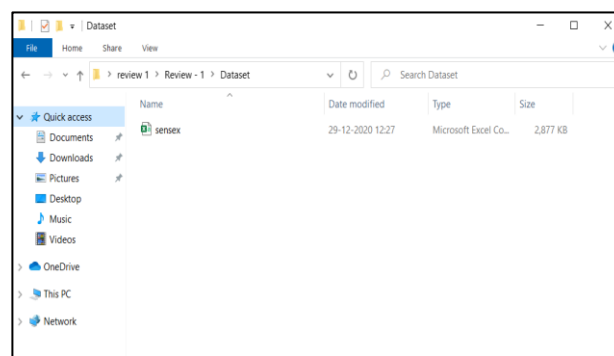
Fig. 2 Generated dataset named 'sensex'
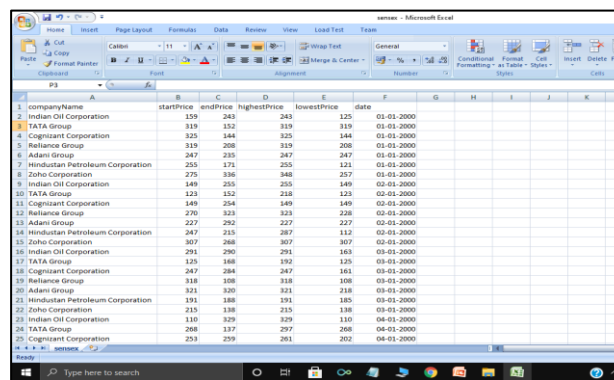
**Fig. 3 File 'sensex' with starting date
'01-01-2000'of All Companies**
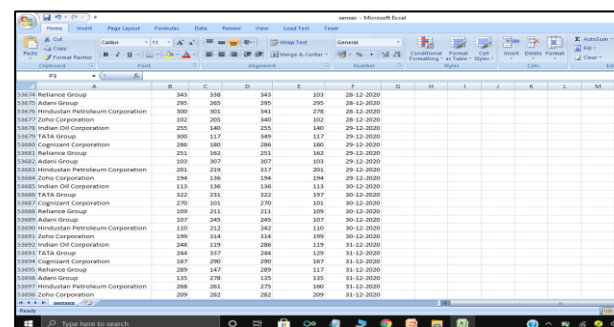
Fig. 4 File 'sensex' with ending date '31-12-2020'of All Companies

Next is the filtering phase. It is done for extracting only necessary information from the dataset. We have utilized Pyspark Data Frame along with its operations like filter and it's built in methods like count, show and performed two kinds of filtering as shown in Fig. 5 and Fig. 6. To filter the company wise Bitcoin prices, Pyspark uses filter ( ) function. This filters rows in the data frame as per the conditions.
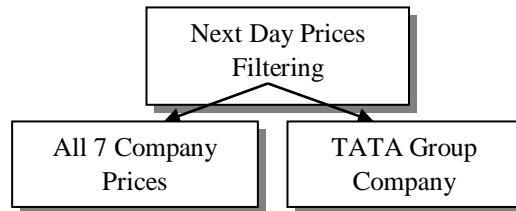
Fig. 5 Filtering Approaches for Next Day Bitcoin prices of All Companies and TATA Group Company
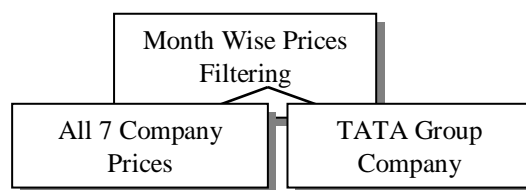
Fig. 6 Filtering Approaches for Month Wise Bitcoin prices of All Companies and TATA Group Company

5.2.   PREDICTION MODULE

Here the individual worker nodes present in the spark cluster can train a subset of models in parallel with other worker nodes. The key mechanism for achieving distributed data processing in Spark is the Data Frame. By loading the data into a Spark Data Frame, the data is distributed across the workers in the cluster. This reduces the overall amount of time required to perform our work. Like filtering, we followed two approaches for prediction i.e. next-day and month.

To predict next day prices of TATA Group, we have converted the Spark Data Frame in to Pandas Data Frame. Then we created three lists which stores actual prices, 10% highest and 10% lowest of the actual prices respectively. Similarly for predicting next day prices of all seven companies, 7 empty lists are created and the values are iterated inside a loop. Meanwhile, the graphs are plotted for those two predictions with the help of matplotlib library. To predict any month prices of all companies and a single company (TATA Group) we took average of all the actual prices present in our dataset and made predictions.

Next is the Graphical User Interface. For effective investor interaction with our project we have developed a GUI with Tkinter. We named our GUI as 'PMIST'. The details present in the GUI includes our project title, organization name, prediction duration i.e. next day or month, about the project and about the project developers.

5.3.   REWARDING MODULE

To appreciate the investor participation with our system, we have designed a rewarding module. In this module, investor mail id is asked and a Bitcoin quiz will be posted in that GUI. If they answer correctly and give their mail id, we will send a credit mail to them stating that he or she is having a 50% offer in Swiggy and 10% for wrong answers. Also a coupon code is given in that mail so that the investors can use that code in Swiggy and enjoy the offer. However it is totally up to the investor to enjoy our rewards without any compulsion. The entire process is given as a flow diagram in Fig. 1.

6.   DISCUSSION

In the study of referenced algorithms we feel that Bitcoin price prediction system with Data Analytics seems to produce the results in a quicker way than any other existing systems.
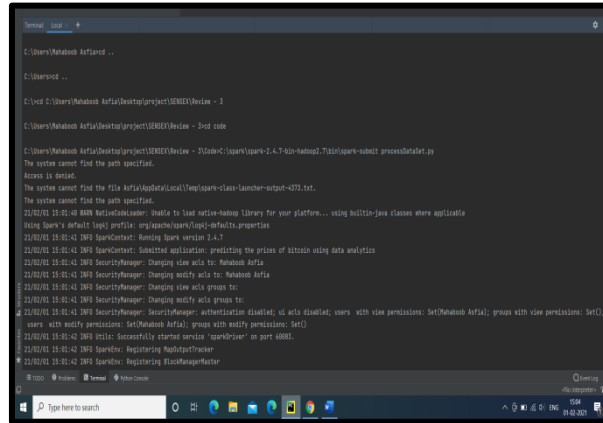
7.   RESULTS

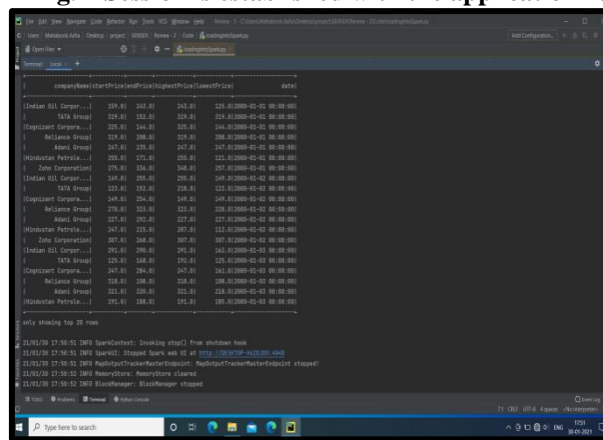**Fig. 7 Session is established with the application name**

Fig. 8 Dataset is loaded in to the Apache Spark environment
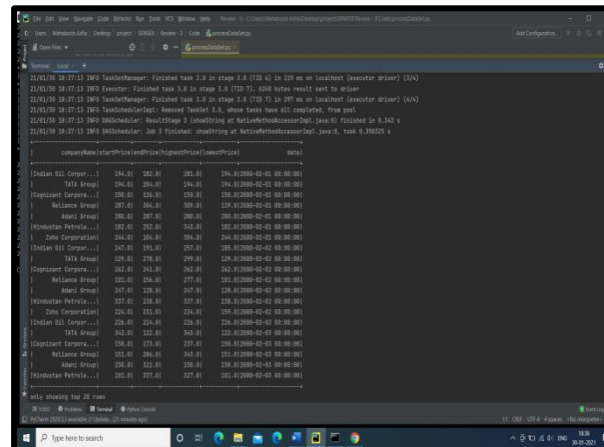

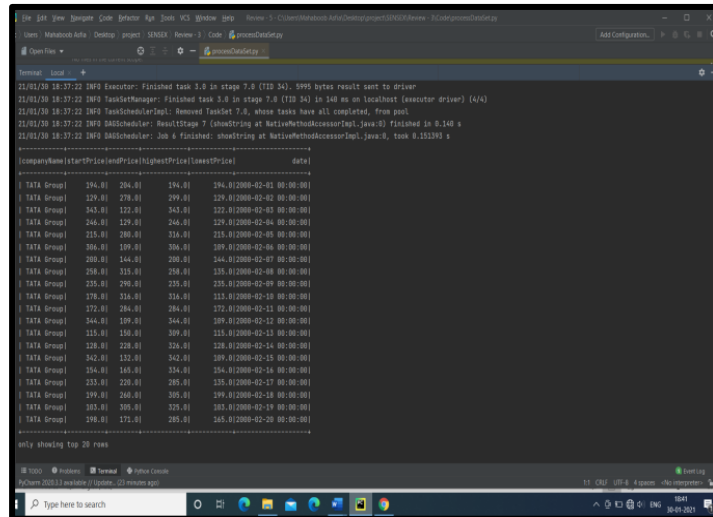**Fig. 9 Bitcoin February Month Prices are filtered for All Companies**

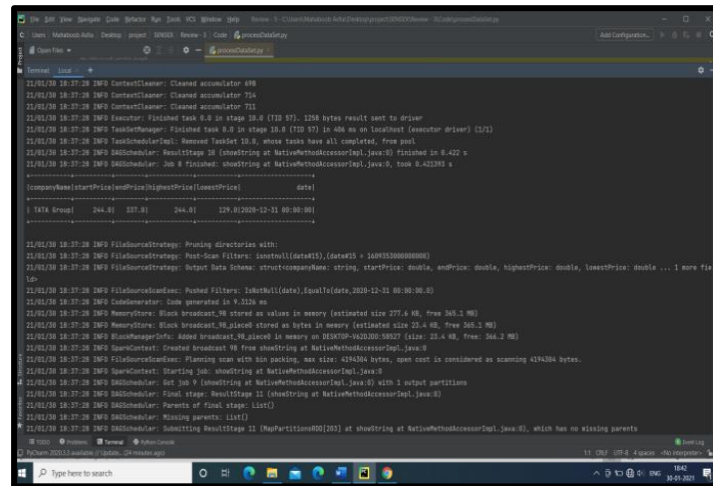**Fig. 10 Bitcoin February Month Prices are filtered for TATA Group**

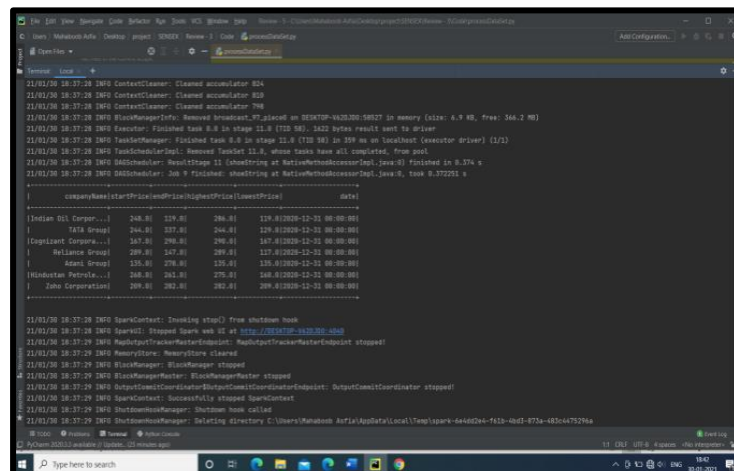**Fig. 11 Today Prices of Bitcoin are filtered for TATA Group**

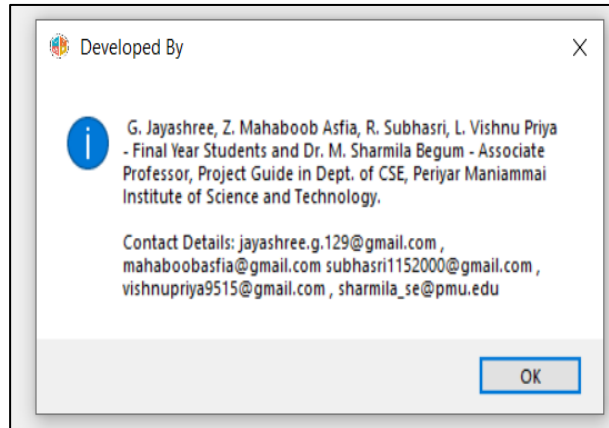**Fig. 12 Today Prices of Bitcoin are filtered for All Companies**

Fig. 13 A GUI named 'PMIST' is designed which shows the list of developers for reference
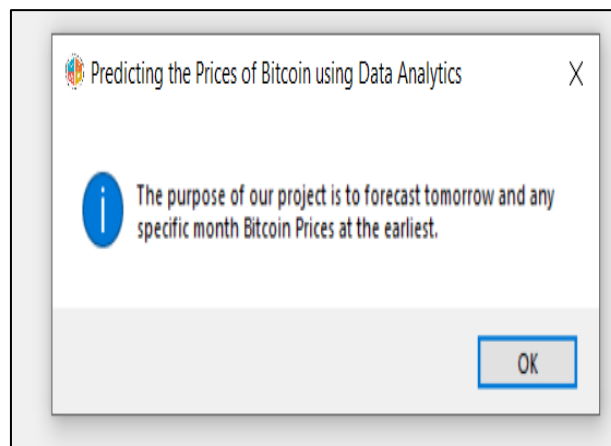
Fig. 14 GUI depicts the purpose of our project
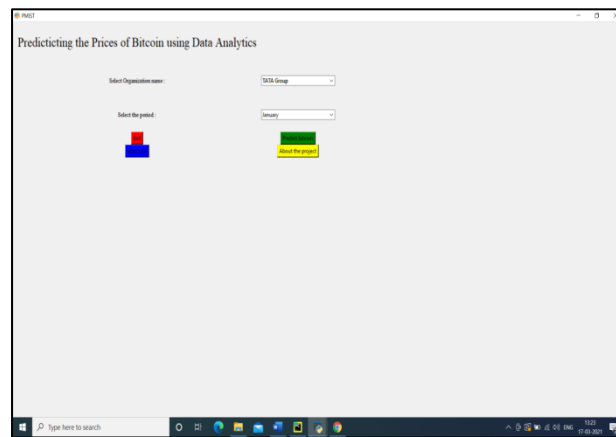
Fig. 15 January Month is given in the GUI to predict the prices of TATA Group
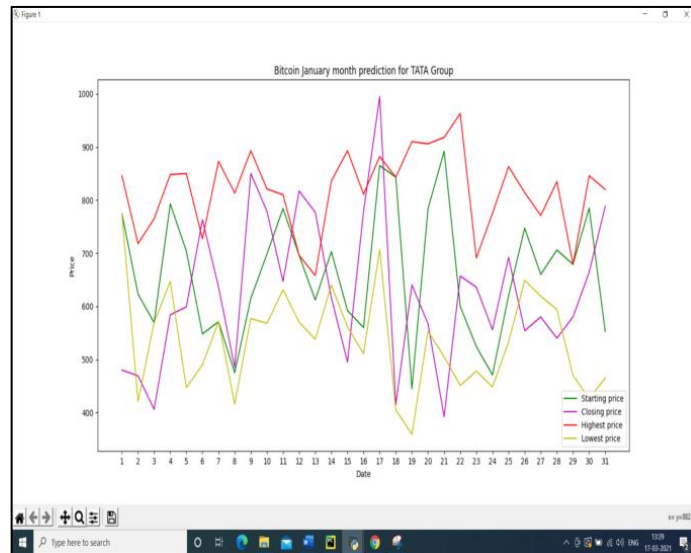
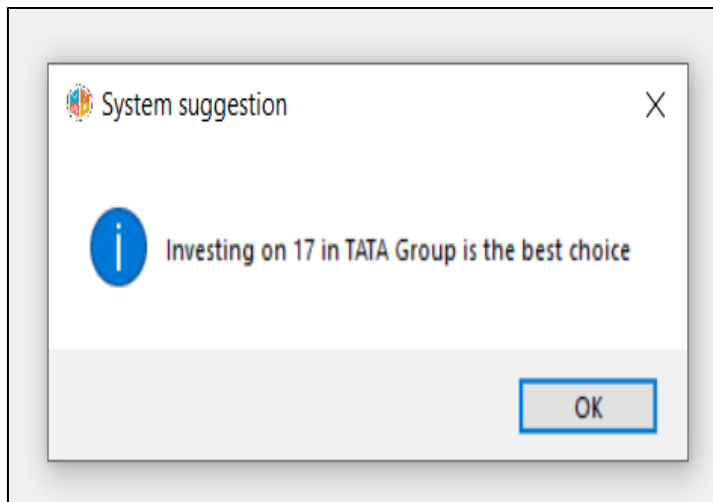**Fig. 16 Prediction results of Bitcoin prices for TATA Group on January**
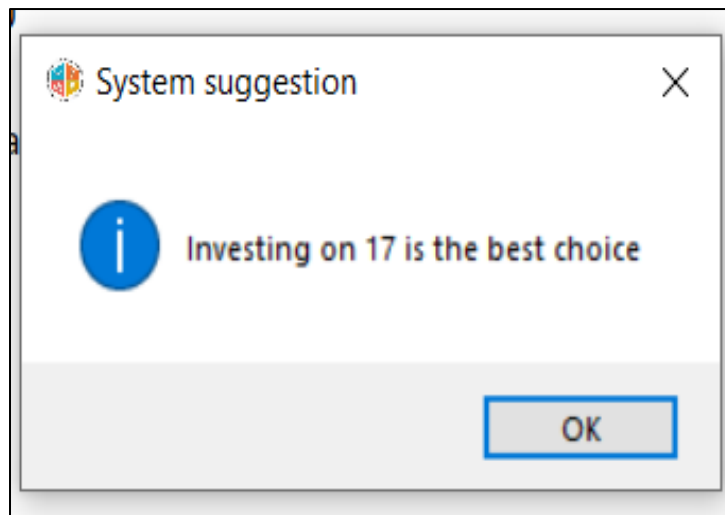
Fig. 17 System suggestion for investor betterment (TATA Group - January)
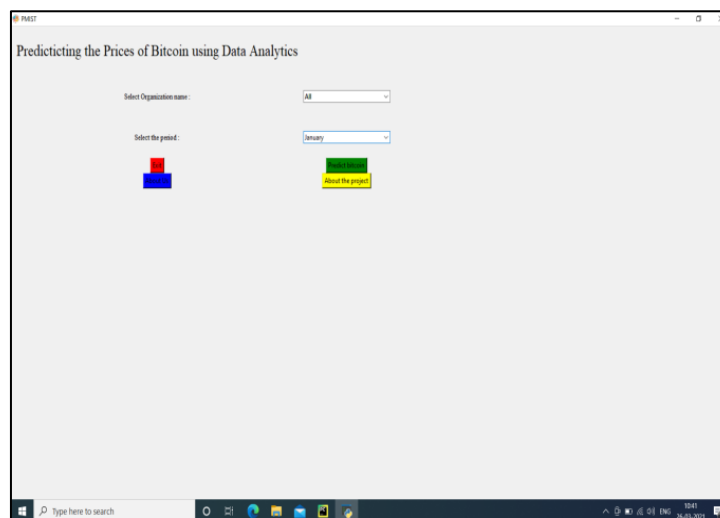
Fig. 18 January Month is given in the GUI to predict the prices of All Companies
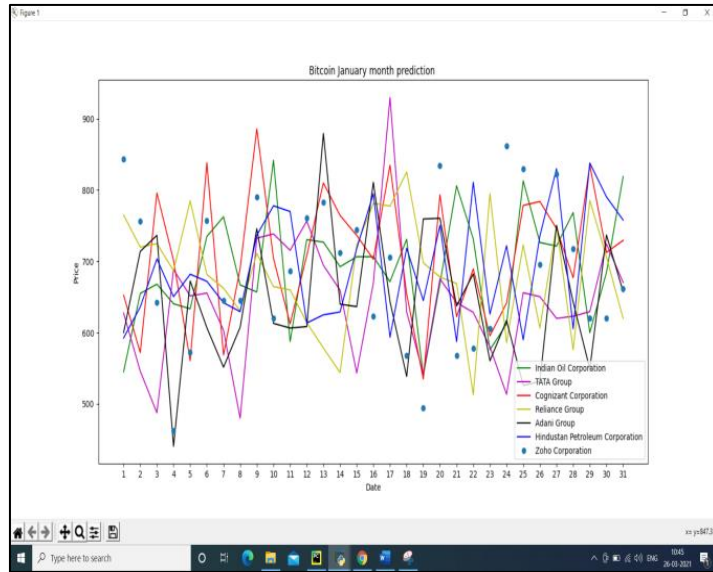
Fig. 19 Prediction results of Bitcoin prices for All Companies on January

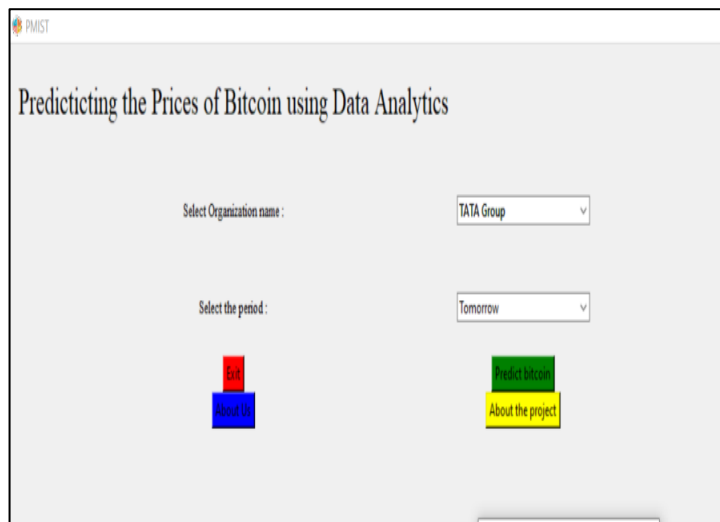**Fig. 20 System suggestion for investor betterment (All Companies - January)**

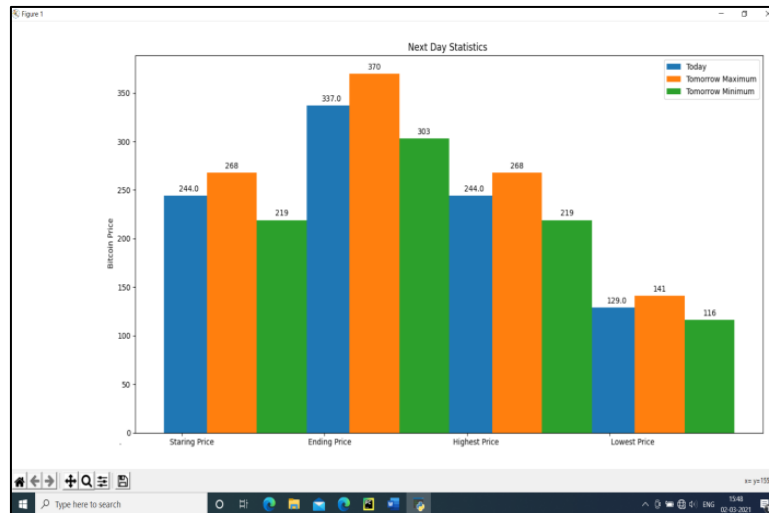**Fig. 21 Tomorrow is selected in the GUI to predict the prices of TATA Group**

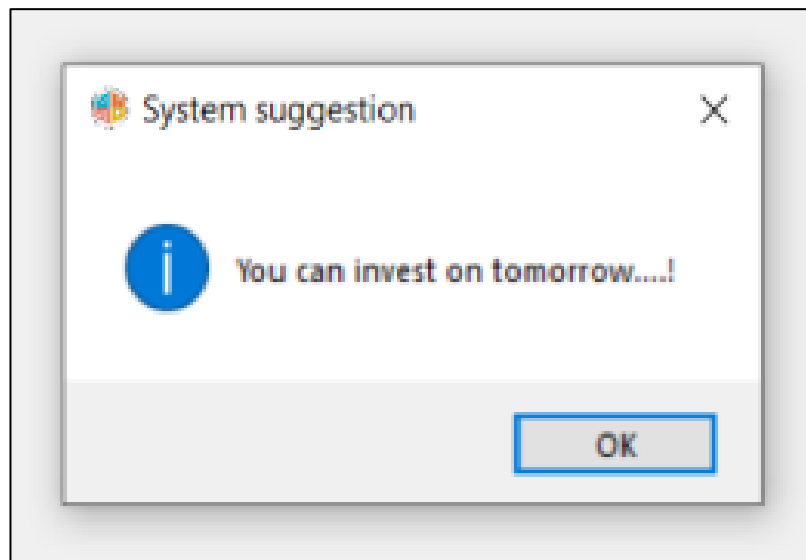**Fig. 22 Prediction results of Bitcoin prices for TATA Group on Tomorrow**

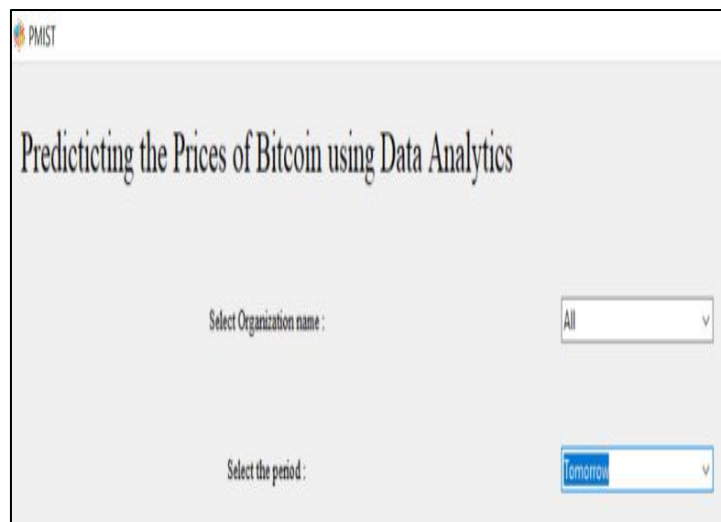**Fig. 23 System suggestion for investor betterment (TATA Group – Tomorrow)**

Fig. 24 Tomorrow is selected in the GUI to predict the prices of TATA Group
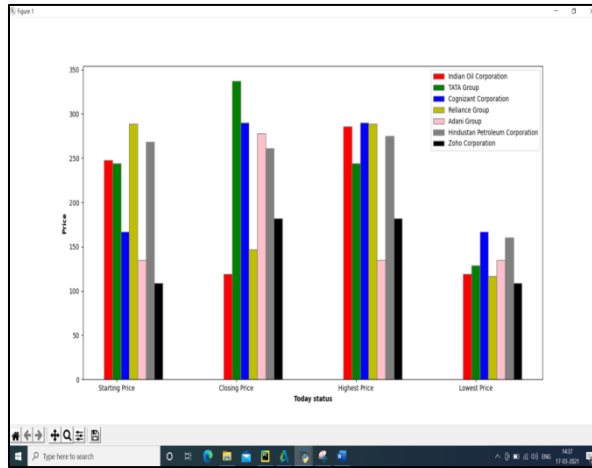
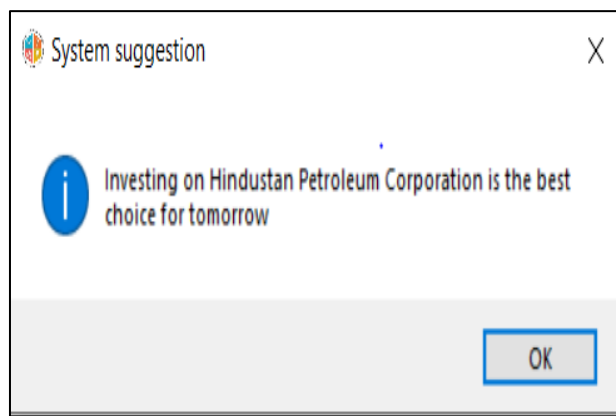**Fig. 25 Prediction results of Bitcoin prices for All Companies on Tomorrow**

Fig. 26 System suggestion for investor betterment (All Companies – Tomorrow)

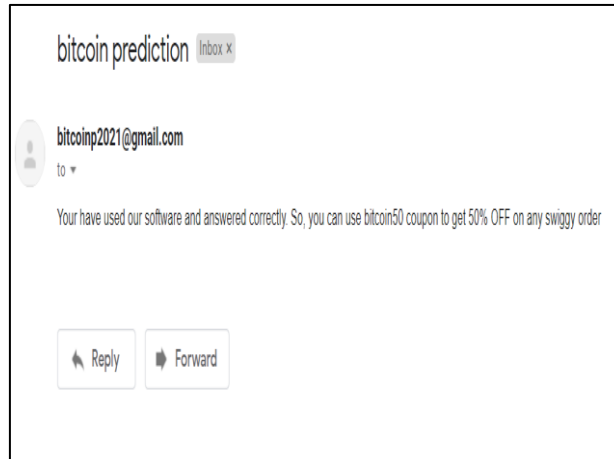**Fig. 27 A Bitcoin quiz with the mail id is asked for rewards**

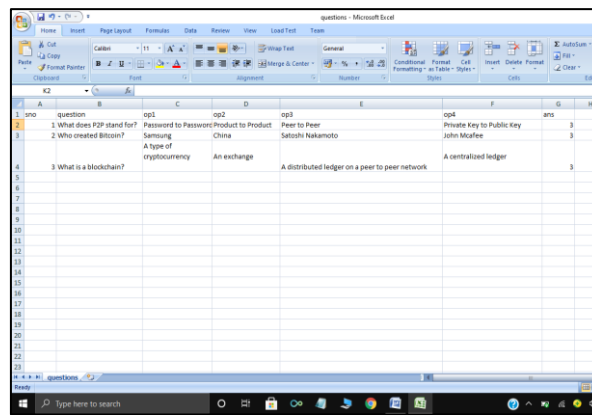**Fig. 28 Swiggy coupon number is mailed for correct answers**

Fig. 29 Quiz questions with answers prepared for the Rewarding System
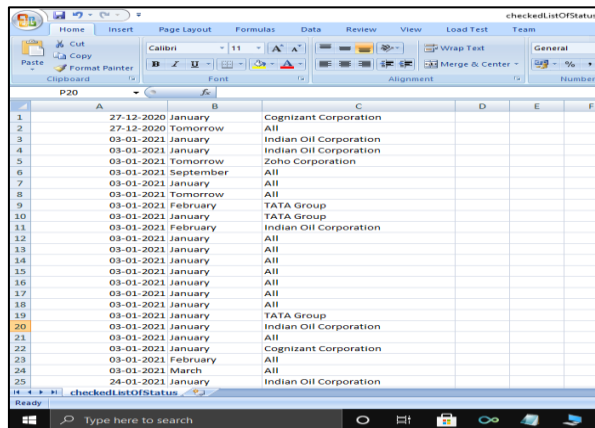

Fig. 30 Search history of Bitcoin future prices by the investors are gets stored in a separate CSV file named 'checkedListOfStatus'

8. CONCLUSION & FUTURE WORKS

The insight of our prediction system contains the numeric input data which allows us to predict the prices of the Bitcoin values periodically at real time with improved and accurate results. Also to attract more investors, a GUI entitled 'PMIST' is created. So, they can visualize the future Bitcoin prices by clicking next day or any specific month to know the prices at that month. The biggest advantage of our project is investor can view the multivariate time series datasets parallely with respect to the firm name (i.e. all prices are displayed within a single click rather than searching the highest, lowest, starting and ending prices separately). This makes the future investors to know at which company the price is high and thus apart from investing at the right time they can also

invest in a right firm. In our study we have included one of the cryptocurrencies namely Bitcoin alone for predicting its price fluctuations. In addition to this, predictions can also be made to the cryptocurrencies like ripple and etherum in the same way and compare with the other existing algorithms in the future.

ACKNOWLEDGMENT

REFERENCES

1. Lekkala Sreekanth Reddy, Dr. P. Sriramya," A Research On Bitcoin Price Prediction Using Machine Learning Algorithms" INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 9, ISSUE 04, APRIL 2020.
2. Ziaul Haque Munim, Mohammad Hassan Shakil, Ilan Alon, " Next-Day Bitcoin Price Forecast" Journal of Risk and Financial Management Received: 30 April 2019; Accepted: 14 June 2019; Published: 20 June 2019.
3. Mohammed Mudassir, Shada Bennbaia, Devrim Unal & Mohammad Hammoudeh, " Time-series forecasting of Bitcoin prices using high-dimensional features: a machine learning approach" S.I. : Data Fusion in the era of Data Science Published: 04 July 2020.
4. Nicola Uras, Lodovica Marchesi, Michele Marchesi, Roberto Tonelli, " Forecasting Bitcoin closing price series using linear regression and neural networks models" arXiv .org Submitted: 4 Jan 2020.
5. Ferdiansyah Ferdiansyah, Siti Hajar Othman, Raja Zahilah Raja Mohd Radzi, Deris Stiawan, " A LSTM-Method for Bitcoin Price Prediction: A Case Study Yahoo Finance Stock Market" 2019 International Conference on Electrical Engineering and Computer Science (ICECOS) , October 2019.
6. Mahboubeh Faghih Mohammadi Jalali & Hanif Heidari, " Predicting changes in Bitcoin price using grey system theory" Financial Innovation Published: 10 February 2020.
7. Yang Li, Zibin Zheng, Hong-Ning Dai, " Enhancing Bitcoin Price Fluctuation Prediction Using Attentive LSTM and Embedding Network" Applied sciences Received: 2 June 2020; Accepted: 8 July 2020; Published: 16 July 2020.
8. Gullapalli, Sneha, " Learning to predict cryptocurrency price using artificial neural network models of time series" k-State Research Exchange, 2018.
9. Suhwan Ji, Jongmin Kim, Hyeonseung Im, " A Comparative Study of Bitcoin Price Prediction Using Deep Learning" Mathematics Received: 12 July 2019; Accepted: 23 September 2019; Published: 25 September 2019.
10. Edwin Sin, Lipo Wang, " Bitcoin price prediction using ensembles of neural networks" 2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), July 2017.
11. Sean McNally, Jason Roche, Simon Caton, " Predicting the Price of Bitcoin Using Machine Learning" 2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP), Cambridge, 2018, pp. 339-343, doi: 10.1109/PDP2018.2018.00060.
12. Azim Muhammad Fahmi, Noor Azah Samsudin, Aida Mustapha, Nazim Razali, Shamsul Kamal Ahmad Khalid, " Regression based Analysis for Bitcoin Price Prediction" December 2018 International Journal of Engineering & Technology.
13. [13] N.I. Indera, I.M. Yassin, A. Zabidi, Z.I. Rizman, " Non-linear Autoregressive with Exogeneous input (narx) bitcoin price prediction model using PSO-optimized parameters and moving average technical indicators" Journal of Fundamental and Applied Sciences, 10 September 2017.
14. [14] H. Jang and J. Lee, "An Empirical Study on Modeling and Prediction of Bitcoin Prices With Bayesian Neural Networks Based on Blockchain Information," in IEEE Access, vol. 6, pp. 5427-5437, 2018, doi: 10.1109/ACCESS.2017.2779181.
15. Thearasak Phaladisailoed, Thanisa Numnonda, " Machine Learning Models Comparison for Bitcoin Price Prediction" 2018 10th International Conference on Information Technology and Electrical Engineering (ICITEE), July 2018.
16. Abraham, Jethin; Higdon, Daniel; Nelson, John; and Ibarra, Juan (2018) "Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis," SMU Data Science Review: Vol. 1 : No. 3 , Article 1.
17. Xiao Li and Weili Wu, " A Blockchain Transaction Graph based Machine Learning Method for Bitcoin Price Prediction" arXiv:2008.09667v1 [q-fin.ST] 21 Aug 2020.
18. Aniruddha Dutta, Saket Kumar, Meheli Basu, " A Gated Recurrent Unit Approach to Bitcoin Price Prediction " arXiv:1912.11166, 24 Dec 2019.

19. Li, Yuze and Jiang, Shangrong, "Hybrid Data Decomposition-Based Deep Learning for Bitcoin Prediction and Algorithm Trading "(January 22, 2020).
20. K. Rathan, S. V. Sai and T. S. Manikanta, "Crypto-Currency price prediction using Decision Tree and Regression techniques," 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2019, pp. 190-194, doi: 10.1109/ICOEI.2019.8862585.