

Bitcoin Price Prediction using Machine Learning

Sakib M¹, Usama S^{2*}, Abhay PS³

¹Department of Computer Science, Aligarh Muslim University Aligarh India

²Kalasalingam Academy of Research and Education, Krishnankoil-626126, Tamil Nadu, India

³Gurukul Kangri Vishwavidyalaya Computer Science Department, Haridwar, India

***Corresponding author:** Usama M, Kalasalingam Academy of Research and Education, Krishnankoil-626126, Tamil Nadu, India, Tel: 9897792486, Email: mohdusama@hust.edu.cn

Received Date: March 30, 2022 **Accepted Date:** May 14, 2022 **Published Date:** May 16, 2022

Citation: Sakib M, Usama S, Abhay PS (2022) Bitcoin Price Prediction using Machine Learning. J Comput Sci Software Dev 2: 1-10

Abstract

Nowadays, Bitcoin has become the most valuable cryptocurrency in the ecosystem of digital marketing. It is a type of money that is virtual, and there is no existence of this money. Using bitcoin, a person can make an anonymous transaction over the internet. Cryptocurrency is a computerized program that allows users to directly trade money value with one another. On the other hand, bitcoin values have been extremely volatile and making them pretty difficult to forecast. It has dropped nearly 10% of its value at the start of 2022 and pulling the wind from the crypto sector. And hence, the goal of this study is to find the most precise and effective model to predict bitcoin values using multiple machine learning models. In this paper, we investigate the price of bitcoin accurately by analyzing various parameters which affect its value. Data is taken from yahoo finance, which is a part of the yahoo network and provides financial reports about the stock market. We will estimate the sign of the daily price change with the greatest accuracy by using the available information.

Keywords: Bitcoin, Machine Learning, Linear Regression, Deep Learning, Forecasting, Prediction

Introduction

Bitcoin is a kind of electronic money that runs without the intervention of any central authority or government monitoring. There is no middle man involved in the transaction of bitcoin. It can only be used for digital payment or merely for investment purposes. Since it is decentralized means, no one owns it (Figure 1). Bitcoin was created in 2008 to address the inherent flaws in the trust-based transaction model, and it was originally defined as a solely peer-to-peer electronic payment system [1].

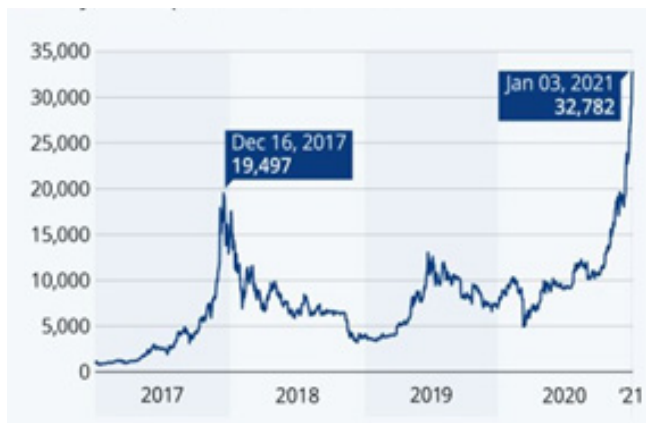


Figure 1: Variation in price of bitcoin at digital marketing

It is still leading the market of cryptocurrencies in terms of market value, user choice, and popularity among users. According to Statist, in 2020, Bitcoin's market capitalization was 66 percent of the entire market value of all cryptocurrencies. However, the percentage of bitcoin is decreasing due to the rising of other cryptocurrencies. In early 2021, the total number of bitcoins exchanged in a single day touched its highest value because of the people taking interest in bitcoin. Around 400,000 transactions were made in a single month, January. As per economists' standards, investors do not regard Bitcoin as a currency; Instead, they see it as fictitious financing equivalent to the web shares of the last century [2]. Investment can be performed on a variety of "Bitcoin exchanges," which are online marketplaces. Bitcoin is not bounded to any currency, and hence the user can buy and sell by using different currencies. Mt Gox is the largest Bitcoin exchange. A digital wallet is maintained for all the users of bitcoin, which is almost the same as a virtual bank. The place where all the timestamp data and transactions are stored is called "Blockchain" [3]. The information stored in the blockchain is encrypted. At the time of the transaction, only the wallet id of a user is made public, but the user's name is hidden [Figure 2].



Figure 2: Usage of cryptocurrencies at digital market

In response to the two most important questions, heated debates have been arisen: Why is bitcoin so valuable? What factors influence bitcoin's price. Bitcoin's value represents investor trust in cryptocurrency in the field of financial innovation [4]. As a result, most previous research has focused on the factors that influence or determine Bitcoin's price. Since bitcoin was first traded, investors have been troubled by price changes due to its intrinsic volatility. It is also necessary to be able to forecast Bitcoin price fluctuations.

Share market price forecasting has evolved over a decades by utilizing readily available high frequency data [5]. However, study on how to forecast its price is still lacking. When applying machine learning algorithms to predict the price of bitcoins, a common question arises, what characteristics should be considered? However, there are other approaches for selecting features [6,7]. Previous studies have relied on the domain knowledge of researcher and have not taken feature dimensions into account all [8,9]. We have tried to resolve this issue by combining the conclusion of previous empirical studies by experts who have a good understanding of the parameters that affect the price of bitcoin. The manuscript is organized as follows: In section one, we have clearly explained about bitcoin followed by all the related work in section two. Section three is about to implementation of this work which includes description of dataset. In section four we have discussed the results of our study. In the last section, its conclusion and some future directions have discussed.

Related Work

When it comes to analyzing the factors of bitcoin price formations, empirical works play significant role. The very first study on the bitcoin predictions was L. Kristoufek [10] work who considered the market of bitcoin to be completely speculative without any fundamentalist, and with the help of Wikipedia and Google trends, they examined the relationship among bitcoin and search queries. Later, for generating the latent source models, Shah et al [11] discussed the Bayesian regression procedure. In this work, they create a basic bitcoin trading strategy based on the prediction algorithms. The author in [12] analyzed all the work that has been done to forecast the US stock market and concluded that the MSE of the forecasting system was as great as the SD of the excess return. The author, on the other hand, shows that a number of basic financial and economic characteristics can anticipate the market excess return. Many researchers are using the open-high-low-close data from Okcoin. Madan et al [13] uses this dataset in minutes time period and separated the data into series of 30, 60 and 120 minutes. On daily bitcoin data, they performed three different binomial classification algorithms such as SVM, Random forest and logistic regression. They predict the price of bitcoin with around 97% accuracy for the next 10 minutes. However, this study does not include a cross-validation step, which could lead to overfitting of the models.

A very big challenge of such studies is the false information circulated through various social media platform like Facebook, twitter, whatsapp...etc, which inflates/deflates prices of cryptocurrency artificially [14]. Liquidity on Bitcoin exchanges is quite limited; as a result, the market is more vulnerable to manipulation. And hence, social media sentiments are not taken into account any longer. In of the study, Roche et al [15] im-

plemented deep learning model such as RNN and LSTM and to compare with these model they have implemented ARIMA model which is well known method for forecasting the time series data. They showed that long term dependencies are better recognizing by long shot term network (LSTM), but it takes longer training time.

Forecasting the price of bitcoin can be compared to other economical time series forecasting task such stock market prediction and foreign exchange marketex. The concept of employing ANN for such task isn't new. For stock price prediction, the MultiLayer Perceptron has been used in several studies [16,17]. In this study author utilized trail-error network parameter search process. Steinkrau et al [18] implemented ANN model on a GPU instead of CPU and reported that an ANN can perform three times faster in training and testing the model.

Implementation

Data Preparation

We have trained our machine learning model by using yahoo finance dataset which is usually trained with the help of the training data set. Everyday opens and closes are captured into a dataset with respect to date and time of the market and these will act as a test data set. We have 09 features in total out of which have 07 independent and 02 dependent variables i.e., day_perc_change & trend in training dataset and 09 independent variables in test dataset. The date, open, close, day_perc_change and trend are all categorical. The sample training data including mean, standard deviation, min and max value of dataset can be seen in Figure 3.

	Open	High	Low	Close	Adj Close	Volume
count	2705.000000	2705.000000	2705.000000	2705.000000	2705.000000	2.705000e+03
mean	11219.325346	11521.833572	10885.835851	11233.473891	11233.473891	1.468531e+10
std	16041.345649	16473.841710	15543.927110	16047.741249	16047.741249	2.004174e+10
min	176.897003	211.731003	171.509995	178.102997	178.102997	5.914570e+06
25%	606.242981	608.458984	602.044983	606.271973	606.271973	7.901180e+07
50%	6292.640137	6409.850098	6202.373535	6300.859863	6300.859863	5.064430e+09
75%	10393.900391	10663.919922	10182.464844	10399.668945	10399.668945	2.458267e+10
max	67549.734375	68789.625000	66382.062500	67566.828125	67566.828125	3.509679e+11

Figure 3: Small sample of the dataset with the description of all the variables used

The closes and avg price is calculated over a period of time, based on that day_perc_change & Trend were calculated and defined (Bear drop, Bull run, Positive, Slightly positive, Negative, Slightly negative). In table1, all the variable names of data-

set are clearly explained. The entire attribute has been defined for a particular time. It is clear that, not a single null value is present in our dataset Figure 4.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2704 entries, 0 to 2703
Data columns (total 7 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Date            2704 non-null   object
1   Open            2704 non-null   float64
2   High            2704 non-null   float64
3   Low             2704 non-null   float64
4   Close           2704 non-null   float64
5   Adj Close       2704 non-null   float64
6   Volume          2704 non-null   int64
dtypes: float64(5), int64(1), object(1)
memory usage: 148.0+ KB
```

Figure 4: showing that no value is null in the dataset

Exploratory Data Analysis

We have done analysis over the data by structuring; visualizing and classifying. The libraries used for data analysis are pandas, matplotlib and seaborn. For the sake of structuring the data we have used pandas and for visualizing we imported seaborn package and finally to represent in a graphical statistical manner we used matplotlib.

Open/close mostly refers to the value of a cryptocurrencies at a time interval, i.e., start of the day refers to the cryptocurrency opens, whereas in parallel the price at which a cryptocurrency closes at a time period, i.e., at the end of the day. The following are the graphs of visualized data Figure 5 (a) & Figure.5 (b)



Figure 5 (a): Opening value of the bitcoin market



Figure 5 (b): Closing value of the bitcoin market

Since the May 2020 halving, the price of bitcoin has increased by 300 percent which shows that it is still considered

as a highly volatile asset having daily price swings of 5 to 10% (Figure 6).

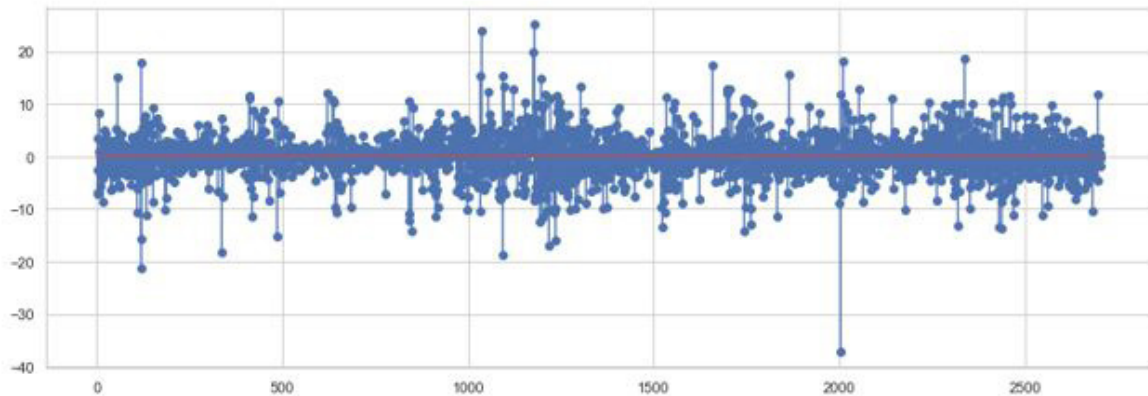


Figure 6: Day per change at digital market of bitcoin

The closes and avg price is calculated over a period of time, based on that day_perc_change was defined. In the world of market investing, the terms “Bull” and “Bear” are frequent terms that were used to refer to various conditions of the market. This represents the general state of the stock markets, i.e., whether they are gaining or declining in value. As an investor, it’s critical to understand how current market conditions affect your investment portfolio.

The trends of the stock are observed based on the daily returns i.e. daily percentage change in closing price of the crypto. The trends are analyzed based on the following conditions. Figure 7(a), Figure 7 (b)

1. If the daily returns are between -0.5 and 0.5 that means very slight change or no change

2. If the daily returns are between 0.5 and 1 that means slight change on the positive side

3. If the daily returns are between -1 and -0.5 that means slight change on the negative side

4. If the daily returns are between 1 and 3 that means change on the positive side

5. If the daily returns are between -3 and -1 that means change on the negative side

6. If the daily returns are between 3 and 7 that means top gains

7. If the daily returns are between -3 and -7 that means top losses

8. If the daily returns are greater than 7 that means bull run (crypto prices are on rise)

9. If the daily returns are lesser than -7 that means bear drop (crypto prices are on decline)

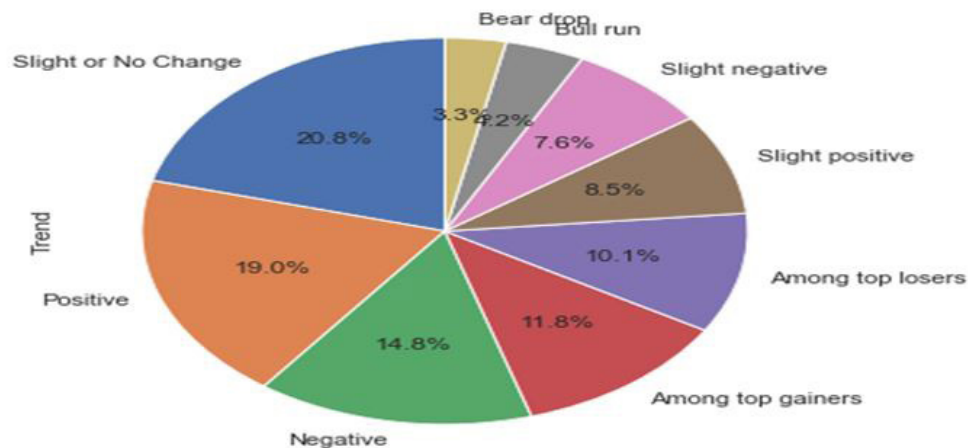


Figure 7 (a): Represents the trend of bitcoin market

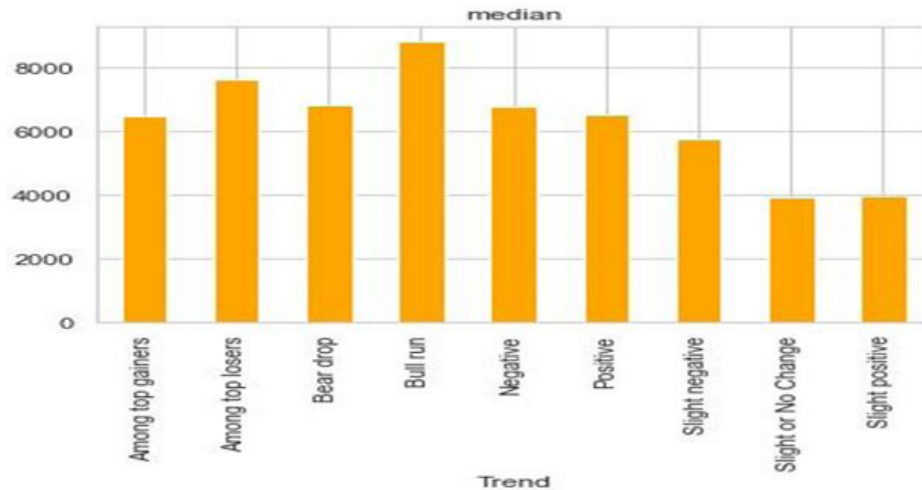


Figure 7 (b): Median for the trend of the bitcoin market

In most of the cases investors withdraw their share from the bear market till the trend to take up reverse, further sending prices lower (Figure 8)

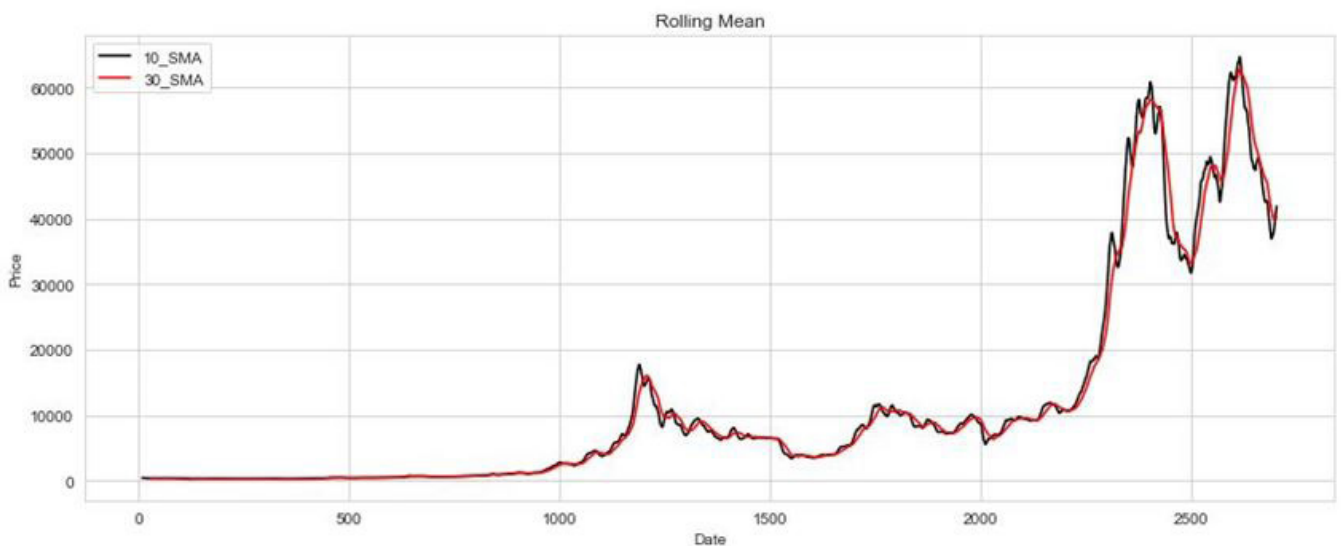


Figure 8: Represents the rolling mean of 10 and 30 days respectively

The rolling mean is defined based on price and number of days respectively. In following figure, the blue curve shows cases the rolling mean of 10 days whereas, green curve showcases the rolling mean of 30 days sequentially.

The upper circuit limit is the highest price a stock can reach on a given day, while the lower circuit limit is the lowest. Similarly, the lower circuit limit is the lowest price that a stock can reach, and when a stock reaches this limit, there will be just sellers and no buyers.

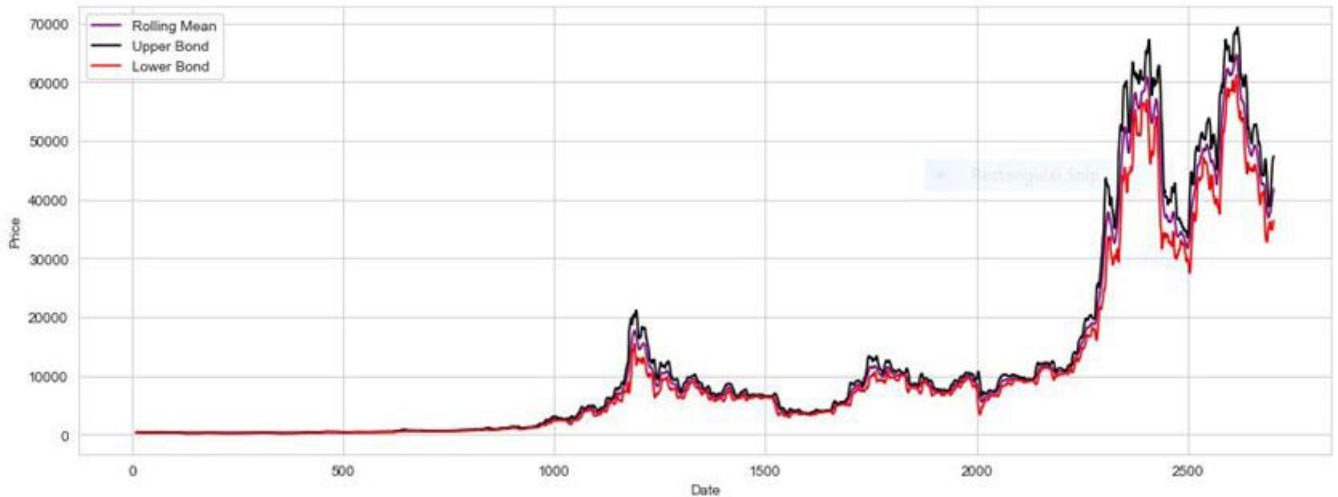
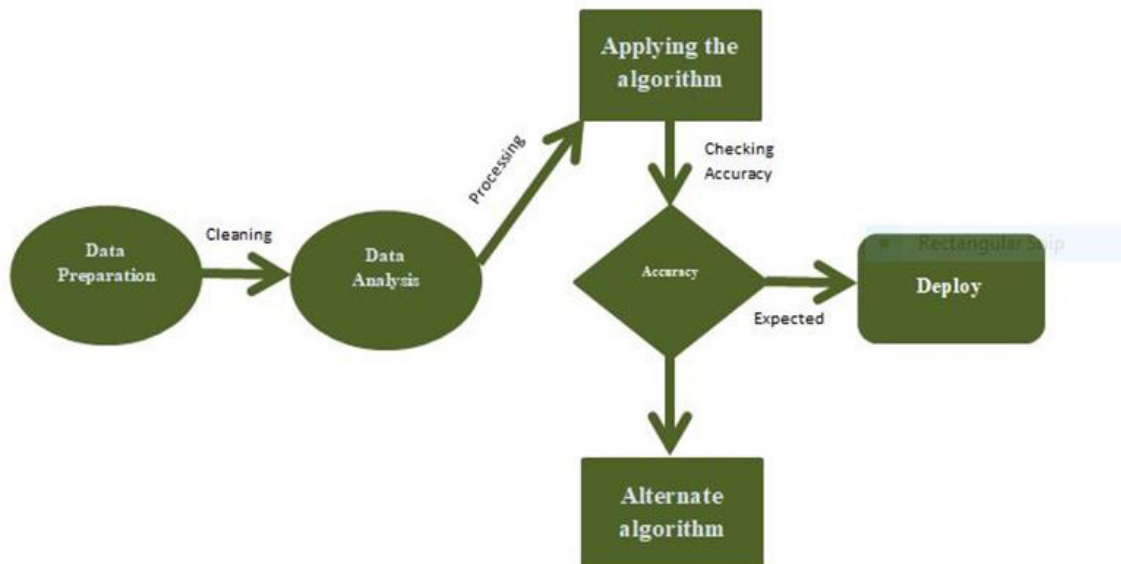


Figure 9: Represents the upper and lower circuit of crypto bitcoin

Architecture



Architecture of the procedure of machine learning algorithm

$$Y = a + bX \quad \text{Eq1}$$

Every model in machine learning must follow an architecture in which main steps remains same but the dataset and choosing methods can be different. All these procedures must be followed for each dataset used to create ML model. The architecture of linear regression which based on supervised machine learning is given above.

In regression analysis we investigate the relationship between a dependent and independent variable. It shows the changes on a dependent variable on the y-axis to the changes in the explanatory variable on the x-axis, by drawing it with the available dataset.

Eq1 is a linear regression where X is an explanatory variable and Y is a dependent variable and b is the slope of the line, a is the intercept. The primary aim is to find the best suitable values for a and b that will result in the best-fitting line for a given dataset values.

Results and Discussion

Below graphs represent the prediction of linear regression model before training and after training the model.

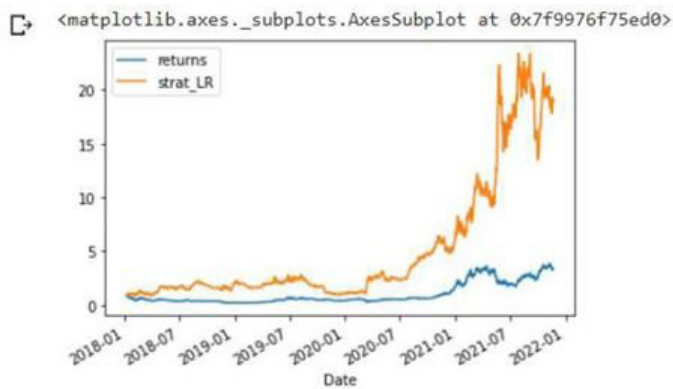


Figure 9(a): Before training the model

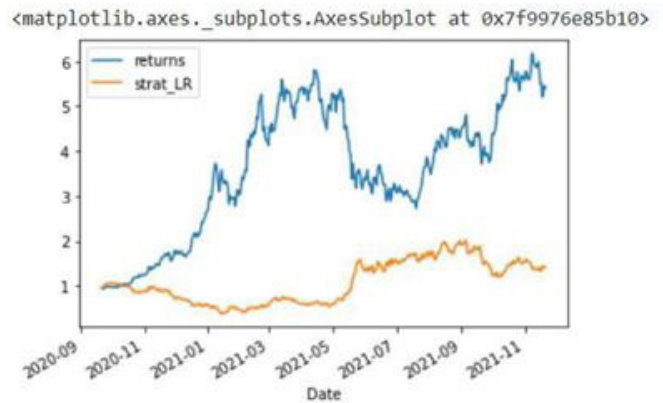


Figure 9(b): After training the model

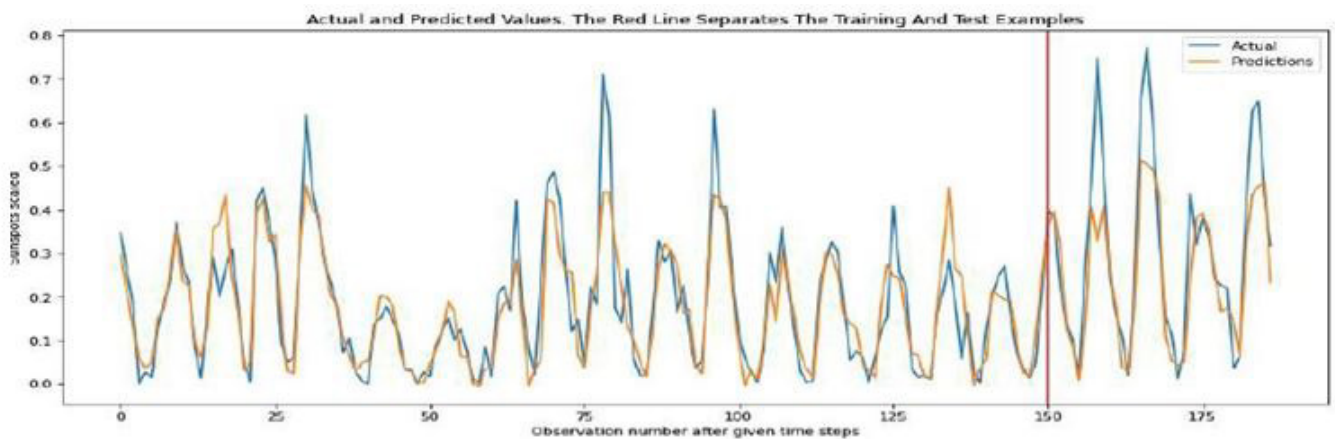
Bitcoin price prediction and its analysis is one of the most promising tasks to complete. A number of factors are affecting its price. Market volatility is also an important factor to consider while predicting its future value. Various dependent and independent parameters are influencing its price. And hence the prediction of rise and fall of Bitcoin price is extremely difficult.

Although, machine learning algorithms make it a little easy to forecast its future price. Since the time machine learning introduced, market prediction advancements have begun to in-

clude such approached in analyzing the data. Machine learning not only saves time and resources but also outperforms people in terms performance.

Without training, the model cannot predict accurate future value of any dataset. But after training the model, we get more accurate result than previous one.

The outcome suggests that utilizing the liner regression technique to predict bitcoin price is very effective. In our study, we have trained 80% data and kept 20% for testing purpose. We got nearly 90% accuracy and beat the existing work.



There is no involvement of government in issuing and or in regulating the bitcoin and hence no government policy can be applied on bitcoin.

But it depends on the supply and the demand of the market, its availability, other cryptocurrencies and the most important sentiments of the investors.

Conclusion and Future work

In this work, we examined machine learning methods linear regression to predict bitcoin future price. We have extracted feature variables of bitcoin like open, close, high, low, etc. These parameters can be highly effective for prediction of bitcoin's price and it should be considered. The majority of our

result outperform than most of the other machine learning algorithms. Although, there are various factor which can affect the price of bitcoin. Social media rumors, government new policies, new laws...etc. can play a big role in the fall and rise of the bitcoin's price. Our study has numerous limitations in terms of sources of data and methods, which indicate future research directions. Consequently, for the best results from all the methods, dataset must always be updated. We plan to look at new methodologies, such as the ARIMA and statistical method the deep learning model RNN, to improve our research.

References

1. Wright CS, (2019) Bitcoin: A Peer-to-Peer Electronic Cash System [Bitcoin: un sistema de efectivo electrónico de igual a igual], J SSRN Electron 1-9.
2. Yermack D (2013) "Is Bitcoin a Real Currency?" J SSRN Electron.
3. Gandal N and Halaburda H (2016) "Can we predict the winner in a market with network effects? Competition in cryptocurrency market," Games 7: 1-21.
4. Mai F, Shan Z, Bai Q, Wang XS and Chiang RHL (2018) "How Does Social Media Impact Bitcoin Value? A Test of the Silent Majority Hypothesis," J Manag Inf Syst 35: 19-52.
5. Demir A, Akilotu BN, Kadiroglu Z, and Sengur A (2019) "Bitcoin Price Prediction Using Machine Learning Methods," 1st Int. Informatics Softw. Eng. Conf. Innov. Technol. Digit. Transform. IISEC 144-147.
6. Qasim Abualigah LM and Hanandeh ES (2015) "Applying Genetic Algorithms to Information Retrieval Using Vector Space Model," Int. J Comput Sci Eng Appl 5 19-28.
7. Abualigah LM, Khader AT and Hanandeh ES (2018) A combination of objective functions and hybrid Krill herd algorithm for text document clustering analysis," Eng Appl Artif Intell 73: 111-125
8. Abualigah LM and Khader AT (2017) Unsupervised text feature selection technique based on hybrid particle swarm optimization algorithm with genetic operators for the text clustering. J Supercomput 73: 4773-4795.
9. Abualigah LM, Khader AT and Hanandeh ES (2018) A new feature selection method to improve the document clustering using particle swarm optimization algorithm J Comput Sci 25: 456-466.
10. Kristoufek L (2013) BitCoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era. Sci Rep 3: 1-7.

11. Shah D and Zhang K (2014) Bayesian regression and Bitcoin 2014 52nd Annu. Allert. Conf. Commun. Control. Comput. Allert. 2: 409-414.
12. Nemes MD and Butoi A (2013) Data Mining on Romanian Stock Market Using Neural Networks for Price Prediction. Inform Econ 17: 125-136.
13. Madan I, Saluja S, and Zhao A (2015) Automated Bitcoin Trading via Machine Learning Algorithms 20.
14. Rechenthin M, Street WN and Srinivasan P (2013) Stock chatter: Using stock sentiment to predict price direction,” Algorithmic Financ. 2: 169-196.
15. Roche J and McNally S (2016) Predicting the price of Bitcoin using Machine Learning Sean McNally Supervisor.
16. White H (1988) Economic prediction using neural networks: The case of IBM daily stock returns 451-458
17. Yoo PD, Kim MH, Jan T (2005) Machine learning techniques and use of event information for stock market prediction: A survey and evaluation,” Proc. - Int. Conf. Comput. Intell. Model. Control Autom. CIMCA 2005 Int. Conf. Intell. Agents, Web Technol. Internet 2: 835-841
18. Steinkrau D, Simard PY, and Buck I (2005) Using GPUs for machine learning algorithms,” in Eighth International Conference on Document Analysis and Recognition (ICDAR’05) 1115-1119.

Submit your manuscript to a JScholar journal and benefit from:

- ✦ Convenient online submission
- ✦ Rigorous peer review
- ✦ Immediate publication on acceptance
- ✦ Open access: articles freely available online
- ✦ High visibility within the field
- ✦ Better discount for your subsequent articles

Submit your manuscript at
<http://www.jscholaronline.org/submit-manuscript.php>