Journal of Informatics
Electrical and Electronics Engineering, 2022,
Vol. 03, Iss. 01, S. No. 004, pp. 1-9
ISSN (Online): 2582-7006

Bitcoin Price Prediction Using Machine Learning Techniques

S. K. Ibrahim[1], Pawan Singh[2]

Amity School of Engineering and Technology, Amity University Uttar Pradesh, Lucknow Campus, India[1,2]
imranshaik1819@gmail.com[1], pawansingh51279@gmail.com[2]

How to cite this paper: S. K. Ibrahim, P. Singh, "Bitcoin Price Prediction Using Machine Learning Techniques," *Journal of Informatics Electrical and Electronics Engineering* (*JIEEE*), Vol. 03, Iss. 01, S No. 004, pp. 1–9, 2022.

http://doi.org/10.54060/JIEEE/003.01.004

Received: 05/04/2022
Accepted: 24/04/2022
Published: 25/04/2022

Abstract

This paper discusses, trying to accurately assess the price of Bitcoin by looking at different parameters affects the value of Bitcoin. In our work, we focus on understanding and seeing the evolution of Bitcoin daily market, a1 and gaining intuition in the most relevant aspects surrounding the Bitcoin price. In the meantime, market capitalization of publicly traded cryptocurrencies exceeds $ 230 billion. The most important cryptocurrency, Bitcoin, is used primarily as a digital value store, and its pricing opportunities have been extensively considered. These features are described in more detail in the following paragraph: details of the main Bitcoin, as described in the paper. Bitcoin is the most expensive digital currency in the market. However, Bitcoin prices have been highly volatile, making it difficult to forecast. As a result, the goal of this research is to find the most efficient and accurate model for predicting Bitcoin prices using various machine learning algorithms. Several regression models with scikit-learn and Keras libraries were tested using 1-minute interval trading data from the Bitcoin exchange website bit stamp from January 1. 2012 to January 8, 2018. The best results showed a Mean Squared Error (MSE) as low as 0.00002 and an R- Square (R2) as high as 99.2 percent.

Keywords

Bitcoin, Prediction, Time-Series, Predictive Model, Deep Learning, Long-Short Term Memory (LSTM), Recurrent Neural Network (RNN), Real-time Data, Mean Absolute Error (MAE), Accuracy.

1. Introduction

1.1 What is Bitcoin?

Satoshi Nakamoto first introduced Bitcoin in October 2008 in his white paper titled "Bitcoin: Peer-to-Peer Electronic Cash System" [1]. Bitcoin is the first decentralized cryptocurrency, and other digital currencies (also known as Altcoins or alterna-

tive virtual currencies) are created by cloning or modifying the Bitcoin mechanism [2]. All transactions governed by cryptography are secure, validated, and recorded in a "blockchain" by a decentralized network [3]. Online payment transactions can be conducted directly between any two willing parties using the concept based on the new electronic cash system, we need to eliminate the trusted third party such as a financial institution. In March 2017, Bitcoin was the largest and most popular cryptocurrency market by market capitalization. In January–February 2017, Bitcoin accounts accounted for 72 percent of all cryptocurrency accounts in the market, and the total number of transactions was 286,419, which was more than all other cryptocurrencies combined [2].

Bitcoin works without central government or banks using peer-to-peer technology. Bitcoin is open source; its plan is public, neither owner nor manage Bitcoin, and anyone can participate. Satoshi Nakamoto, an unknown creator, was introduced digital money as open-source software in 2009 digital money because it controls the creation once 16 money transfers using cryptography. Users post payments on the network through digital streaming 5 1 signed messages. by digitally signed distribution miners are reassuring donors and markers by date transaction on blockchain, shared social network. to earn transaction funds once just made bitcoins.

1.2 Prediction

The value of Bitcoin fluctuates in the same way that the value of any other stock does. Many algorithms are used to forecast stock market prices using stock market data. However, the elements that determine Bitcoin are not the same. As a result, forecasting the value of Bitcoin is necessary to make sound investment decisions. Unlike the stock market, the Bitcoin price is not affected by Business events or government intervention. As a result, we believe it's necessary to use machine learning [4][5] technology to forecast the Bitcoin price to forecast its value.

2. Related Work

This study analyses data using two libraries, scikit- learn and Keras, to create machine learning models. The TensorFlow library is also used in this study to generate data flow graphs.

2.1. Scikit-learn

Scikit-learn is an open-source data mining analysis package. Python is used to analyze and build models based on algorithms of machine learning such as classification, regression, and clustering. Scikit-learn can also be used to prepare data in a variety of ways, including normalization, standardization, and cleaning outlier or missing data [6].

2.2. TensorFlow

TensorFlow is an open-source deep learning framework developed by Google. It is used to train Neural Network (NN) models and predict results by collaborating with many Graphical Processing Unit (GPU), allowing powerful deep learning and NN algorithms to be implemented. This framework can also be used in other fields such as speech recognition, computer vision, robotics, and so on. When graphs are made up of node groups, TensorFlow can generate data flow graphs for processing. Figure 1 depicts an example of data flow processing [7].

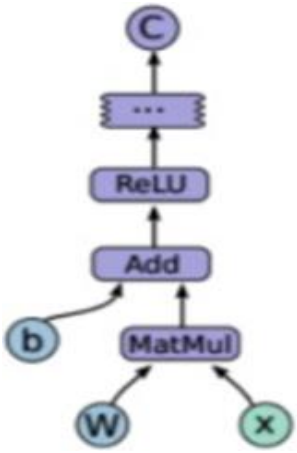

Figure 1. An example of data flow processing [7]

2.3 Keras

Keras is an open-source package for high-level NN. It provides Python API for NN programming. It can also be used with the TensorFlow, CNTK, and Theano libraries [6]. Keras is used to create machine learning, neural network, and deep learning models. Keras is easier to build and understand when codes are divided into parts. The components of generating models typically include neural layers, cost functions, an optimizer, and activation functions. Python is used to create new defined functions or classes.

3. Machine Learning

The study of how a computer programmer can learn and adapt to new data without human intervention is known as machine learning by using statistical model algorithms to analyze and draw conclusions from data patterns. Machine learning algorithms-based systems have the potential to learn from experience or historical data. Machine learning can be divided into three categories:

3.1 Supervised learning

The algorithm in supervised learning deals with the labelled dataset and provides a solution key that the algorithm can utilize to calculate its accuracy on training data. [7] in this machine is trained on a dataset that contains both input and output. So, once the training is completed and the machine has attained a certain level of learning, the machine is finally deployed. Supervised learning is further subdivided into two categories: and regression.

Supervised learning examples: For example, suppose you have a 2-year-old niece who is learning to speak. You want to teach her the distinction between a dog and a cat. So, what are you going to do? You will show her videos of dogs and cats, or you will bring a dog and a cat to show her in person so she can understand how they differ. Both cats and dogs have 4 legs and a tail.

- Dogs have long mouths while cats have smaller mouths.
- Dogs come in small to large sizes. But Cats, are always small.
- cat's meow while Dogs bark.

This is how u train the system first by giving some data. so that the computer will differentiate between the dogs and cats.
Applications of Supervised learning: Cortana, Siri, and Alexa are voice assistants that trained using our voice.

3.2. Unsupervised learning

In unsupervised learning, this is the technique in which there is no need to teach the model or supervise the model. The model itself works on its own to discover the insights and patterns that were undetected before. It deals with unlabeled data. Based on the observations the values are predicted in the future.

Examples of unsupervised learning:

- Apriori Algorithm
- K-means Algorithm
- Hierarchical Clustering.

Consider the example of a newborn baby whose family members have a dog. You see this dog. A few weeks later, one friend brings another dog closer and tries to play with the baby. Baby had never seen this dog before. However, you can see that many features (2 ears, eyes, movements in four legs) are like those of his favorite dog. He sees a new animal, like a dog. This is unattended reading, where you can be taught by anyone, but you can learn from the data (in this case the dog data.) This would have been guarded reading if a friend had told the child that he was a dog.

4. Data Set

The main database includes Bitcoin prices taken at intervals between October 16, 2015, and October 1, 2021.

	Date	Price	Open	High	Low	Vol.	Change %
0	Oct 16, 2021	61,635.6	61,669.2	62,329.2	61,123.2	96.15K	-0.06%
1	Oct 15, 2021	61,672.5	57,348.2	62,892.8	56,874.8	121.48K	7.54%
2	Oct 14, 2021	57,345.8	57,370.4	58,506.5	56,866.8	59.84K	-0.06%
3	Oct 13, 2021	57,380.1	56,014.5	57,697.4	54,314.4	73.62K	2.44%
4	Oct 12, 2021	56,015.9	57,480.6	57,635.7	54,155.4	74.85K	-2.54%
...	...	...	...	...	...	...	...
361	Oct 20, 2020	11,913.5	11,753.9	12,029.4	11,685.7	94.59K	1.36%
362	Oct 19, 2020	11,753.4	11,507.1	11,823.3	11,413.9	73.04K	2.14%
363	Oct 18, 2020	11,506.9	11,362.1	11,506.9	11,349.8	31.51K	1.27%
364	Oct 17, 2020	11,362.1	11,321.8	11,402.7	11,274.4	29.86K	0.35%
365	Oct 16, 2020	11,322.0	11,503.2	11,542.2	11,224.6	71.42K	-1.57%

366 rows × 7 columns

Figure 2: data set display

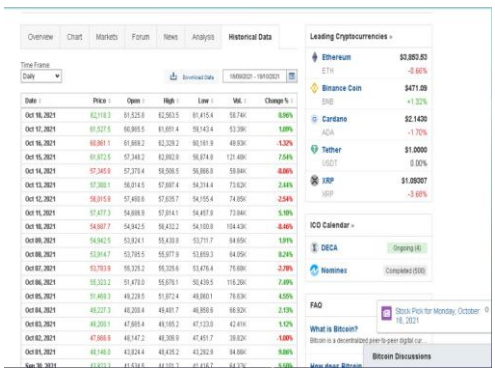

Figure 3: Data set from the website.

Dataset is taken from https://in.investing.com/crypto/bitcoin/historical-data

We have set the close price of Bitcoin as our target for our project. The dataset is divided into two parts: 80 percent for training and 20 percent for testing.

5. Methodology

There are too many models tested 2 for their ability to predict Bitcoin price volatility. Logistic Regression and Support Vector Machine are two separate models.

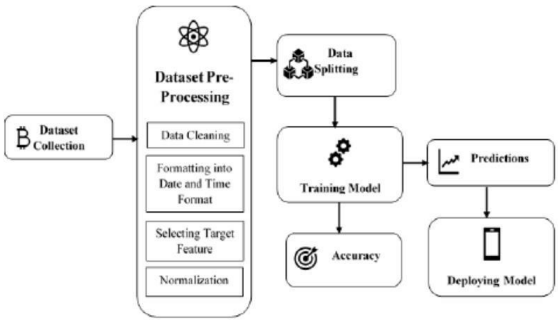

Figure 4: data pre-processing process

5.1. Data Cleaning

In this study, models of machine learning are trained using Bitcoin transaction data from the Bitstamp website and published on the Kaggle website. However, this study focuses on 1-minute interval trading exchange data rates in USD from January 1, 2012, to January 8, 2018. CSV files are used to store the datasets.

5.2. Feature Selection

Features of the datasets from A are as follows:

Features and definition:

Close: Latest trade
Open: Opening trade
High: highest trade during day
Low: lowest trade during day
Price: Bitcoin Price
Volume: total trade volume of day in BTC

When predicting the Weighted price, the scikit-learn library is used to create models with only the features Close, Open, High, and Low.

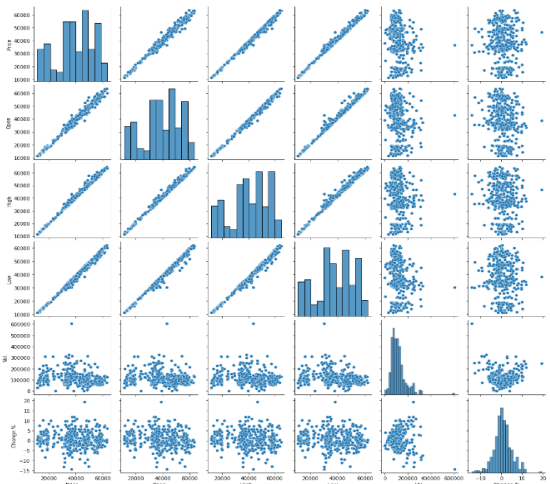

Figure 4: Correlation pair plot coefficient for each parameter of features

Figure 5: Heat map correlation values for each parameter of features.

5.3 Data Preparation

The Min-Max scaler is used to convert data from 0 to 1 by using the data's minimum and maximum values [12]. In 4.2, the MinMaxScaler in scikit-learn can use equation 1 to transform data of selected features.

$$x_{sc} = \frac{x - \min(x)}{\max(x) - \min(x)}$$

Increasing the number of dimensions from two to three by including a time dimension allows the data to use the LSTM and GRU modelling.

6. Data Visualization

Figure shows that the price increased rapidly from April 2019 to July 2019 and reached its highest peak, which is above 12000 USD. The curve then gradually declines, with a visible dip in the graph from January 2020 to April 2020. The period of lockdown has also had an impact on the Bitcoin price. The rise of price again in April 2020. Since April, the price has been steadily rising with minor fluctuations.

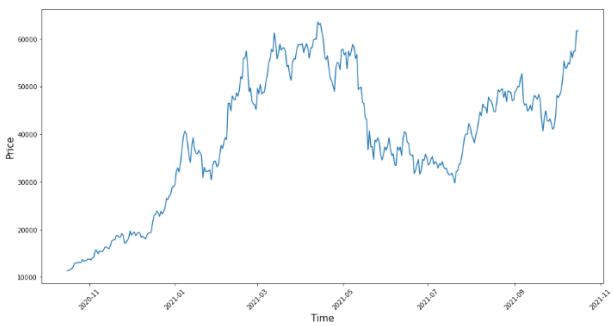

Figure 6: Data plotted and visualized

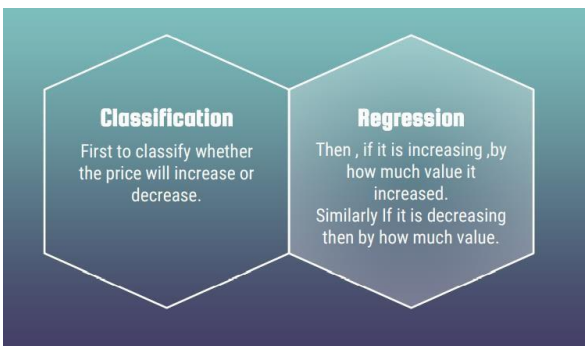

Figure 7: classification vs regression

Basic Approach:

We chose regression machine learning in this study due to the continuous values of the Bitcoin price. With scikit-learn library, logistic regression is used for prediction and for deep learning-based regression models, Keras library was used to create LSTM models.

6.1 Logistic Regression

It is a mathematical process of analyzing a database in which one or more variables influence the outcome. The result is measured using different variables (only two possible outcomes). Given a collection of independent variables, it predicts a binary result (1 / 0, Yes / No, True/ False).

It is a model of categorical predictive regression. It employs Estimation based on Maximum Likelihood to calculate the chances that Logistic Regression will be applied to a given class.

$$h_\theta(x) = g(\theta^T x) = \frac{1}{1 + e^{-\theta^T x}}$$

where x denotes the input and θ parameter to be learned.

6.2 Support Vector Machine

Like logistic regression, the support vector machine approach generates a binary classification model while making relatively few assumptions about the dataset. The classifier is generated by optimization.
where we have taken input as x and w,b are the parameters to be learned The value of wTx + b is used to make predictions.

$$\begin{aligned} \min_{\gamma,w,b} \quad & \frac{1}{2}||w||^2 \\ \text{s.t.} \quad & y^{(i)}(w^T x^{(i)} + b) \geq 1, \; i = 1, \ldots, m \end{aligned}$$

6.3 Normalization

The following step is to normalize the data. The process of rescaling original data without changing its behavior or nature is known as normalization. The Min-Max Normalization technique is used here.

- **Min - Max Normalization:**

  It applies linear transformation to actual data. It preserves the relationship between the values of real records. In the future, if the input values exceed the normalization range, an error known as an out-of- bound-error may occur. [8]

- **Long Short – Time Neural Network[LSTM]**

  The recurrent neural network is a neural network with a directed graph sequence that connects the output to the input. This input connects the results of the number point to the next point in time and tracks neuron-related information, resulting in the next event. The value is proportional to the value of earlier time points' production.

  Long-Short Term Memory (LSTM) is a type of Recurrent Neural Network that can learn data from long sequences, a challenge that conventional RNNs struggle with, and is useful for some types of prediction that allow the network to maintain knowledge over long periods of time. The shape of all repeating neural networks is represented

by a sequence of RNN modules. This repeating module would have a simple structure with regular RNN, like a tan layer. LSTMs have a similar shape to this chain, but the repetitive module has a different structure. Instead of a single layer, there are four that interact in a unique way.

- **Mean Absolute Error**

  Mean Absolute Error is a model evaluation metric that is commonly used with regression models. The total error of the model in relation to the test set is the definition of all the guessing error values in all cases in the test set. Each prediction error shows the difference between the actual events and the actual estimated prices. [9]

7. Results and discussion

The model is implemented using the LSTM algorithm.

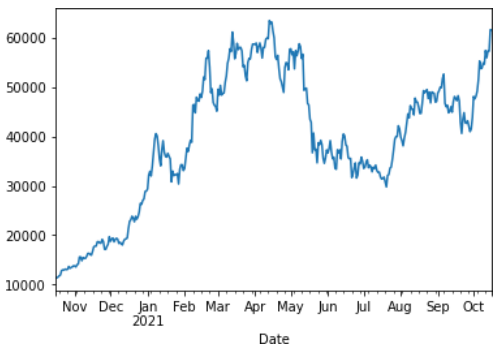

Figure 8: growth of future bitcoin price

Linear Regression results are finalized, as shown in Figure 1. Shows the outcome of using Linear Regression with the day as a feature to forecast the market price of Bitcoin over the last 365 days. Linear Regression has demonstrated that a simplified model of Bitcoin market price can be produced with minor complexity details, albeit with larger error due to the regression line's linear constraint. As the price of cryptocurrencies such as Bitcoin rises exponentially, the error of Linear Regression will worsen because of underfitting. Moving on to the second Linear Regression model, where we use Blockchain size as our main feature, we notice some striking similarities.

Code of the project: Github (2019) Bitcoin Price Prediction [online] available at: https://github.com/SkIbrahim18/Bitcoin-Price- Prediction

8. Conclusion

Bitcoin is a decentralized cryptocurrency that operates on a peer-to-peer basis. It employs cryptography for security and anonymity. Bitcoins, unlike traditional currencies, are entirely virtual. Bitcoin's value fluctuates like that stock prices do. Many algorithms on the stock market value are used for price forecasting. Bitcoin's valuation must also be anticipated to make the best investment decision. This paper uses a 3-Layer LSTM Model and real-time data to forecast the future price of Bitcoin for the next 100 days starting today. The model provides an excellent prediction with an error rate of less than 0.5. The idea was to include a framework that could analyze real-time data and provide investors with a sense of direction to help them make decisions. In this project, we attempted to predict Bitcoin prices using two deep learning methodologies. This work focuses on the advancement of project- based learning in the f of computer science engineering, considering

problem definition, progression, student assessment, and the use of hands-on activities based on the use of a deep learning algorithm to develop an application that can predict bitcoin price.

References

[1]. G. Neil, H.Halaburda. "Can we predict the winner in a market with network effects? Completion in cryptocurrency market," Games, vol.7 no.3, 2016, pp.16. vol.7 no.3, 2016, pp.16. cryptocurrency market," Games, vol.7 no.3, 2016, pp.16.

[2]. N. Mangla and P. Rathod, "Unstructured data analysis and processing using Big Data tool -Hive and machine learning algorithm -linear regression," *Iaeme.com*. [Online]. Available: https://iaeme.com/MasterAdmin/Journal_uploads/IJCET/VOLUME_9_ISSUE_2/IJCET_09_02_006.pdf. [Accessed: 27-Mar-2022]

[3]. P. Melville and V. Sindhwani, "Recommender Systems," in Encyclopedia of Machine Learning and Data Mining, Boston, MA: Springer US, 2017, pp. 1056–1066.

[4]. S. Kumar *et al.*, "Novel method for safeguarding personal health record in cloud connection using deep learning models," *Comput. Intell. Neurosci.*, vol. 2022, p. 3564436, 2022.

[5]. A. Singh, P. Singh (2020) Image Classification: A Survey. Journal of Informatics Electrical and Electronics Engineering, Vol. 01, Iss. 02, S. No. 2, pp. 1-9, 2020.

[6]. S. Yadav, P. Singh (2020) Web Application and Penetration Testing. Journal of Informatics Electrical and Electronics Engineering, Vol. 01, Iss. 02, S. No. 3, pp. 1-11, 2020.

[7]. A. K. Sahani and P. Singh, "Web Development Using Angular: A Case Study". *Journal of Informatics Electrical and Electronics Engineering*, Vol. 01, Iss. 02, S. No. 5, pp. 1-7, 2020.

[8]. L. Buitinick, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. Vanderplas, A. Joly, B. Holt, and G. Varoquaux, "API design for machine learning software: experiences from the scikit-learn project," arXiv preprint arXiv:1309.0238, 2013.

[9]. M. Devin, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shelns, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viegas, O. Vinyals, p. Warden, M. Wattenburg, M. Wicke, Y. Yu and X. Zheng "Tensorflow: Large-Scale machine learning on heterogeneous distributed systems," arXiv preprint arXiv: 1603.04467, 2016.

[10]. S. Squarepants, "Bitcoin: A peer-to-peer electronic cash system," *SSRN Electron. J.*, 2008.

[11]. E. Feyen, J. Frost, L. Gambacorta, H. Natarajan, and M. Saal, "Fintech and the digital transformation of financial services: implications for market structure and public policy," *Bis.org*, 2021. [Online]. Available: https://www.bis.org/publ/bppdf/bispap117.pdf. [Accessed: 27-Mar-2022].