IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 2, February 2022

DOI: 10.17148/IJARCCE.2022.11214

Bitcoin Price Prediction

Varad Kolte[1], Shardul Kulkarni[2], Shreya Langar[3], Devyani Avhale[4], Prof. Savitri B. Patil[5]

G H Raisoni College of Engineering and Management, Wagholi, Pune, Maharashtra (India)[1-5]

Abstract: Bitcoin is an innovative payment network and a new kind of money. It is open-source; its design is public; nobody owns or controls Bitcoin and everyone can take part. Bitcoin uses peer-to-peer technology to operate with no central authority or banks; managing transactions and the issuing of bitcoins is carried out collectively by the network. Recently Bitcoin has received a lot of attention from the media and the public due to its recent price hike. As Bitcoin has been viewed as a financial asset and is traded through many cryptocurrency exchanges like a stock market, many researchers have studied various factors that affect the price of Bitcoin and the patterns behind its fluctuations using various analytical and predictive methods. In recent years, Bitcoin is the most valuable in the cryptocurrency market. However, prices of Bitcoin have highly fluctuated which make them very difficult to predict. Hence, this research aims to discover the most efficient and highest accuracy model to predict Bitcoin prices using machine learning algorithms. Using the available information through the dataset, we will predict the sign of the daily price change with highest possible accuracy.

Keywords: Bitcoin, Price Prediction, Blockchain, Cryptocurrency, Random Forest, Predictive Analysis

INTRODUCTION

Bitcoin (₿) is a decentralized digital currency, without a central bank or single administrator, that can be sent from user to user on the peer-to-peer bitcoin network without the need for intermediaries. Transactions are verified by network nodes through cryptography and recorded in a public distributed ledger called a blockchain. The cryptocurrency was invented in 2008 by an unknown person or group of people using the name Satoshi Nakamoto. The currency began use in 2009 when its implementation was released as open-source software. Bitcoins are created as a reward for a process known as mining. They can be exchanged for other currencies, products, and services, but the real-world value of the coins is extremely volatile. Users choose to participate in the digital currency for a number of reasons: ideologies such as commitment to anarchism, decentralization and libertarianism; convenience; using the currency as an investment; and pseudonymity of transactions. Increased use has led to a desire among governments for regulation in order to tax, facilitate legal use in trade and for other reasons (such as investigations for money laundering and price manipulation).

I.METHODS AND ALGORITHM

Random Forest (RF) models are quite possibly the most famous gathering method utilized in order. These models expect to address for the issue of over-fitting in conventional choice trees. This won't be shrouded top to bottom; however, choice trees will quite often learn on sporadic ways of information. RF models train various profound choice trees on various parts of the dataset determined to diminish the general change. In addition to the fact that RF was the most reliable model generally speaking, yet it was the speediest to prepare.

In this project, the random forest model is trained to predict the closing price on the next day. At a high level, random forests are collections of decision trees used for classification and regression tasks. Once grown, each decision tree classifies an unlabeled point by casting a vote, and the random forest reports the label or value with the most votes. A random forest is a type of ensemble model, which averages the predictions of many different "reasonably good" models to produce a prediction that better estimates the true hypothesis. Ensemble models are highly successful as machine learning tools, because they avoid the chance dependent pitfalls of many singular models. For example, gradient descent methods can get stuck in local minima, but combining many models increases the chance that one will find the global minimum. Alternatively, even if none of the models in the ensemble produce the true hypothesis, averaging every prediction can lead to a prediction that more closely matches the underlying truth.

Random forest also can be summarized in the perspectives of Map-Reduce step by step like this:

1. Read train data from a CSV file.
2. For n Trees call n mappers
3. Create 90% subset of the training data with replacement
4. When mapper is running it takes these data
5. After receiving data, each mapper starts to build a tree and produce a prediction for the test dataset.
6. Pass the test data and label as key and value to Reducer
7. Reducer counts the majority label according to the key.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 2, February 2022

DOI: 10.17148/IJARCCE.2022.11214

8. Write the results to the output file.

This representation can be seen at the figure-1

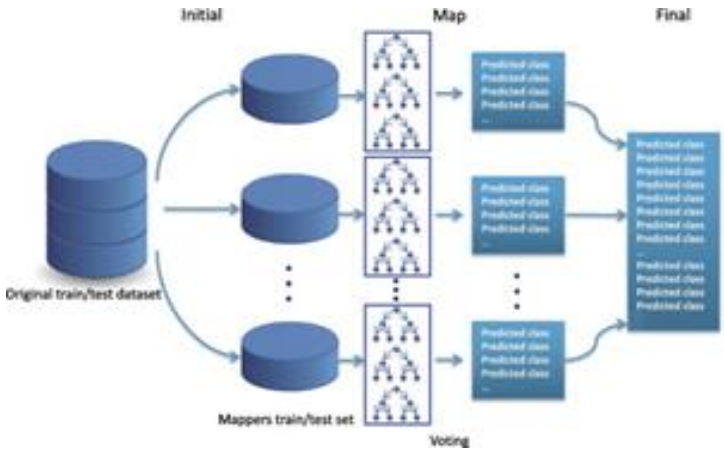

Fig. 1. Random Forest Map-Reduce

The proposed system implements Machine algorithm to build the model to predict the price of bitcoin based on a historical dataset available on online databases. In the proposed model, the bitcoin price prediction can be done using the random forest. The tool used for the project is Jupyter notebook.

The procedure to be followed for the proposed system is given as follows:

1. First, collect the historical dataset from Kaggle which is available on an online database.
2. Arrange the data into the data frame according to the problem definition, so as to get the analysis correct and produce the results which are efficient to meet the goals of the system.
3. The data-pre-processing is performed to missing values for the attributes, this done to reduce the noise and inconsistency in the data.
4. Then we build the mode for the data-set using random forest to predict values of bit-coin on a daily basis.
5. Test the predictions.

This representation can be seen at the figure-2

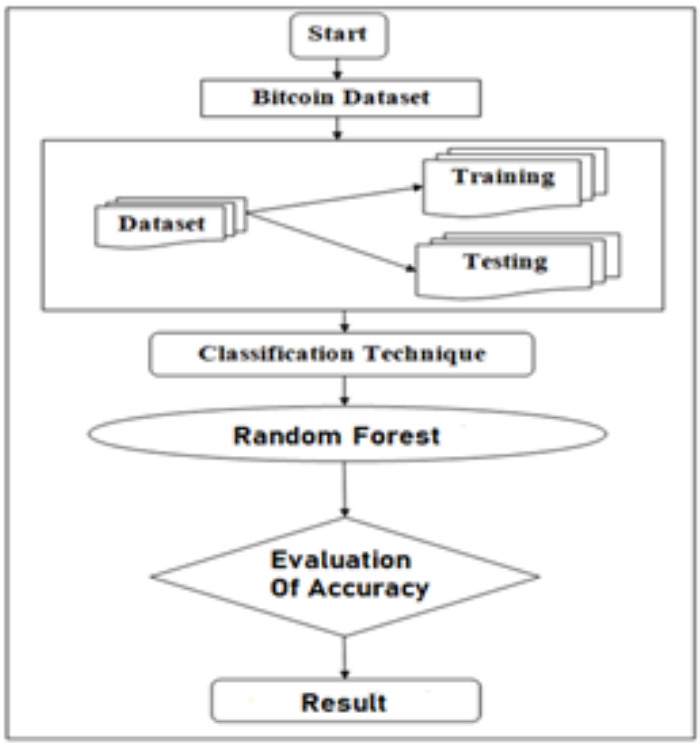

Fig. 2 Flow of algorithm

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 2, February 2022

DOI: 10.17148/IJARCCE.2022.11214

II.DATASET AND REPRESENTATION

Several Bitcoin data sets are available online to download for free. Most of them provide data related to the price of Bitcoin on a day-to-day basis. In this project Kaggle Bitcoin Historical Data is used. The dataset is from January 2012 to March 2021. This dataset gives access to Bitcoin exchanges and daily Bitcoin values. These values:

1. OPEN PRICE: The open represents the first price traded during the candlestick.
2. HIGH PRICE: The high is the highest price traded during the candlestick.
3. LOW PRICE: The low shows the lowest price traded during the candlestick.
4. CLOSE: The close is the last price traded during the candlestick.
5. Volume (BTC): Volume, in BTC, traded in the stock market during a given measurement interval.
6. Volume (Currency): Volume, in USD, traded on the stock market during a given measurement interval.
7. Weighted Price: Measure of the average price.

To predict the closing price of Bitcoin one day ahead, in each of the day data sets, columns close are shifted up by one (1) unit. And also, the features column is shifted down by seven (7) units to use 7-days historical values for prediction. The data representation can be seen at figure-3.

Timestamp	Open	High	Low	Close	Volume_(BTC)	Volume_(Currency)	Weighted_Price	Open_b_1	High_b_1	Low_b_1	...	Low_b_6	Close_b_6
2020-10-13 00:00+00:00	11563.62	11568.44	11563.62	11564.11	0.019977	231.019791	11564.305833	11412.15	11419.61	11412.15	...	10719.24	10720.74
2020-10-14 00:00+00:00	11394.73	11394.73	11377.37	11377.37	2.456657	27964.786790	11383.268769	11563.62	11568.44	11563.62	...	10656.55	10656.55
2020-10-15 00:00+00:00	11335.86	11339.16	11335.86	11339.16	0.018018	204.295733	11338.336446	11394.73	11394.73	11377.37	...	10853.05	10858.85
2020-10-16 00:00+00:00	11463.65	11464.44	11463.65	11464.44	0.016837	193.022212	11464.202940	11335.86	11339.16	11335.86	...	11049.99	11054.04
2020-10-17 00:00+00:00	11370.76	11370.76	11368.00	11369.85	0.388916	4421.226388	11368.085861	11463.65	11464.44	11463.65	...	11355.87	11355.89

Fig. 3 Dataset

III.RESULT

A. Performance Metrics

Root mean square error (RMSE) is a frequently used measure of the differences between values (sample or population values) predicted by a model or an estimator and the values observed. RMSE is a measure of how spread-out prediction errors are. It can be formulated as in fig-4.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (Predicted_i - Actual_i)^2}{N}}$$

Fig. 4 RMSE Formula

In the above formula, n is the number of exemplars in the data set.
The **R2** score is a very important metric that is used to evaluate the performance of a regression-based machine learning model. It is pronounced as R squared and is also known as the coefficient of determination. This metric gives an indication of how good a model fits a given dataset. It indicates how close the regression line (i.e., the predicted values plotted) is to the actual data values. The **R squared value lies between 0 and 1** where 0 indicates that this model doesn't fit the given data and 1 indicates that the model fits perfectly to the dataset provided.

B. Experimental Result

The result of random forest model trained on a data set from 2012 to 2021 (starting from January 2012 to March 2021) and tested in the quarter of 2021(January 2021 – March 2021). Models use window size 7 and 1000 epochs during training and testing. The illustration of the result can be seen in figure-5.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 1, January 2022

DOI: 10.17148/IJARCCE.2022.11108

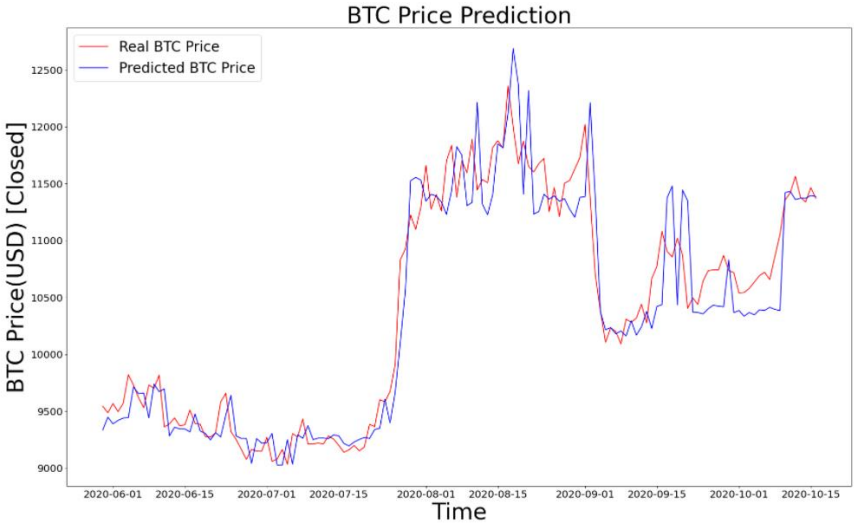

Fig. 5 Result

IV.CONCLUSION

Bitcoin can be seen as the most commonly used digital currency today, attracting investor's attention. In recent years, Bitcoin has grown exponentially which resulted in increasing demand and its recognition by reputable companies such as Amazon, Microsoft, Overstock, DISH Network, Intuit, and even PayPal. In countries such as Japan, Netherlands, Canada and the United States, you can pay Bitcoin at restaurants, malls and other large and small businesses. On the other hand, the prediction of Bitcoin price is not an easy task since it is a new and unstable market. In this project, we utilize it for predicting the real-valued quantity, the price of Bitcoin. Based on this price prediction method, we devise a simple strategy for trading Bitcoin. The strategy is able to nearly double the investment in less than 60-day period when run against real data. The machine learning algorithms will improve that feature idea of crypto currencies. It will also improve the market price of global investments.

This work presents an application of random forest model using Pyspark framework for making one day ahead prediction of closing price of cryptocurrency Bitcoin. It is observed that the bitcoin market is not stable. Since the Bitcoin market is open to manipulation, the price of bitcoin may be hard decrease or hard increase. It may be necessary to provide data that the model can predict manipulations to better capture these ups and downs. Examples of these data are tweets about bitcoin, bitcoin news, and telegram bitcoin groups.

ACKNOWLEDGMENT

We would like to thank **Mrs. Savitri B. Patil** for her guidance and support. We're also thankful to the department of Information Technology and GH Raisoni College of Engineering and Management, Wagholi, Pune for providing essential research facilities.

REFERENCES

[1] Yakup Görür (2018). BITCOIN PRICE DETECTION WITH PYSPARK USING RANDOM FOREST.
[2] Amjad, M. J. & Shah, D. (2017). "Trading Bitcoin and Online Time Series Prediction". Proceedings of the Time Series Workshop at NIPS 2016, (pp. PMLR 55:1-15).
[3] Areekul, P., Senjyu, T., Urasaki, N., & Yona, A. (2010). Next day price forecasting in a deregulated market by combination of Artificial Neural Network and ARIMA time series models. 2010 5th IEEE Conference on Industrial Electronics and Applications. doi:10.1109/iciea.2010.5514828 .
[4] Temesgen Awoke Muniye, Minakhi Rout, Lipika Mohanty, Suresh Satapathy (2020). Bitcoin Price Prediction and Analysis Using Deep Learning Models. Communication Software and Networks, Proceedings of INDIA 2019 (pp.631-640).
[5] Theerasak Phaladisailoed, Thanisa Numnonda (2018). Machine Learning Models Comparison for Bitcoin Price Prediction. Conference: 2018 10th International Conference on Information Technology and Electrical Engineering (ICITEE).
[6] Satoshi Nakamoto. 2008. Bitcoin: A peer-to-peer electronic cash system. (2008).
[7] Thomas G. Dietterich. Ensemble methods in machine learning. In Proceedings of the First International Workshop on Multiple Classifier Systems, MCS '00, pages 1–15, London, UK, UK, 2000. Springer-Verlag.

