Bitcoin Price Prediction using ARIMA Model

Poongodi M

School of Computing Science and Engineering,
Vellore Instititue of Technology,
Chennai, Tamil Nadu, India
E-mail: m.poongodi@gmail.com

Vijayakumar V

School of Computing Science and Engineering,
Vellore Instititue of Technology,
Chennai, Tamil Nadu, India
E-mail:vijayakumar.v@vit.ac.in

Naveen Chilamkurti

Acting Head of Department,
Computer Science and Computer Engineering,
La Trobe University, Melbourne, Australia
E-mail: xiao.z.gao@gmail.com

Abstract: Bitcoin is a highly volatile cryptocurrency with rising popularity. It is a turning point in the way currency is seen. Now the currency, rather than being physical is becoming more and more digital. Bitcoin has removed the central party and has given the control to the users. Due to high variance of solo mining, the number of users joining top most famous Bitcoin mining pools are increasing due to the fact that users together under a Bitcoin pool will have a higher chance of generating next block in the Bitcoins blockchain by reducing the variance and earning the mining reward. Furthermore,emerging mining farms with strong mining resources and fast processing power is another trend towards centralization. This trend clearly illustrates that the pure, decentralized protocol of Bitcoin is going towards centralization in its distribution network, where any kind of centralization should be considered carefully due to the 51% attack. The centralization caused by bitcoin cloud wallets giving new users easy access to the bitcoin network should not be ignored either. This might cause various security issues in the bitcoin technologies due to the hackable applications and websites. There has been much research on the bitcoinâŁ™s decentralised network, and the various price prediction techniques. In this research paper we are doing a survey on the technology lying underneath bitcoinâŁ™s network and the various machine learning predictive algorithms. We collected the dataset on bitcoin blockchain from April 28th, 2013 to July 31 st , 2017 which is publicly available on https://coinmarketcap.com and applied the ARIMA model for price prediction of bitcoin.

1   Introduction

Satoshi Nakamoto, an anonymous person from japan introduced Bitcoin in 2009. It was the first successful attempt in developing a virtual currency and is now probably the most controversial virtual currencies[3]. Many virtual currencies like e-Gold(1996-2014) and Liberty Reserve(2006- 2013) were introduced but were not successful in developing in the market due to the laundering concerns. Other currencies include DigiCash started in 1990s and were close to achieving a global level of success. However, it fell prey to a mixture of internal strife and lack of ignition that stopped it from becoming a widespread online payment mechanism.Bitcoin represented about 41% of the total estimated cryptocurrency capitalization in 2017. The idea behind bitcoin is removing the central trusted authority from the fund transfer system and hence removing the commission and trusting a cryptography algorithm in making transactions.Transactions are made on a peer-to- peer network (i.e. completely decentralized). The payment is cheap due to the absence of middlemen, there is minimal commission in terms of mining reward to the miners. This makes the system secure, fast and reduces the large commissions paid to the central authorities. It can be used to shop online but much of the hype is about trading in it for profits.

2   Bitcoin Market

2.1   Analogy to Stock Market

The Bitcoin markets financial analog is a stock market.To maximize financial reward, the field of stock market prediction has grown over the past decades, and has more recently exploded with the advent of high-frequency, low-latency trading hardware coupled with robust machine learning algorithms [4]. Thus, it makes sense that this prediction methodology is replicated in the world of Bitcoin, as the network gains greater liquidity and more people develop an interest in investing profitably in the system. To do so, it is necessary to leverage machine learning technology to predict the price of Bitcoin.

2.2   Bitcoin as Investment

The community using bitcoin for investment purposes and for trading / making payments will benefit from the precision of the algorithm by getting to know the right time for investing their money in the bitcoin and the time to use bitcoin for purchase. This will help retain the decentralized nature of the network in the coming future and shake off the threat of 51% attack. The bitcoin cloud wallets are another source the pose a threat of centrality to the p2p network of bitcoin. The risk of wallets being hacked and coins stolen is another risk. The wallets store information of many users of that particular wallet and may manipulate them to invest at the wrong time for their own benefit.There are many wallets emerging in the market promising users easy and secure way to invest in bitcoin, gaining a huge commission in the process leaving the users exploited. By having a prediction technique, the risk of exploitation will be avoided and may even encourage more users to the network, reducing centralisation risks.

2.3 Barro's Model

It is a bitcoin pricing model[2] under perfect market conditions. In this model, bitcoin is assumed to possess currency value and is exchanged with traditional currencies which are under central bank control. In equation 1 The total bitcoin supply, S_B , is represented by

$$S_B = P_B \text{ B} \tag{1}$$

Where P_B denotes the exchange rate between bticoin and dollar, and B is the total capacity of Bitcoins in circulation. The total Bitcoin demand D_B depends on the price of goods or services, P; the economy size of Bitcoin, E; and the frequency at which Bitcoin is used for purchasing goods or services (shown in equation 2).

$$D_B = \text{P E / V} \tag{2}$$

In the perfect situation when the supply and demand are the same

$$P_B \ = \frac{PE}{VB} \tag{3}$$

From equation 3 it implies that in perfect situations the price of bitcoin is directly proportional on the price of goods and services at the certain time, economy size of Bitcoin and is inversely proportional to the velocity or frequency of bitcoin transactions and the total capacity of Bitcoin.

3 Working of Bitcoin

3.1 Blockchain and Proof of Work

BlockChain is the technology used behind bitcoin making the system decentrailsed and secure at the same time. It is a data structure that makes it possible to create a digital ledger of data and share it among a network of independent parties. It uses cryptography to allow each participant on any given network to manage the legder in a secure way. The removal of central authority is one of the most important and powerful aspects of blockchains.

When data is recorded in a blockchain, its extremely difficult to change or remove it. When someone wants to add a record to a blockchain, also called a transaction or an entry, users in the network who have validation control verify the proposed transaction.

Bitcoin, unlike other fiat currencies with central banks, aims to achieve complete decentralisation working on the principles of peer-to- peer network by using blockchain technology. The basic goals bitcoin tries to achieve are:

1. Removing the central authority

2. Faster mode to transfer (immediately)

3. Cheaper than the regular currencies

Blockchain is the only available technology that can achieve all three goals at once. Along with blockchain, proof- of-work is also being used to make the network more secure and much more difficult for people to tamper with. This automatically limits the difficult of the problem for the min- ers to ensure that it can be solved within approximately 10 minutes. PoW also provides incentives to motivate the miners to participate in maintaining the ledger and maintaining the value of Bitcoin by paying Bitcoin for those who validate the block and are able to add the block to the blockchain before anyone else. Blockchain helps transfer the currency from one individual to another through its p2p network [6].

3.2 Bitcoin Mining

The blockchain's public ledger is open for viewing all the bitcoin transactions. As soon as a transaction is made, everyone on the network gets to know of the transaction and hence can validate the transaction. The bitcoin network has a distributed ledger (i.e. everyone on the network, or every miner has his own ledger). In the event of a transaction such as the one shown in Fig 1, the person sending the bitcoin broadcasts the transaction in the network. The other nodes or the miners compete with each other to validate the unvalidated transaction, validate it and add it to the block. To validate the transaction the miner checks the ledger for whether the sender is having enough funds to make the transaction, after validating he has to find a unique key (hash) for the block and link it to the chain (i.e. The ledger). As the miner generates the key and links the block to his own ledger (and claims the reward), he broadcasts the validated block along with the key for others to add it to their ledger and then look for the next transaction to work on.
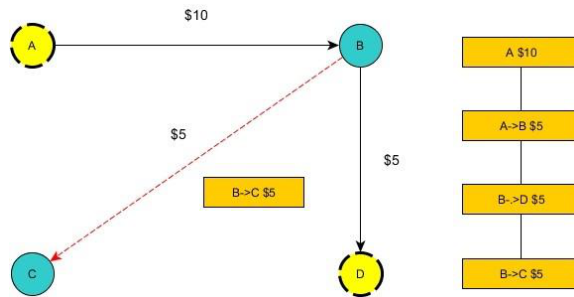

Figure 1 Credibility of every node in the network

3.3 Benefits of Bitcoin

In the regular currency transaction, the bank acts as the third party (Fig 2) leading to the following issues in the system:

- The central trusted authority takes a huge commission,
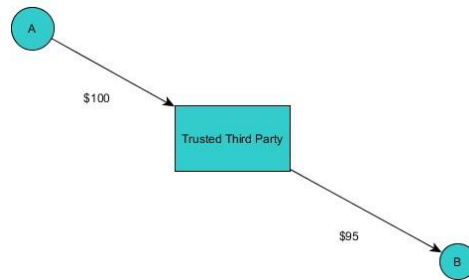
- The transaction takes long time,

Figure 2 Centralized currency transfer

Blockchain helps in eliminating the central authority and distributing control of the system equally among the users. This is done with the help of an open ledger system explained below. Fig 3. shows the bitcoin transaction system in which all the transactions are available to view by everyone on the network and can be validated by anyone, thus removing the central authority from the equation and making the transactions faster (i.e. within no time), and without much transaction costs because of the absence of the central authorities.

Transaction is made by the sender over the network which creates a block having the information like the IDs of the sender, receiver, and the amount transferred, time stamp and the hash of the previous block to link it to the blockchain.
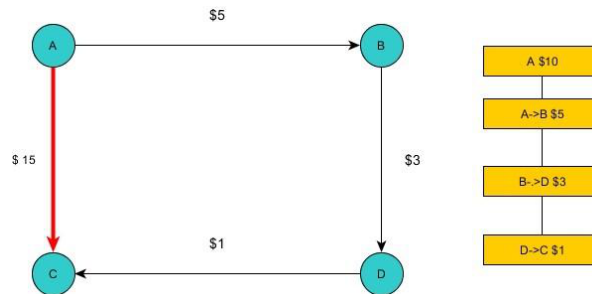
Figure 3 Open Ledge

4 Possible Issues

Bitcoin faces some serious threats or risks, which include 51% problem, centralisation, and considerable time taken to form the agreed blockchain until the longest chain is selected after generation of several blocks. This condition causes time delay in making the transaction. There is a capacity limit to the blockchain. 51% problem[1] arises when majority of the total participants in the network are occupied by a specific group with some specific purpose, they might be able to cause discrepancy in the ledger.

The network is itself doesn't provide total anonymity as the IP addresses [1] of the users can be traced at the time of transaction and can be used to track the location. The online

wallets coming up increase the risk to a much higher level. It is the blockchain which, due to its algoritm is difficult to hack. But with these wallets coming up the risk of being hacked has increased as hackers can hack into the wallets and change the queries to alter the amount user actually wanted to transfer before it is even added to the blockchain. This will lead to the problem of lost bitcoins. Tor network can be used by the users to anonymize themselves further to stay protected. This anonymization is good for the people making large transactions, but along with this it brings in the fact that being anonymous makes is easier for the criminals to move money.

The final risk is that of centralisation caused by the online wallets who provide people with the facility of investing in bitcoin and making transactions in bitcoin. Though there is no other middle man in the network there companies act as a central authorities and charge commission from the people in return for their services. This slowly lead to centralization of the network by having many accounts under them, they become somewhat like a central authority themselves. There is not much time left when bitcoin will be accepted worldwide and become a common currency for the people. If some company providing the online wallet services gains monopoly over the market by that time, the network will move towards centrality and bitcoin will no longer be much different that the other fiat currencies.

5   Predictive Analysis

The price prediction is done using the blockchain information and macroeconomic factors publicly available [2]. The dataset of blockchain contains all the information of the bitcoin, therefore it is a huge dataset hence improving the accuracy of the prediction technique. Machine learning technologies like linear regression methods, SVRs, and Bayesian neural networks(BNN) are used to predict the price variation on providing a proper dataset. Increasing use of Bitcoin as a mode of payment in the online market in the coming future requires a sophisticated price prediction technique due to its highly volatile nature. More and more use of this technology has proven it to be a strong long term and volatile short term investments. A widespread acceptance is on its way among the online services, strengthening the foothold of cryptocurrency and increasing its value in the market. The gigantic mining farms having greater control over the bitcoin blockchain is conflicting the idea of decentralized currency that lead to the development of bitcoin in the first place. These mining farms cause a great threat to decentralisation of bitcoin, making it centralized. This gives certain level of control on bitcoins price in their hands making it a risky investment for others.

We researched on various algorithms used in the past by others in the field of price prediction of bitcoin and compared the accuracy of their results.

For time series analysis [2], nonlinear methods, such as kernel regression model, exponential autoregressive models, artificial neural network(ANN), and support vector regression, have exhibited improved predictive performance for various time series data.

5.1   Regression Model

Linear Regression [5] is a predictive model that formulates a best-fitting straight line through the scalar dependent variables (target) and independent variables (predicator). Thebest-fitting is called a regression line. Linear fit occurs by minimizing the mean squared error between the predicted and actual output.

Logistic Regression is a predictive regression model in which the dependent variables are categorical (For ex, event of success or failure). It used maximum likelihood estimation to formulate the probabilities in which Logistic Regression will take on a particular class, with an iterative algorithm.

5.2 Support Vector Machine (SVM)

Support Vector Machine [5] is a supervised machine learning algorithm which is used for both, classification and regression problems. Linear decision boundaries are augmented to more complex boundary shape through kernel implementation. SVM obtains decision boundary by creating a margin which maximizes the functional and geometric margins between classes. SVMs have gained popularity for their impressive performance in classification. It can also be augmented for regression problems as Support Vector Regression (SVR) optimizing response variable distance from the decision boundary.

5.3 Neural Network

Neural Network are a family of learning methods inspired by biological neural networks by modeling a system of interconnected neurons, which are tuned based on iterative learning. Feedforward neural networks connect a multidimensional input into one or more hidden layers of neurons before predicting an output. Using Bayesian regression approach allows us to utilize the existence of patterns for the purpose of better prediction without explicitly nding them [7]. Baeysian neural network (BNN) is a transformed multi layer perception (MLP) which is a general term for ANNs [2]. The networks have been successful in many applications such as image recognition, pattern recognition, natural language processing and financial time series. The structure of BNN is constructed with a number of processing units classified into three categories: an input layer, an output layer, and one or more hidden layers. A BNN is a non-linear version of ridge regression, which is largely based on the Bayesian theory of neural networks. BNN [2] is a machine maximizing the value of posterior through an application of the Bayes theory.

5.4 Sentiment Analysis

An empirical model incorporating multiple economic and technological variables was built and also extended the existing literature by taking Twitter sentiment into account [6]. After dealing with issues of stationarity and cointegration, several regression models were estimated indicating that the Twitter sentiment ratio has a positive short-run impact on Bitcoin prices. The frequency of transactions is very effective feature for HYIP classication, the classication performance slightly degrades when AC is applied and it has been shown that the proposed pattern extraction for summarizing transactions can effectively distinguish the difference of usage of Bitcoin by classes [8].

6 ARIMA

Auto Regressive Integrated Moving Average model is one of the most general class of models for forecasting a time series which can be made to be stationary by differencing, in conjunction with nonlinear transformations such as logging or deflating. Random-walk and random-trend models, autoregressive models, and exponential smoothing models are

all special cases of ARIMA models. For our research purpose we used time series ARIMA model without trends and seasonalities as the prediction technique to predict Bitcoin's closing rate.

Seasonality is a certain kind of autocorrelation pattern which occurs repeatedly after regular intervals of time for example on monthly, quarterly bases. Seasonality must be corrected before training the model as the model might consider various seasonal patterns which will give false results. A trend is a long-term increase or decrease in the level of the time series. If trends are left unanalyzed , they lead to misreading and misinterpretation the dataset. Removing the seasonality refers to removing regular peaks and valleys from the sequence of data-points.

Seasonal ARIMA model is denoted by ARIMA(p,d,q)(P,D,Q) m with P, D and Q as its seasonal autoregressive, differencing and moving average parts and m as the number of periods in each season. while the non-seasonal ARIMA model is classified as ARIMA(p,d,q) model, where:

- p is number of autoregressive terms (i.e. order or AR model),

- d is the number differences needed for stationarity (i.e. degree),

- q is the number of lagged forecast errors in the prediction equation (i.e. order of moving-average model).

7   Results

Bitcoin being a really volatile currency makes it difficult to predict. We compared the closing price of Bitcoin each day with the ones predicted by us and got a satisfactory accuracy of 49%.
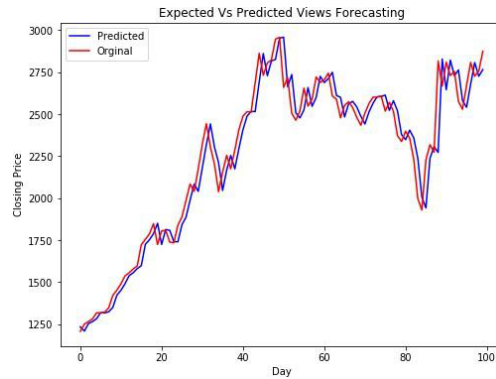

Figure 4   Closing Price-Time(days) graph

Fig 4. shows the Bitcoin price plotted against time (days) on the dataset of Bitcoin market stats from April 28th, 2013 to July 31st, 2017 while Fig 5. shows the results obtained from the hourly based analysis of the actual price and predicted price. The red line shows the actual price and the blue shows the predicted price which clearly explains about the

volatality of the currency if measured on short term (hourly) instead of slightly longer periods. This caused the huge variation in the predicted and actual price.

Further removing the trends and seasonalities makes the predictions even better as the time series models tend to pick up some regularities in patterns along time giving misleading results.

8 CONCLUSION

Bitcoin is an alternate for the fiat currencies nearing to being used globally in every community but to the benefits it provides. Even after its shortcomings bitcoin market is being preferred for investment and is slowly turning into a network being used for making payments too. A few shortcomings being the major ones like restrictions by the government due to the tax invasion.

Various price prediction models are being used for predicting the price of this extremely unpredictable crypto-currency. We concluded that from all the features of the Blockchain network the features which influence the price of bitcoin the most, are:

1. Total bitcoin transactions,

2. Frequency of transactions,

3. Current bitcoin price,

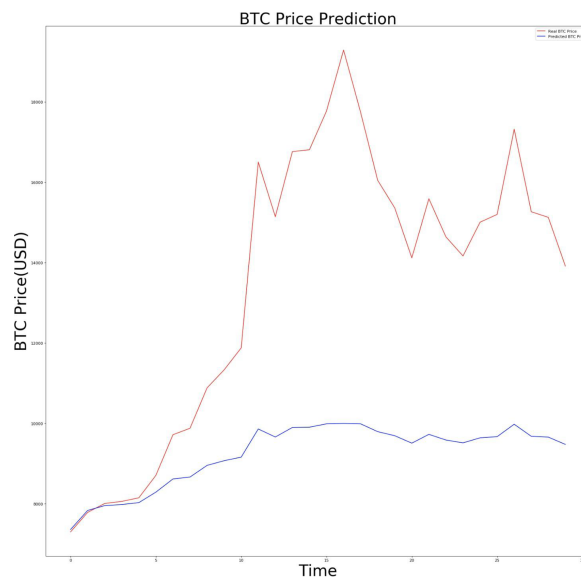4. Market capital present in the form of bitcoin;

Figure 5 Price-Time(hours) graph

Neural Networks gave the most satisfying results in the past, now on trying to determine the price using time series ARIMA model it gave similar results. It gives improved results on removing the seasonality occurring in the model.

References

[1] Sahar Mirzayi, Mohammad Mehrad BITCOIN, An SWOT Analysis

[2] Huisu Jang, Jaewook Lee An Empirical Study on Modeling and Prediction of bitcoin Prices with Bayesian Neural Networks Based on Blockchain Information

[3] Paraskevi Katsiampa Volatility estimation for bitcoin: A comparison for GARCH models

[4] Isaac Madan, Shaurya Saluja, Aojia Zhao Automated Bitcoin Trading via Machine Learning Algorithms

[5] Alex Greaves, Benjamin Au USING THE BITCOIN TRANSACTION GRAPH TO PREDICT THE PRICE OF BITCOIN

[6] Ifigeneia Georgoula, Demitrios Pournarakis, Christos Bilanakos, Dionisios N. Sotiropoulos, George M. Giaglis Using Time-Series and Sentiment Analysis to Detect the Determinants of Bitcoin Price

[7] Devavrat Shah, Kang Zhan Bayesian regression and Bitcoin

[8] Kentaroh Toyoda, Tomoaki Ohtsuki, and P. Takis Mathiopoulo Iden- tication of High Yielding Investment Programs in Bitcoin via Trans- actions Pattern Analysis