International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 03 | Mar 2019 www.irjet.net p-ISSN: 2395-0072

Predicting Bitcoin Prices using Convolutional Neural Network algorithm

Kanchana.A[1], Priyadharshini.G[2], Muthuselvi.R[3], Prathiba.K[4]

[1]Associate Professor, Department of computer science and Engineering, Panimalar Engineering College Chennai – 600 123, Tamilnadu, India
[2,3,4]U. G. Scholar, Department of computer science and Engineering, Panimalar Engineering College, Chennai – 600 123, Tamilnadu, India

---

Abstract — *The Predict the crypto cash costs precisely mulling over different parameters that influence the Bit coin esteem. For the primary period of our examination, we expect to comprehend and recognize every day drifts in the Bit coin advertise while picking up understanding into ideal highlights encompassing Bit coin cost. Our informational index comprises of different highlights identifying with the Bit coin cost and instalment arrange through the span of five years, recorded day by day. For the second period of our examination, utilizing the accessible data, we will anticipate the indication of the day by day value change with most noteworthy conceivable exactness.*

Key Words: Crypto Currency, Bitcoin, Machine Learning, Deep learning, Linear algorithm.

I. INTRODUCTION

Most likely one of the greatest things in 2017, Bitcoin developed by around 800% that year, held a market top of around 250 billion dollars, and started overall enthusiasm for crypto monetary standards. Fundamentally they're advanced monetary forms that utilization complex PC calculations and encryption to create more money and to secure exchanges. What's truly cool about crypto monetary forms is that they use a system of thousands of PCs that forward individuals' exchanges to what's known as a square chain.Once an exchange is in the square chain, it's failing to come out once more; this shields crypto monetary standards from twofold spends.So it's quite evident that crypto monetary forms are a cool better approach to burn through cashThere's a great deal of information identified with Bitcoin—37 distinctive qualities of it on bitcoin.com (counting value, square chain measure, advertise top, and so on.). This information has been gathered since July of 2010, so there wound up being around 60 thousand distinct information focuses to process.

II. LITERATURE SURVEY

In 2014, Bissi et al Network Layer The second bunch of crypto cash related research centers around the system layer.Most of the papers gathered in this segment look at the Bitcoin shared system,just Bissias et al.[2014] included Litecoin as option to Bitcoin into their research.Anish Dev [2014] names additionally other themes derivates of Bitcoin. In 2014, Gervais et al Simplified installment check (SPV) customers are an extra importantconcept to cultivate Bitcoin as an option for e-Business exchanges. As explicit gadgets (likemobile telephones) have a constrained measure of information stockpiling and can't store the total blockchain.SPV customers enable friends to separate Bitcoin exchanges applicable for the customer while outsourcingtransaction approvals to all the more dominant system peers (Gervais et al. 2014a). Gervais et al.show that these "channels bring about genuine protection spillage in existing SPV customer implementations"(2014a) and propose a lightweight adjustment of the SPV customers. In 2014, El Defrawy and Lampkins Ecosystem Layer most of writing looked atexamined the crypto cash biological system. Like in the past segments, Bitcoin is the dominantCC inspected. El Defrawy and Lampkins (2014), Malone and O'Dwyer (2014) and Taylor (2013) notice other crypto monetary standards like Litecoin, however base their exploration on Bitcoin.BenSasson et al. (2014) present Zerocash as an option for decentralized mysterious installments.

In 2013, Danezis et al.The created convention from Jayasinghe et al. "ensures solid fairnesswhile protecting namelessness of the customer and the shipper" (2014). Andrychowicz et al. (2014) engineer a convention to verify multiparty lotteries without a confided in power which is builton the Bitcoin convention. Just crafted by Danezis et al. (2013) and Miers et al. (2013) are builton Zerocoins, a crypto cash for mysterious decentralized exchanges. The convention uses"modern strategies dependent on quadratic number juggling programs bringing about littler evidences andquicker check" (Danezis et al. 2013).

III. RELATED WORKS

The objective is to find out with what precision can the course of Bit-coin cost in USD can be anticipated. The value information is sourced from the Bitcoin Price Index. The assignment is accomplished with changing degrees of achievement through the execution of a Bayesian streamlined repetitive neural system (RNN) and Long Short Term Memory (LSTM) organize.The LSTM accomplishes the most noteworthy grouping exactness of 52% and a RMSE of 8%.

- Less prediction accuracy
- Not applicable for all crypto curriencies
- Not a real time analysis

IV. PROPOSED SYSTEM

we intend to comprehend and recognize day by day inclines in the Bit coin showcase while picking up understanding into ideal highlights encompassing Bit coin cost. Our informational collection comprises of different highlights identifying with the Bitcoin cost and installment organize through the span of five years, recorded every day. For the second period of our examination, utilizing the accessible data, we will fore see the indication of the everyday value change with most elevated conceivable precision.

1. Data collection.

2. Data preprocessing.

3. Machine learning algorithm training

4. Training and testing

	bt_Close	bt_Volume	bt_close_off_high	bt_volatility	eth_Close	eth_Volume	eth_close_off_high	eth_volatility
688	0.000000	0.000000	-0.560641	0.020292	0.000000	0.000000	-0.418477	0.025040
687	-0.002049	-0.170410	0.250597	0.009641	-0.011498	0.239937	0.965898	0.034913
686	-0.009946	0.092475	-0.173865	0.020827	0.025190	0.978201	-0.317885	0.060792
685	-0.002855	0.060603	-0.474265	0.012649	0.006810	0.680295	-0.057657	0.047943
684	-0.005457	-0.048411	-0.013333	0.010391	0.002270	0.066829	0.697930	0.025236
683	-0.012019	-0.061645	-0.003623	0.012782	0.002991	0.498534	-0.214540	0.026263
682	0.054613	1.413585	-0.951499	0.069045	-0.006349	2.142074	0.681644	0.040587
681	0.043515	0.570968	0.294196	0.032762	0.040890	1.647747	-0.806717	0.055274
680	0.030576	-0.110282	0.814194	0.017094	0.040937	0.098121	-0.411897	0.019021
679	0.031451	-0.007801	-0.919598	0.017758	0.054014	0.896944	-0.938235	0.025266

Constant information gathered fromTwitter, kaggle, UCI, Data.gov.Collection of information is one of the major and most critical undertakings of any AI ventures. Since the information we feed to the calculations is information. In this way,the calculations productivity and precision relies on the accuracy and nature of information gathered. So as the information same will be the output.The information is inputed as a .csv record.

This technique will change the information from a variety of shape (n x m), where n speaks to the quantity of days and m speaks to the quantity of highlights identifying with bitcoin, to a tensor of shape (n-w x d x m), where d speaks to the quantity of days to take a gander at in each example of information and w speaks to the window measure. This will be rehearsed by using a period game plan change to change the principal show into a great deal of windows data (window_size = 50).

The information will at that point be standardized by taking a gander at every window and partitioning each an incentive in the window by the primary estimation of the window and after that subtracting one. For instance, the standardization strategy will change the arrangement of information [4,3,2] into [0, - 0.25, - 0.5]. These qualities are gotten by separating all qualities in the information by the primary esteem, for this situation 4, at that point 1 is subtracted from each subsequent esteem (for example 3 would progress toward becoming (3/4)- 1, or - 0.25).

The unnormalized bases are kept in order to get the original values back for the testing data. This is necessary to compare the model's predictions of prices with the true prices.

After normalization, the first 90% of the data is used in training the model, and the last 10% will be used to test the model. These data will be stored in X_train, Y_train, X_test, and Y_test. The preparation information will be rearranged with the end goal that the request of days in every window stays steady, yet the request of the windows will be arbitrary. At long last, a rundown of the costs before every day Y_test is drawn from will be accumulated so as to create measurements about the model's expectations, including accuracy, review, and F1 score. Moreover, these costs can be utilized to recognize whether the model anticipated an expansion or lessening in cost.

V. COLUMNS DESCRIPTION

Annual Hash Growth: Growth in the total network computations over the past 365 days.

Block Height: The total number of blocks in the blockchain.

Block Interval: Average amount of time between blocks.

Block Size: The storage size of each block (i.e. megabytes).

BlockChain Size: The storage size of the blockchain (i.e. gigabytes).

Daily Blocks: Number of blocks found each day.

Chain Value Density: The value of bitcoin's blockchain, in terms of dollars per megabyte.

Daily Transactions: The number of transactions included in the blockchain per day.

Difficulty: The minimum proof-of-work threshold required for a bitcoin miner to mine a block.

Fee Percentage: Average fee paid as a percentage of transaction volume.

Fee Rate: Average fee paid per transaction.

Two-Week Hash Growth: Growth in the total network computations over the past 14 days.

Hash Rate: The number of block solutions computed per second by all miners.

Market Capitalization: The market value of all bitcoin in circulation.

Metcalfe's Law - TX: A variant of Metcalfe's Law in which price is divided by n log n number of daily transactions.

Metcalfe's Law - UTXO: A variant of Metcalfe's Law in which price is divided by n log n number of unspent transaction outputs.

Miner Revenue Value: The amount of dollars earned by the mining network.

Money Supply: The amount of bitcoin in circulation.

Output Value: The dollar value of all outputs sent over the network.

Output Volume: The amount of Bitcoin sent over the network.

Bitcoin Price: The amount of dollars a single bitcoin is worth.

Quarterly Hash Growth: Growth in the total snetwork computations in the past 90 days.

Total Transactions: The running total number of transactions processed by the Bitcoin network.

Transaction Amount: The average amount of bitcoin moved per transaction.

Fees Value: The dollar value of mining fees.

Transaction Fees: The amount of bitcoin paid to miners in fees.

Transaction Size: The average data size of a transaction.

Transaction Value: The average dollar value moved in each transaction.

Transactions per Block: The number of transactions in each block.

Average UTXO Amount: The average amount of bitcoin contained in each unspent transaction output.

UTXO Growth: The net number of unspent transaction outputs created.

UTXO Set Size: The total number of unspent transaction outputs.

Average UTXO Value: The average dollar value of each uspent transaction output.

Velocity - Daily: The proportion of the money supply transacted each day.

Velocity - Quarterly: The proportion of the money supply transacted each day, computed on a rolling-quarter basis.

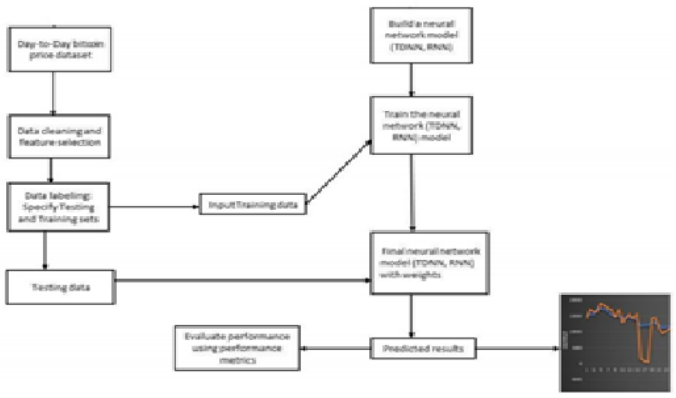
Velocity of Money: How many times the money supply changes hands in a given year.

Miner Revenue: The amount of bitcoin earned by the mining network, in the form of block rewards and transaction fees.

Collecting the data is one task and making that data useful is another vital task. Data collected from various means will be in an unorganized format and there may be lot of null values, in-valid data values and unwanted data. Cleaning all these data and replacing them with appropriate or approximate data and removing null and missing data and replacing them with some fixed alternate values are the basic steps in preprocessing of data.

Even data collected may contain completely garbage values. It may not be in exact format or way that is meant to be. All such cases must be verified and replaced with alternate values to make data meaning meaningful and useful for further processing.Data must be kept in an organized format.

VI. ARCHITECTURE DIAGRAM



Finaly after processing of data and training the very next task is obviously testing. This is where performance of the algorithm, quality of data, and required output all appears out. From the huge data set collected 80 percent of the data is utilized for train-ing and 20 percent of the data is reserved for testing. Training as discussed before is the process of making the machine to learn and giving it the capability to make further predictions based on the training it took. Whereas testing means already having a predefined data set with output also previously labeled and the model is tested whether it is working properly or not and is giving the right prediction or not.If maximum number of predictions is right then model will have a good accuracy percentage and is reliable to continue with otherwise better to change the model.

VII. SIGNIFICANCE TEST

This significance test will check to see if the data provides convincing evidence that the F1 score achieved by the model is statistically greater than the average F1 score obtained by guessing.

If conditions are met, the test will use a 1 sample t test for proportions (F1 score is a proportion of precision and recall). The conditions are as follows: a random sample, normality, and large population size. The null and alternative hypothesis must also be stated. The following variables are used in the checking of the conditions and in later calculations:

p_0: the population proportion (in this case, an F1 score of 0.5)

n: the sample size (in this case, there are 267 prices used in the testing data)

p̂: the sample proportion (in this case, the model's F1 score of 0.584905660377)

z; the z statistic of the test

z^*; the z-score value in a normal distribution

H_o; the null hypothesis

H_a; the alternative hypothesis Random

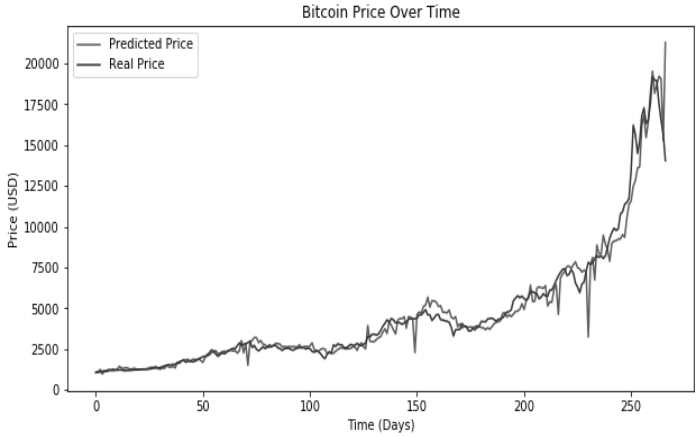
This can be disregarded in this instance, since it is necessary that the model predict on the most recent days of data

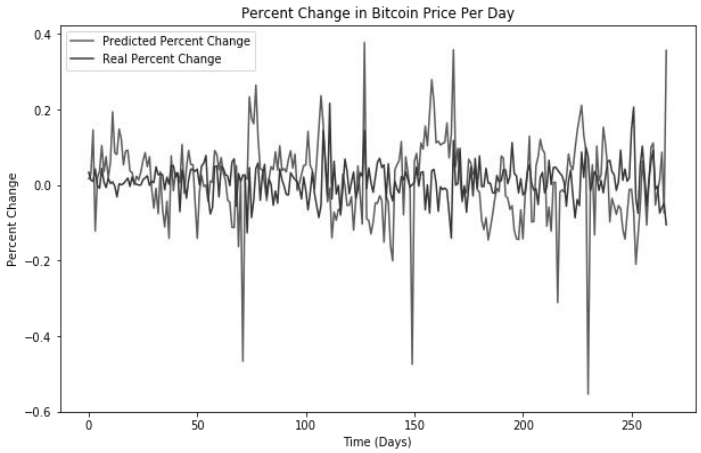
Normality

The p value is 0.0028. Since the p value, 0.0028, is less than α=0.01, the null hypothesis is rejected.There is sufficient evidence to show that the F1 score of the model is statistically greater than the F1 score obtained by guessing. If the proportions really were the same, then the probability of getting a sample F1 score as extreme as this is 0.0028. This is statistically significant at the α=0.01 significance level, and any other reasonable significance level (i.e. α=0.05). The null hypothesis, H_o: p=p_0 is rejected. This means that the idea that the model obtained an F1 score of .584905660377 through guessing is rejected.

Since the null hypothesis has been rejected, there is the probability of making a Type I error. This would be if the idea that the model obtained such a high F1 score by change is rejected, when in reality it is true. However, since it was tested at the α=0.01 significance level, the probability of making this type of mistake is only 0.01 (or 1%), since P (Type I Error) = α and α = 0.01.

VII. Experimental results





IX. CONCLUSION AND FUTURE ENHANCEMENT

A powerful web application can be developed where inputs are not given directly instead student parameters are taken by evaluating students through various evaluations and examining. Technical, analytical, logical, memory based, psychometric and general awareness,interests and skill based tests can be designed and parameters are collected through them so that results will be certainly accurate and the system will be more reliable to use.Also decision trees have few limitations like over fitting, no pruning, lack of capability to deal with null and missing values and few algorithms have problem with huge number of values. All these can be taken into consideration and even more reliable and more accurate algorithms can be used. Then the project will be more powerful to depend upon and even more efficient to depend upon.

X. REFERENCES

[1] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," Available:http:// bitcoin.org/ bitcoin.pdf , 2008.

[2] G. O. Karame, E. Androulaki, and S. Capkun, "Double-spending fast payments in bitcoin," in Proceedings of the 2012 ACM Conference on Computer and Communications Security, ser. CCS '12. New York,NY, USA: ACM, 2012, pp. 906–917.

[3] E. Heilman, A. Kendler, A. Zohar, and S. Goldberg, "Eclipse attacks on bitcoin's peer-to-peer network," in Proceedings of the 24th USENIX Conference on Security Symposium, ser. SEC'15. USENIX Association,2015, pp. 129–144.

[4] C. Decker and R. Wattenhofer, "Bitcoin transaction malleability and mtgox," in ESORICS 2014: 19th European Symposium on Research

in Computer Security. Springer International Publishing, 2014, pp.313–326.

[5] A. Maria, Z. Aviv, and V. Laurent, "Hijacking bitcoin: Routing attacks on cryptocurrencies," in Security and Privacy (SP), 2017 IEEE Symposium on. IEEE, 2017.

[6] I. Eyal and E. G. Sirer, "Majority is not enough: Bitcoin mining is vulnerable," in Financial Cryptography and Data Security: 18th International Conference. Springer Berlin Heidelberg, 2014, pp. 436–454.

[7] A. Sapirshtein, Y. Sompolinsky, and A. Zohar, "Optimal selfish mining strategies in bitcoin," in Financial Cryptography and Data Security:20th International Conference, FC 2016, Christ Church, Barbados,February 22–26, 2016. Springer Berlin Heidelberg, 2017, pp. 515–532.

[8] K. Nayak, S. Kumar, A. Miller, and E. Shi, "Stubborn mining:Generalizing selfish mining and combining with an eclipse attack,"in 2016 IEEE European Symposium on Security and Privacy (EuroSP), 2016, pp. 305–320.

[9] I. Eyal, "The miner's dilemma," in Proceedings of the 2015 IEEESymposium on Security and Privacy, ser. SP '15. Washington, DC, USA: IEEE Computer Society, 2015, pp. 89–103.

[10] J. Bonneau, A. Miller, J. Clark, A. Narayanan, J. A. Kroll, and E. W.Felten, "Sok: Research perspectives and challenges for bitcoin and cryptocurrencies," in 2015 IEEE Symposium on Security and Privacy,May 2015, pp. 104–121.

[11] WikiLeaks, "Donation request via cyrptocurriencies," Available: https:// shop.wikileaks.org/donate.

[12] W. F. Slater, "Bitcoin: A current look at the world most popular,enigmatic and controversial digital cryptocurrency," in A Presentation for Forensecure 2014, April 2014.

[13] "Status about bitcoin technoogy was obtained from-what 2016 holds for bitcoin business," Available: http://www.coindesk.com/what-2016-holds-for-bitcoin-businesses/.

[14] M. T. Alam, H. Li, and A. Patidar, "Bitcoin for smart trading in smart grid," in the 21st IEEE International Workshop on Local and Metropolitan Area Networks, April 2015, pp. 1–2.

[15] Y. Zhang and J. Wen, "An iot electric business model based on the protocol of bitcoin," in 2015 18th International Conference on Intelligence in Next Generation Networks, Feb 2015, pp. 184–191.

[16] S. Huckle, R. Bhattacharya, M. White, and N. Beloff, "Internet of things, blockchain and shared economy applications," Procedia Comput. Sci., vol. 98, pp. 461–466, Oct. 2016.

[17] A. Lei, H. Cruickshank, Y. Cao, P. Asuquo, C. P. A. Ogah, and Z. Sun "Blockchain-based dynamic key management for heterogeneous intelligent transportation systems," IEEE Internet of Things Journal, no. 99,2017.

[18] M. Mettler, "Blockchain technology in healthcare: The revolution starts here," in 2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom), Sept 2016, pp. 1–3.

[19] K. Biswas and V. Muthukkumarasamy, "Securing smart cities using blockchain technology," in 2016 IEEE 18th International Conference on High Performance Computing and Communications (HPCC/SmartCity/DSS), Dec 2016, pp. 1392–1393.

[20] F. Tschorsch and B. Scheuermann, "Bitcoin and beyond: A technical survey on decentralized digital currencies," IEEE Communications Surveys Tutorials, vol. 18, no. 3, pp. 2084–2123, 2016.

[21] S. your wallet, "The bitcoin wiki," Available: https:// en.bitcoin.it/wiki/Securing your wallet, Mar. 2014.

[22] G. Andresen, "Bip 16: Pay to script hash," Available: https:// github.com/bitcoin/ bips/ blob/ master/ bip-0016.mediawik, Jan. 2012.

[23] J. R. Douceur, "The sybil attack," in the First International Workshop on Peer-to-Peer Systems, ser. IPTPS '01. London, UK: Springer-Verlag, 2002, pp. 251–260.

[24] A. Back, "Hashcash - a denial of service asure"Available:http://www.hashcash.org/papers/hashcash .pdf , 2002.

[25] D. E. III and T. Hansen, "US secure hash algorithms (sha and shabased hmac and hkdf)," Available: http://www.ietf.org/ rfc/ rfc6234.txt, 2011.