ISSN : 2394-2975 (Online)
ISSN : 2394-6814 (Print)
Vol. 7, Issue 1 (Jan. - Mar. 2020)
International Journal of Advanced Research in Education & Technology (IJARET)

Bitcoin Price Prediction Using Machine Learning

[I]Atul Chaudhary, [II]Arush Agarwal
[I,II]Dept. of CSE, Meerut Institute of Engineering Technology, Meerut, UP, India

Abstract
In this paper, we tried to estimate the Bitcoin price precisely taking into consideration various parameters that affect the Bitcoin value. In our work, we pointed to understand and identify daily changes in the Bitcoin market while obtaining insight into most appropriate features surrounding Bitcoin price. We will predict the daily price change with highest possible accuracy. The market capitalization of publicly traded cryptocurrencies is currently above $230 billion. Bitcoin, the most valuable cryptocurrency, serves primarily as a digital store of value, and its price predictability has been well-studied. These characteristics are outlined in the following subsection; the underlying details of Bitcoin, as they are described in depth in the cited papers.

Keywords
Bitcoin, Cryptocurrency, Context layer, Time stamp, Blockchain.

I. Introduction

A. Bitcoin

Bitcoin uses a peer-to-peer technology to operate with no central authority or banks. Bitcoin is open-source; its design is public, nobody owns or controls Bitcoin and everyone can take part. Digital currency brings into use as open source software in 2009 by pseudonymous creator Satoshi Nakamoto It is a cryptocurrency, so-called because it uses cryptography to control the creation and transfer of money. Users send payments by broadcasting digitally signed messages to the network. Participants known as miners verify and timestamp transactions into a shared public database called the blockchain, for which they are rewarded with transaction fees and newly minted bitcoins. Conventionally "Bitcoin" capitalized refers to the technology and network whereas "bitcoins" lowercase refers to the currency itself. Bitcoins can be obtained by mining or in exchange for products, services, or other currencies.

B. Prediction

The Bitcoin's value varies just like any other stock. There are many algorithms used on stock market data for price forecast. However, the parameters affecting Bitcoin are different. Therefore, it is necessary to foretelling the value of Bitcoin so that correct investment decisions can be made. The price of Bitcoin does not depend on the business events or intervening government authorities, unlike the stock market. Thus, to forecast the value we feel it is necessary to leverage machine learning[6][7] technology to predict the price of Bitcoin.

II. Literature Survey

Bitcoin Price Prediction Using Ensembles of Neural Networks[5]. Here they explored the relationship between the features of Bitcoin and the next day change in the price of Bitcoin using an Artificial Neural Network ensemble approach called Genetic Algorithm based Selective Neural Network Ensemble, they have constructed the neural network using Multi-Layered Perceptron. To better understand the practicality and its effectiveness in real world application, the entity was used to predict the next day direction of the price of Bitcoin given a set of approximately 200 features of the cryptocurrency over a span of 2 years. Over a span of 50 days, a trading strategy based on the ensemble was compared against a –previous day trend following trading strategy through back-testing. The former trading strategy generated almost 85% returns, outperforming the –previous day trend following trading strategy which produced an approximate 38% returns and a trading strategy that follows the single, best MLP (Multilayer Perceptron) model in the ensemble that generated approximately 53% in returns.

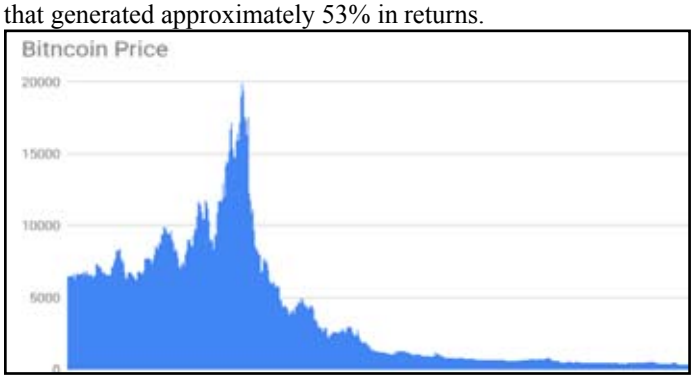

Fig.1 : Bitcoin Price History

III. Datasets

The primary dataset consists of the price of Bitcoin sampled at approximately one-hour intervals between October 10, 2015 and March 01, 2019.

Table 1 : Sample Dataset

Date	Open	High	Low	Close	Volume
2018-10-25 23:00:00	6405.08	6408.66	6405.08	6408.66	0.04456267
2018-10-25 22:00:00	6397.5	6408.66	6394.84	6405.08	23.39852208
2018-10-25 21:00:00	6396.56	6402.63	6393.99	6397.5	35.4535606

The above dataset is downloaded from https://www.cryptodatadownload.com

IV. Methods

Multiple models were assessed on the task of predicting the directionality of change in Bitcoin price.Classification models like Logistic Regression,Support Vector Machine. Other models were based on regression algorithms, such as the Autoregressive

ISSN : 2394-2975 (Online)
ISSN : 2394-6814 (Print)

integrated moving average (ARIMA). Models based on a Recurrent neural network(RNN) implemented and tested.
All of the models were assessed on how well they performed on the task, and these results are analysed. The impetus for trying such a large number of models was to analyze how the assumptions underlying each of the respective models could affect the models performance. The methods underlying these models and their assumptions are briefly summarized below.

A. Logistic Regression

It is a statistical method for examine a dataset in which there are one or more individualistic variables that determine an outcome. The outcome is measured with a divided variable (only two possible outcomes). It is used to predict a binary outcome (1 / 0, Yes / No, True / False) given a set of independent variable. It is a predictive regression model in which the dependent variable is categorical. It uses Maximum Likelihood Estimation to formulate the probabilities in which Logistic Regression will take on a particular class.

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

where x is the input and θ the parameter that must be learned.

B. Support Vector Machine

Like logistic regression, the support vector machine algorithm yields a binary classification model while making very few assumptions about the dataset. The classifier is obtained by optimizing:

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

where x is the input and w,b are parameters that must be learned. Predictions are made by analyzing the value of wTx + b.

C. Auto Regressive Integrated Moving Average (Arima)

ARIMA is a model used for time series analysis and forecasting. The model is used on time series data which will be transformed into a stationary time series; the predictions are a linear regression upon features including time differences and moving averages. The implementation used is from the Statsmodels package (Seabold and Perktold, 2010). In ARIMA, the data is difference that is, the price features are transformed to the difference between prices.

- p: number of autoregressive term.
- d: number of non-seasonal differences needed for stationary
- q: number of logged forecast errors in the prediction equation

$$\left(1 - \sum_{k=1}^{p} \alpha_k L^k\right)(1 - L)^d X_t = \left(1 - \sum_{k=1}^{q} \beta_k L^k\right)\epsilon_t$$

Let L be the lag operator, in the above equation and p,d,q are hyper-parameters over which we optimized. At each time t, we train a model using the price history to predict the price at time t and use the sign of the change in price as a prediction.

D. Recurrent Neural Networks(Rnn)

The RNN (Recurrent neural network) was first developed by scientist Elman. The RNN is structured similarly to the MLP(multi-layer perceptron), with the exception that signals can flow both forward and backwards in an iterative manner. In order to facilitate both the backward and forward flow an addition layer has been added called the **Context Layer.** In addition to passing input between layers, the output of each layer is fed to the context layer to be fed into the next layer with the next input. In this context, the state is overwritten at each timestep. This offers the benefit of allowing the network to assign particular weights to events that occur in a series rather than the same weights to all input as with the MLP.
We used Long short-term memory(LSTM)cells. We tried different numbers of units for the layers, training times, and batch sizes. We have implemented the neural networks with both Keras and TensorFlow.

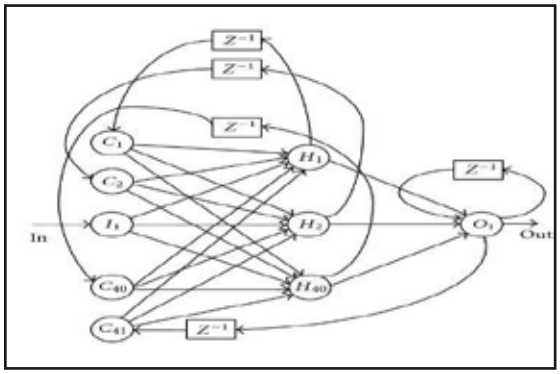

Fig. 2 : Diagram of RRN

V. Conclusions

We considered previous Bitcoin transaction in which price and timestamps are the attributes used to predict the bitcoin price for future. We used four methods for price predictions such as Logistic regression, Support vector machine, RNN and ARIMA. Prediction accuracies for these four methods are listed in the Table 2. Among the four methods, ARIMA performs well for next days predictions but performs poor for longer terms like given last few days price predict next 5-7 days prices. RNN perform consistently upto 6 days. Logistic regression-based model's assumptions were not violated, it is only able to classify accurately if a separable hyperplane exists.

Table 2 : Bitcoin Price Change Predictor Accuracies

Method	Accuracy
Logistic Regression	47%
SVM	48%
ARIMA	53%
RNN	50%

References

[1] *Charles Bovaird. 2017. Why the crypto market has appreciated more than 1,200% this year. Forbes Magazine [Online; posted 17- November-2017]. https://www.forbes.com/sites/cbovaird/2017/11/17/whythe-crypto- market-has-appreciated-more-than1200-this-year.*
[2] *Marshall Van Alstyne. 2014. Why bitcoin has value. Communications of the ACM 57(5):30–32. https://doi.org/10.1145/2594288.*
[3] *Kareem Hegazy and Samuel Mumford. 2016. Comparitive automated bitcoin trading strategies. http://cs229.stanford.edu/proj2016/report/MumfordHegazyCompar*

itiveAutomatedBitcoinTradingStrategiesreport.pdf.

[4] Satoshi Nakamoto.2008. Bitcoin: A peer-to-peer electronic cash system.

[5] Lipo Wang and Edwin Sin School of Physical and Mathematical Sciences Nanyang Technological University Singapore presented ―Bitcoin Price Prediction Using Ensembles of Neural Networks‖

[6] Neha mangla,‖Unstructured Data Analysis And Processing Using Big Data Tool-Hive And Machine Learning Algorithm-Linear Regression‖,International Journal of Computer Engineering & Technology (IJCET) Volume 9, Issue 2, March-April 2018, pp. 61–73, Article ID: IJCET_09_02_006

[7] Neha Mangla,‖Machine Learning Approach for Unstructured Data Using Hive,‖International Journal of Engineering Research‖, Volume 5,Issue 4, Pages 790-991, 2016.