International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 05 | May 2021 www.irjet.net p-ISSN: 2395-0072

Bitcoin Price Prediction using Machine Learning

Mr. Shivam Pandey[1], Mr.Anil Chavan[2], Miss. Dhanashree Paraskar[3], Prof. Sareen Deore[4]

[1-2]Student, Dept. Information and Technology, A. C. Patil college of engineering, Maharashtra, India
[4]Associate Professor, Dept. Information and Technology, A. C. Patil college of engineering, Maharashtra, India

---

Abstract - *In this paper, we attempt to predict the Bitcoin price accurately taking into consideration various parameters that affect the Bitcoin value. For the first phase of our survey, we aim to understand and identify daily trends in the Bitcoin market while gaining insight into optimal features surrounding Bitcoin price. For the second phase of our survey, using the available information, we will predict the sign of the daily price change with highest possible accuracy.*

Key Words: Arima model, Bitcoin, Bitcoin prediction, Blockchain, crypto currency, Linear regression, Random Forest regressor,machine learning.

1 .INTRODUCTION

A Bitcoin:

Bitcoin is a crypto currency which is used worldwide for digital payment or simply for investment purposes. Bitcoin is decentralized i.e. it is not owned by anyone. Transactions made by Bitcoins are easy as they are not tied to any country. Investment can be done through various marketplaces known as "bitcoin exchanges". These allow people to sell/buy Bitcoins using different currencies. The largest Bitcoin exchange is Mt Gox. Bitcoins are stored in a digital wallet which is basically like a virtual bank account. The record of all the transactions, the timestamp data is stored in a place called Blockchain. Each record in a blockchain is called a block. Each block contains a pointer to a previous block of data. The data on blockchain is encrypted. During transactions the user's

Sample paragraph, The entire document should be in name is not revealed, but only their wallet ID is made public. At present, the prices of these cryptocurrencies do not have a significant amount of studies and research as compared to traditional trading markets. However, the number of studies is steadily increasing as the popularity of Bitcoin is surging.

B Prediction:

The Bitcoin's value varies just like a stock albeit differently. There are a number of algorithms used on stock market data for price prediction. However, the parameters affecting Bitcoin are different. Therefore it is necessary to predict 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 unlike the stock market. Thus, to predict the value we feel it is necessary to

1.1 LITERATURE SURVEY

Bitcoin is a new technology hence currently there are few price prediction models available. [1] deals with daily time series data, 10-minute and 10-second time-interval data. They have created three time series data sets for 30, 60 and 120 minutes followed by performing GLM/Random Forest on the datasets which produces three linear models. These three models are linearly combined to predict the price of Bitcoin.

According to [2] the author is analysing what has been done to predict the U.S. stock market. The conclusion of his work is the mean square error of the prediction network was as large as the standard deviation of the excess return.evidence that several basic financial and economic factors have predictive power for the market excess return.

In [3], instead of directly forecasting the future price of the stock, the authors predict the trend of the stock. The trend can be considered as a pattern. They perform both short term predictions (day or week predictions) and also long-term predictions (months). They found that the latter produced better results with 79% accuracy. Another interesting approach the paper reflects is the performance evaluation criteria of the network. Based on the predicted output the performance evaluation algorithm decides to either buy, sell or hold the stock.

From[4],a comparison between Multi-Layer Perceptron (MLP) and Non-linear autoregressive exogenous (NARX) models is made. They conclude that MLP can also be used for stock market prediction.

1.2 PROBLEM DEFINITION

Cryptocurrencies such as Bitcoin, Ethereum, etc. generated significant attention in 2017. Cryptocurrencies have significant volatility as there is rampant speculation. Given the high variance in prices, can data science methods be used to model the market dynamics?

There are many directions this project could take. **Trading Strategy** Can an effective trading strategy be found? We are looking for a demonstration of sound data science principles here.

Market Analysis Given there is now option trading on certain cryptocurrencies, is it possible to create a volatility index for cryptocurrencies such as VIX? Is variance of this market infinite and therefore not predictable? Are there any rational reasons for investing that you can justify using data science?

Arbitrage Given the number of different currencies and different markets, how efficient is the market? Are there arbitrage opportunities? Can evidence be found of arbitrage?

2 PROPOSED SYSTEM

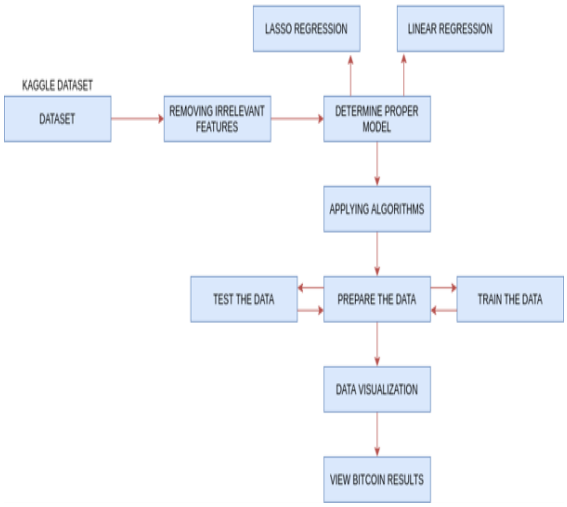
A .AutoRegressive model

Autoregressive models operate under the premise that past values have an effect on current values, which makes the statistical technique popular for analyzing nature, economics, and other processes that vary over time. Multiple regression models forecast a variable using a linear combination of predictors, whereas autoregressive models use a combination of past values of the variable.

B Linear Regression

Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data. One variable is considered to be an explanatory variable, and the other is considered to be a dependent variable. For example, a modeler might want to relate the weights of individuals to their heights using a linear regression model.

2.1 WORKFLOW



2.2 IMPLEMENTATAION METHODS

1. The first thing we have to do is retrieve the historical data of Bitcoin which can be downloaded as a convenient CSV file from Yahoo Finance. Once we have that, we can begin by formatting the CSV file as a Pandas DataFrame. Then, we use that same DataFrame for the rest of our plotting and calculations.

2. Prepare the data for modeling by making the data stationary. We do this by simply differencing the data and testing for stationarity by using something called the Dickey-Fuller test. We are aiming for a P-Value of less than the critical value of 5%, or simply trying to get as close to zero as possible. For even a lower P-value, we'll take the log of the prices, then difference the log instead of just differencing the prices.

3. Next, we'll have to plot the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF). Since we are working with daily data, the ACF shows us which day in the past correlates the most with the current day with respect to the days in between. PACF shows us which day in the past correlates directly to the current day by ignoring the days in between.

4. By knowing the PACF and ACF, we now better understand our dataset and the parameters to potentially choose. Now, we can move on to modeling our data by using the SARIMA model.

5. In order to get the best performance out of the model, we must find the optimum parameters. We do this by trying many different combinations of the parameters and selecting the one with the relatively lowest AIC score. Don't worry, we wrote a function that will do this for us.

6. Depending on your computer, the process of finding the best parameters may take awhile. For some like us, we'll have to settle for the best parameters limited by our computer's specifications. Unfortunately, not all computers are equal and some models will perform better based on the computer that is running them.

7. Now that we have our parameters, let's go ahead and train and fit the model to Bitcoin's prices.

8. Now we can get to the part that we really want to know about — Predicting Bitcoin's future prices! We do this by forecasting from the present day and seeing where it might go in the future.

2.3 RESULT AND DISCUSSION

The first step towards Bitcoin prediction is database collection. For our paper we have collected database from the following sources:

A Kaggle Kaggle, a subsidiary of Google LLC, is an online community of data scientists and machine learning practitioners. Kaggle allows users to find and publish data sets, explore and build models in a web-based data-science environment, work with other data scientists and machine learning engineers, and enter competitions to solve data science challenges.

Kaggle got its start in 2010 by offering machine learning competitions and now also offers a public data platform, a cloud-based workbench for data science, and Artificial Intelligence education. Its key personnel were Anthony Goldbloom and Jeremy Howard. Nicholas Gruen was the founding chair succeeded by Max Levchin. Equity was raised in 2011 valuing the company at $25 million. On 8 March 2017, Google announced that they were acquiring Kaggle.

The next step is database normalization. We basically perform this step to achieve consistency i.e. reduce or eliminate duplicate data, insignificant points and other redundancies. For normalizing our data, we have used five different techniques.

A.1 Box Cox transformation: A Box Cox transformation is a transformation of non-normal dependent variables into a normal shape. Normality is an important assumption for many statistical techniques; if your data isn't normal,applying a Box-Cox means that you are able to run a broader number of tests. : The function used is:-

data(λ)=(data^λ−1)/λ ....λ is not = 0

data(λ)=log(data) ...λ is = 0

A.2 Stat model : A statistical model is mathematicalmodel that embodies a set of statistical assumptions concerning the generation of sample data (and similar data from a larger population). A statistical model represents, often in considerably idealized form, the data-generating process.Here are the results obtained after implementing various normalization techniques:

Resampling:-Resampling is the method that consists of drawing repeated samples from the original data samples. The method of Resampling is a nonparametric method of statistical inference.

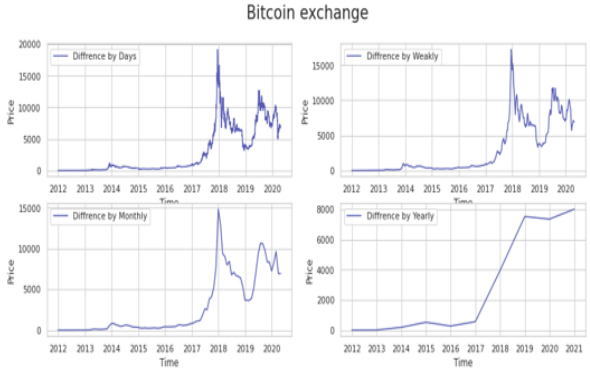


fig 2.3.1 Result of resampling

Histogram:- A histogram is a graphical display of data using bars of different heights. In a histogram, each bar groups numbers into ranges. Taller bars show that more data falls in that range. A histogram displays the shape and spread of continuous sample data.

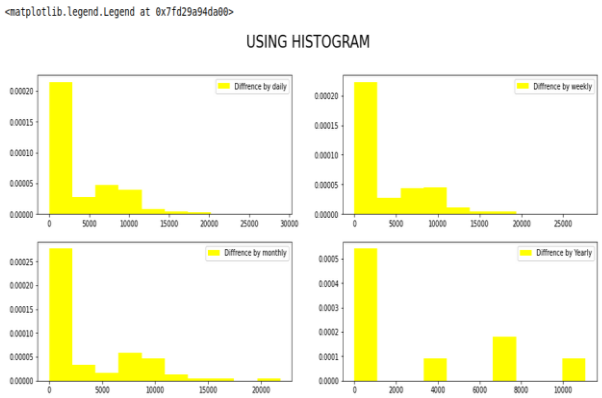


Fig 2.3.2 Result of histogram

Bar:- A bar chart or bar graph is a chart or graph that presents categorical data with rectangular bars with heights or lengths proportional to the values that they represent. The bars can be plotted vertically or horizontally.

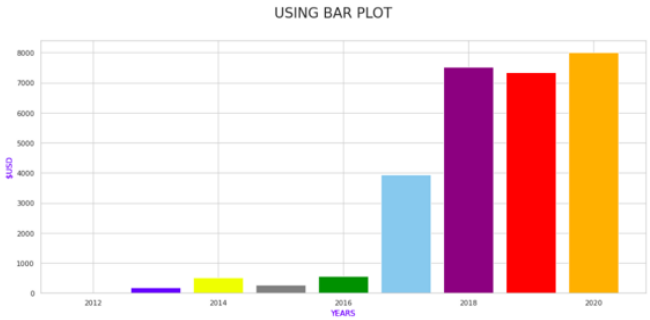


Fig 2.3.3 Result after bar plot

Stationarity:-stationarity tests such as the KPSS test that consider as null hypothesis H0 that the series is stationary, and unit root test.

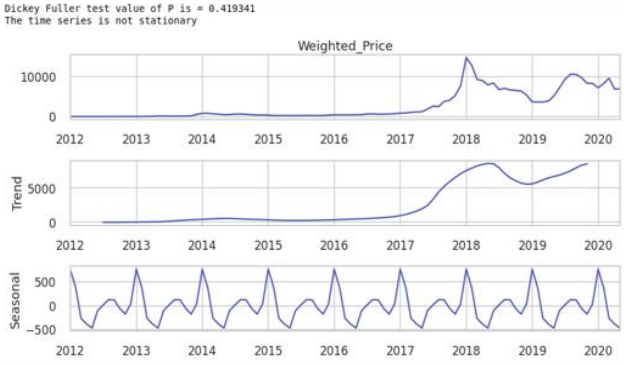


Fig 2.3.4 Result after stationarity

Box-Cox transformation:- The Box-Cox transform is a configurable data transform method that supports both square root and log transform, as well as a suite of related transforms.

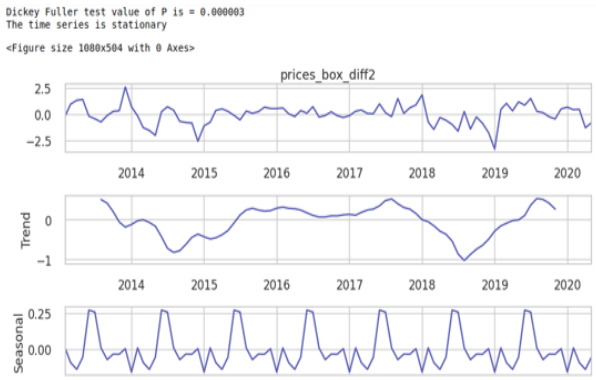


Fig 2.3.5 Result of Correlation

Final Prediction:-

Fig. 2.3.6 Result of Prediction

3. CONCLUSION

Predicting the future will always be on the top of the list of uses for machine learning algorithms. Here in this project we have attempted to predict the prices of Bitcoins using two deep learning methodologies. This work focuses on the development of project based learning in the field of computer science engineering, by taking into account the problem definition, progression, student assessment and use of hands on activities based on use of learning algorithm to develop application.

4. REFERENCES

[1] Siddhi Velankar, Sakshi Valecha , Shreya Maji, "Bitcoin Price Prediction using Machine Learning", International Conference onAdvanced Communications Technology(ICACT) , Dept of Electronics and Telecommunication, Pune Institute of Computer Technology, Pune Maharashtra, India,February 2018.

[2] Issac Madan, Shaurya Saluja, Aojia Zhao,"Automated Bitcoin Trading via Machine Learning Algorithms",Department of Computer Science , Stanford University, Stanford , 2015.

[3] Brian Vockathaler, " The Bitcoin Boom: An In Depth Analysis of The Price Of Bitcoins", Thesis, University Of Ottawa, Ontario, Canada, June 2017

[4] Huisu Jang and Jaewook Lee, "An Empirical Study on Modelling and Prediction of Bitcoin Prices with Bayesian Neural Networks based on Blockchain Information," in IEEE Early Access Articles, 2017.

[5] F. Andrade de Oliveira, L. Enrique ZÃ¡rate and M. de Azevedo Reis; C. Neri Nobre, "The use of artificial neural networks in the analysis and prediction of stock prices," in IEEE International Conference on Systems, Man, and Cybernetics, 2011.

[6] M. Daniela and A. BUTOI, "Data mining on Romanian stock market using neural networks for price prediction". informatica Economica.