IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 11, November 2017

Analysis of Bitcoin using Linear Regression and Data Mining Techniques

Abhyudit Bisht[1], Puru Agarwal[2]

Department of Computer Science and Engineering, SRM University Chennai India[1]

Department of Computer Science and Engineering, SRM University Chennai India[2]

Abstract: Cryptocurrency also referred as virtual currency is a digital asset which is used as a medium of exchange using high level cryptography. After the introduction of Bitcoin, hundreds of alternate cryptocurrencies started emerging and so did the need to analyse them. As of September 2017, there were over 1100 digital currencies in existence. We investigate the use of data mining techniques in analysing and predicting attributes like Volume, Market Cap. The cryptocurrency has many parameters like Open, High, Low, Close whose values are highly volatile and change every day. On available data set collected between 2013 and 2017 we perform Machine Learning algorithm and predict the values for Market Cap and various other attributes. This paper develops a model using linear regression to predict the future values which can be useful in many ways in determining the market behaviour.

Keywords: cryptocurrency, Bitcoin.

I. INTRODUCTION

The emergence of cryptocurrency might not be a new phenomenon but during recent years there have been huge rise in the valuation of these cryptocurrencies. The entire cryptocurrency system collectively is responsible for the production of the decentralized cryptocurrency. These cryptocurrencies are designed to gradually decrease the production of money. This can be done by placing an ultimate cap on the total amount of money in circulation. With the use of cryptocurrency the third parties like Lawyers, Notary can be eliminated helping in immediate settlement.

There are two methods to predict price development
1. Fundamental Analysis: This approach examines the the underlying forces of an economy. Factors like government regulations and media influence play a major role in deciding the cryptocurrency prices.

2. Technical Analysis: This methodology is used for forecasting the direction of prices through the study of past market data. The historical data that has been collected for many years is used as a training set for prediction.

II. LITERATURE SURVEY

Literature survey plays a very important role in the project development. It represents a study of previously existing material on the topic of the report. These may include existing theories about the topic ,research done , challenges being faced and ongoing work. Literature survey also helps in following the best practices in project development and understanding the risk and feasibility of the project. If the resources, time and money are not available
for the project development the risk is higher. Literature survey also gives light on various tools, platforms and operating systems suitable for project development and research.

1. In [1] the author explains how the cost of mediation increases the transation cost and how it limits the minimum practical transation size. The author also lays emphasis on the need for an electronic payment system based on cryptographic proof instead of trust and allowing any two willing parties to transact directly with each other without the need for a trusted third party.
2. In [2] the book the author discusses techniques such as least squares method which helps in performing mathematical regression analysis that finds the line of best fit for the dataset. This also provides a visual demonstration of the relationship between the data points.
3. In [3] the author has introduced a wide variety of models to perform data analysis. As the dataset consists of various factors contributing to change in one attribute over a period of time, we find it best to choose linear and multiple regression for data analysis.
4. In [4] the author has mentioned the two approaches to predicting price development, this dataset provides information only for technical analysis as fundamental analysis examines the underlying forces of the economy.
5. The weka tool provided by the University of Waikato [5] has been used to develop models by using classify and associate functions and further visualizing them to get a better understanding of the data.

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 11, November 2017

III. WORK FLOW DIAGRAM

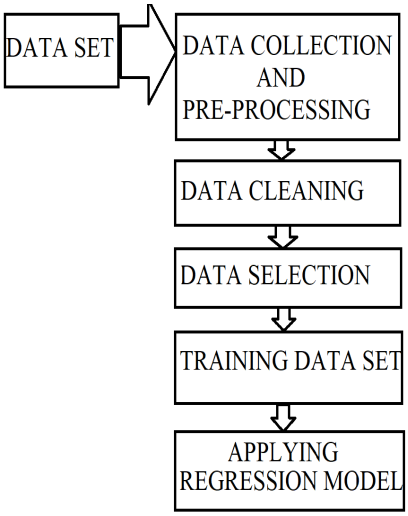

Fig 1. Work Flow Diagram

In this approach the paper is completed in five stages. Data collection and pre-processing, data cleaning, data selection, training data set and applying linear regression model. The responsible parameters are considered taken from the kaggle datasets. These data values are collected from April 2013 to October 2017 and then perform the linear and multiple regression to generate a regression model to predict the future values.

A. Data collection and pre-processing

The data used for this work was collected from kaggle which provides free access to various datasets. In this approach paper only the bitcoin values are collected from the historical dataset.

B. Data cleaning

The data cleaning phase includes converting the values in integers and removing symbols like commas, that will cause trouble in analysis of the data.

C. Data selection

The data is selected by taking into account any missing values or values of incorrect type.

D. Training dataset

The training method adopted for this research was percentage split that train on a percentage of the dataset, cross validate on it and test on the remaining percentage.

E. Applying regression model

Techniques like linear and multiple regression are applied to prepare a regression model using which the future values for the attributes will be predicted.

IV. DATA SETS

Data mining is an accumulation of utilized procedures that examine information from different perspective and various point of view. There are numerous organizations that provide free access to large database and information on cryptocurrency.

	Date	Open	High	Low	Close	Volume	Market.Cap
1	3-Oct-17	4408.46	4432.47	4258.89	4317.48	1288020000	73181300000
2	2-Oct-17	4395.81	4470.23	4377.46	4409.32	1431730000	72963200000
3	1-Oct-17	4341.05	4403.74	4269.81	4403.74	1208210000	72047300000
4	30-Sep-17	4166.11	4358.43	4160.86	4338.71	1207450000	69136600000
5	29-Sep-17	4171.62	4214.63	4039.29	4163.07	1367050000	69219200000
6	28-Sep-17	4197.13	4279.31	4109.70	4174.73	1712320000	69633200000
7	27-Sep-17	3892.94	4210.05	3884.82	4200.67	1686880000	64579200000
8	26-Sep-17	3928.41	3969.89	3869.90	3892.35	1043740000	65161000000
9	25-Sep-17	3681.58	3950.25	3681.58	3926.07	1374210000	61061100000
10	24-Sep-17	3796.15	3796.15	3666.90	3682.84	768015000	62954300000
11	23-Sep-17	3629.92	3819.21	3594.58	3792.40	928114000	60190000000
12	22-Sep-17	3628.02	3758.27	3553.53	3630.70	1194830000	60152300000
13	21-Sep-17	3901.47	3916.42	3613.63	3631.04	1411480000	64677600000

Showing 1 to 13 of 1,620 entries

Fig 3. Attributes of bitcoin data set

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 11, November 2017

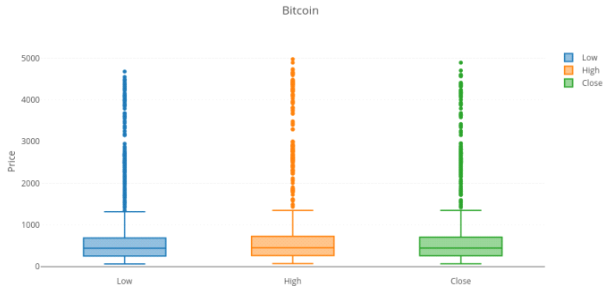

Fig 4. Box plot of bitcoin data set

V. LINEAR REGRESSION

Linear regression model is developed using the data and correlation between the data is analysed. The shows a R-square factor of 0.40 . The model is used to calculate the Y intercept and slope of the trend line which can be used to determine future values as well.

Fig 5. Linear Regression model between Market cap. and date

VI. PROPOSED WORK

The presented work takes into account only few attributes that were present in the given dataset, but there are various other attributes and factors that determine the price and eventually the trend the bitcoin will follow. These factors may be news announcements , government and central bank statements which has shown to cause a corresponsing upward or downward movement in price. In future these factors can also be taken into consideration for predicting the changes that might occur in the future.

VII. CONCLUSION

In this paper we used data mining techniques and regression algorithms for predicting the parameters like Volume , Market cap. using the crypto-currency historical data. The data used from Kaggle which consisted information from April 2013 to October 2017. Given enough data the observed trend over time could be studied and important patterns in change in these parameters for bitcoin or any other crypto-currency can be identified. This work is important for analysis of crypto-currencies because the variation in these attributes which affect the price of bitcoin can be studied using these data mining techniques.

REFERENCES

1. Nakamoto, S. : Bitcoin: A Peer-to-Peer Electronic Cash System (2008) Google Scholar
2. Mosteller F; Tukey JW : Data analysis and regression: a second course in statistics
3. Data analysis using regression and multilevel/hierarchical models, by Gelman, A., & Hill, J
4. https://www.coindesk.com/information/understanding-bitcoin-price-charts/
5. https://www.cs.waikato.ac.nz/ml/weka/
6. Regression and interpretation low R-squared, Kenshi Itaoka
7. Excel-Data-Analysis-for-Dummies, Stephen L.Nelson, E.C.Nelson

BIOGRAPHIES

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 11, November 2017

Abhyudit Bisht currently pursuing B.Tech Computer Science and Engineering in SRM University. His research interests include Data Mining and Big Data.

Puru Agarwal currently pursuing B.Tech Computer Science and Engineering in SRM University. His research interests include Artificial Intelligence and Machine learning.