DATA MINING AND VISUALIZATION

Coursework

Topic:
"Forecasting a Futuristic cryptocurrency analysis in terms of Cost"

By
Kiran Babu Basnet (B00728243)
Joshna Devi Rakonda(B00619262)

Lecturer: Dr. Naeem Ramzan

UNIVERSITY OF THE
WEST *of* SCOTLAND

UWS

School of Engineering and Computing
University of the West of Scotland

April 26, 2022

ABSTRACT

Cryptocurrency is one of the popular digital currencies. The first cryptocurrency was Bitcoin, which is founded in 2009 and is most popular these days. Bitcoin gains trust in users by publicly disclosing the full creation and transaction history. In return, the transaction history faithfully records the whole spectrum of cryptocurrency user behaviours. This report analyzes and summarizes the existing trends on knowledge in the bitcoin's transactions using data mining tools and techniques. Specifically, we classify the existing data sets into three aspects, i.e., growth rate, time and future prediction linking, the analysis of collective data sets, and the study of behaviours of the data sets. For each aspect, we present the problems, by taking the data from 2015 AD to 2021 AD, Applying linear regression methodologies and visualizing the major findings obtained in various forms of graphical representation like scatter plot , bar diagram, histograms. The time-series charts are plotted using ggplot2 – R programming library for graphing plots. This model can predict what would be the nature of the Bitcoins market in the future. The Mean Absolute Error calculated between the actual and predicted values is been seen as the uncertainty quantification method which has been used for opinion mining. Furthermore, an enumeration of transaction data parsing and visualization tools for future prediction is shown. Finally, we outline the several gaps and trends for future investigation in this research area.

1. INTRODUCTION

1.1 Background

As of 2020, more than 7000 cryptocurrencies are actively trading in more than 20000 online exchanges. Their total market capitalization has exceeded USD 300 billion [1]. Although these cryptocurrencies are not backed by any tangible assets, they gain trust from users by publicly

1.3 Problem

This is a kind of transitional time for the entire world economically i.e., the transition from our conventional monetary system to new digital, technology based crypto economic demography. It has only the history of a decade, but its impact is like a history of thousand's year. Cryptocurrency and its parent technology viz Blockchain has started in 2008 AD and introduced its first currency in 2009 AD viz Bitcoin is considered as Blockchain 1.0 after that its gradual development and optimization found that will become the hot ball in the hand of today's world. In the specific circumstances we took consider the crypto history from 2015 AD to 2021 AD, so that we will able to figure out the each ups and down to the crypto world and also help to visualize the upcoming days for the world's digital economy.

1.4 General Interest

Though the cryptocurrency has not been fully understand by all the people, Only a handful of people has somehow deep understanding of it but everyone has interest to know about it and it is becoming a way of life in future for every citizen. The general understanding of crypto, it's working

disclosing the full creation and transaction history in peer-to-peer blockchain networks. Each transaction in the blockchain consists of transferring a virtual value from a virtual identity, i.e., a blockchain address or a set of addresses, to another. Among these variety of cryptos Bitcoin is a market leading and known to every individual. The sizes of transaction records are quickly expanding. The total transaction volumes of Bitcoin is highest at the middle of 2021. Although technically challenging in extracting, transforming, and analyzing, these transaction histories have given us an unprecedented opportunity to study the panorama of human behaviour in a complex economic environment. Reid and Harrigan [4] conducted the first study on the entire cryptocurrency transaction history (up to mid-2021), revealed emerging structure from the Bitcoin flow network, and demonstrated the transaction history and being a futuristic technological behaviours data mining from cryptocurrency transactions has emphasized into a large body of research and been successfully applied in assisting multiple law enforcement actions, including ceasing the then-largest darknet market Silk Road in late era [5] and arresting suspects in a major theft from the then-largest cryptocurrency exchange.

1.2 Motivation

The main motive of the generation of this report is to study of crypto currency that will grow in future. To know why cryptocurrency is being so popular and to know which cryptocurrency is growing and how are the people benefitted using cryptocurrencies and to know how people are predicting the future and investing. Direct access to the data is going to help, So going through the methodical study and draw a meaningful conclusion based on data. Data mining and its tool are not only the theoretical approach, but it is a proven facts with mathematical proof, It is not only the probability, but its actuality.

, impact , popularity, flexibility and many other unseen dimension on futuristic aspect in multi-disciplinary view.

1.5 Method of Solving

It is said that "Human mind can only be able to do visualization" Our mind is not for any sorts of theoretical letter and descriptive data. Performing the Data Pre-processing like importing the required libraries and loading them, importing the csv files to analyze the data frames, Handling the missing data, Visualizing the data, finding the relationship between the variables, splitting it into train and test data for a model. The graphical depiction of information and data is known as data visualization. Data visualization tools make it easy to examine and comprehend trends, outliers, and patterns in data by employing visual components like charts, graphs, and maps. Data visualization tools and technologies are critical in the Big Data environment for analyzing enormous volumes of data and making data-driven choices.
Professionals' ability to utilize data to make choices and use graphics to communicate stories about when data informs the who, what, when, where, and how is becoming increasingly valuable. While conventional schooling often creates a distinction between creative storytelling and

technical analysis, the new professional world encourages people who can bridge the gap: data visualization stands during analysis and visual storytelling.

Before making a final story, every parameter should properly cleaned so we follow the process of data cleaning , present it into the best format and visualize it and apply the regression more specifically linear regression so see which method suits our data sets.

For the best use case of available data sets we try to visualize data in every possible aspect

- Bar Chart
- Box-and-whisker Plots
- Bubble Cloud
- Bullet Graph
- Cartogram
- Circle View
- Dot Distribution Map
- Gantt Chart
- Heat Map
- Highlight Table
- Histogram
- Matrix
- Network
- Polar Area

It is absolute that all method can't be implemented in single set of data due to its inferences if data field and parameter. So, for our data set we try to best apply qualitative mean method with scatters plot and bar plot.

Different time-series methods such as the Exponential Weighted Averages, Generalized Autoregressive Conditional Heteroskedastic Model, Structural Time Series models were used to observe the future of bitcoin. Various uncertainty quantification methods such as Mean Absolute Error and Root Mean Square Error were used to determine the accuracy of the predictions.

2. Data Sets

As this is the comparative study among various crypto coin such as Aave, BinanceCoin, Bitcoin,Cardano,ChainLink,Cosmos,CriptocomCoin,Dog ecoin,Eos,rium,Iota,Litecoin,Monero,NEM,Polkadot,Sol

Fig. 1 Data Set for Ethereum

All the formatted data set has mainly six calculative, alphanumeric value under consideration. which are listed below
- Highest value
- Lowest value
- Opening value
- Closing value

Common general types of data visualization:
- Charts
- Tables
- Graphs
- Maps
- Infographics

There are many specific application of data visualization specific examples of methods to visualize data:
- Area Chart

- Radial Tree
- Scatter Plot (2D or 3D)
- Streamgraph
- Text Tables
- Timeline
- Tree map
- Wedge Stack Graph
- Word Cloud
- And any mix-and-match combination
 in a dashboard!

ana,Stellar,Tether,Tron,Uniswap,USDCoin,WrappedBitc oin and XRP

All the data set are in similar in structure and their respective fields

Fig. 2 Data Set for Bitcoin

- Volume
- Market Capitalization

2.1 Data Pre-Processing

In this step, we've imported libraries, datasets, handled with missing values, removed unwanted columns. Proposed data set are being considered as valid and some preliminary pre-processing like eliminating null and

unassigned values and rounding the largest value by using i.na() to know the missing values in dataset.

2.2 Primary Working Interest

The primary working interest for the real data sets is to find out the best result based on possible available data sets and apply the mining technique for visualize and forecast for the future.

2.3 Visualizing the Output Array

After training the models over layers of predictions for the closing prices of the successive days are made. An analysis of how this validates with real time data is done. Certain Uncertainty Quantification methods such as Root Mean Square Error and Mean Absolute Error are visualized as heat maps.

A visualization of the data obtained from Twitter for this purpose is carried out. Certain graphs such as line charts plotted for Sentiment versus Time Frame, Bar Charts showing analysis of popularity of Bitcoin or Litecoin can be used as measures of validations. For the purpose of visualizations, we have used ggplot2 - a visualization package.

2.4 Architecture Design

We try to follow the standard practice of data mining and visualization technique. The there are many methodologies to analyze the data but all the necessary step of collecting data source based on secondary real time data set via reliable data source followed by data cleaning. which involves rounding the decimal value by using floor and ceiling function, eliminating the unassigned value and deleting the unwanted data. The major step is data analysis which clustering, classification, outlier detection and association analysis. The final stage is to apply the result based o prevention application , prevention strategies and prediction.
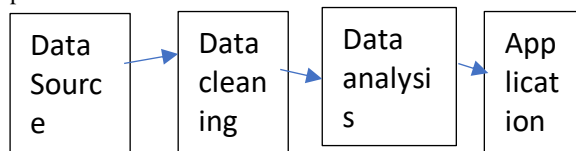

Fig.3 Schema Diagram of Data Mining

2.5 Data Models

In cryptocurrency and blockchain design, Bitcoin and Ethereum are the two milestones, i.e. blockchain 1.0 and 2.0. The transaction record data structures were adopted and updated by the majority of cryptocurrencies [10]. Unspent transaction output (UTXO)-based data models are used by Bitcoin and its derivatives, whereas account-based data models are used by Ethereum and its derivatives.

2.6 Exploratory Data Analysis

2.6.1 Raw Data
We consider the secondary source data set as authentic one and real time data in the precise subscription being given by

https://www.kaggle.com/datasets/sudalairajkumar/crypto currencypricehistory

2.7 Statistical Analysis

We've used describe function to perform the statistical functions by installing psych package.

Fig.3 shows the mean, median, sd, variance of the dataset
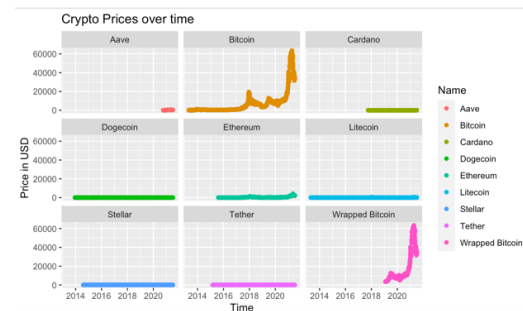
Fig.4 Scatterplot showing the Cryptocurrency prices over time

We need to remove unneeded variables from the model to minimize data complexity and dimensionality. To avoid over-fitting, we use a training and validation set to analyses the combination of features. We begin by choosing a feature and training it with the training set. The validation feature set is then used as input for the trained model. And for casting process ins seen in the following figure.
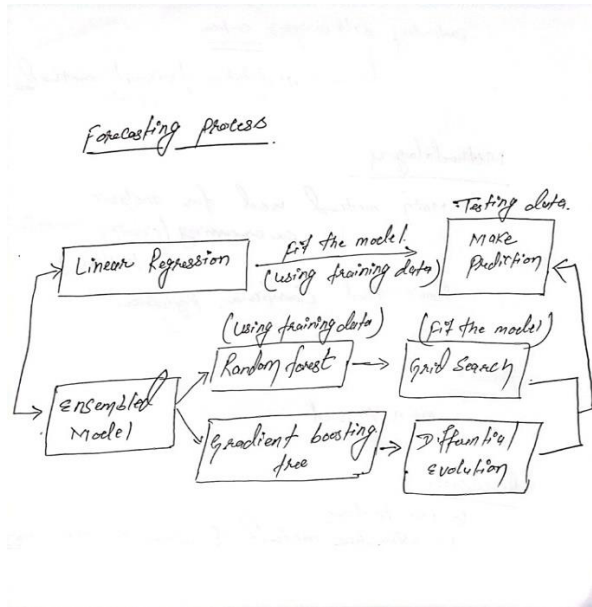
Fig.5 Forecasting process

Result

The linear regression is performed on predicted and actual models to know the bitcoin prediction. We've used prophet functions to forecast the data.

Fig.6 shows prediction of April 29th data using the models.
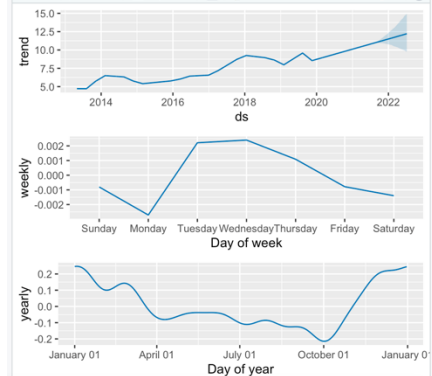
Fig.7 shows the yearly, weekly and trends of bitcoin data
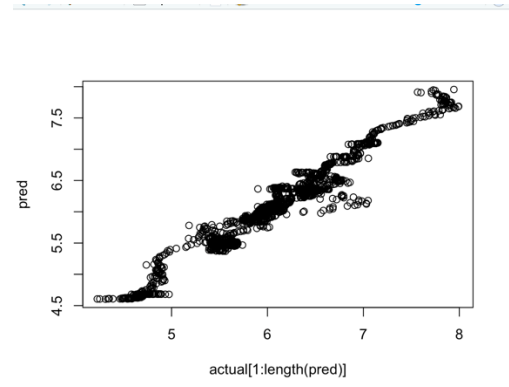
Fig.8 shows the scatterplot of predicted vs actual bitcoin data with linear regression

Discussion / Future Improvement

We still have a lot of ideas to finish and enhance due to time constraints. The following is a list of upcoming improvements:

➢ In the future we plan to apply more advanced feature selection technique known as autoencoder
➢ We'd want to employ deep learning networks like Long-Short Term Memory (LSTM) and Generative Adversarial Network (GAN) for more efficient and accurate forecasting because financial data is usually non-stationary and non-linear.
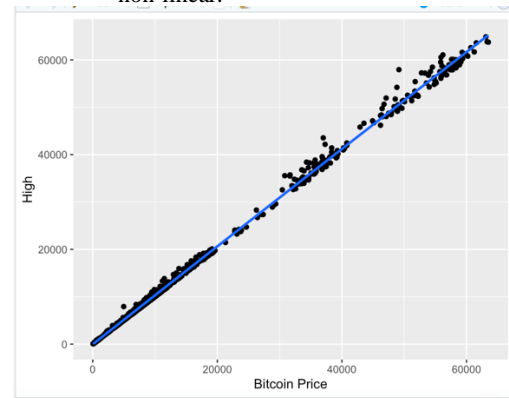
Fig.9 shows the plot when polynomial regression is applied

```
Console  Terminal ×  Jobs ×
R  R 4.1.2 · ~/
> #applying polynomial regression to check the changes
> best = lm(bitcoin$High ~ poly(bitcoin$Close,2, raw=T), data=bitcoin)
> summary(best)

Call:
lm(formula = bitcoin$High ~ poly(bitcoin$Close, 2, raw = T),
    data = bitcoin)

Residuals:
    Min     1Q  Median     3Q    Max
-1606.2 -105.7   22.3   35.2  7368.1

Coefficients:
                              Estimate Std. Error t value
(Intercept)                   -3.837e+01 1.003e+01 -3.826
poly(bitcoin$Close, 2, raw = T)1 1.036e+00 1.875e-03 552.840
poly(bitcoin$Close, 2, raw = T)2 -1.347e-07 3.707e-08 -3.634
                              Pr(>|t|)
(Intercept)                   0.000133 ***
poly(bitcoin$Close, 2, raw = T)1 < 2e-16 ***
poly(bitcoin$Close, 2, raw = T)2 0.000284 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 396.9 on 2988 degrees of freedom
Multiple R-squared: 0.9988,    Adjusted R-squared: 0.9988
F-statistic: 1.285e+06 on 2 and 2988 DF, p-value: < 2.2e-16
```

Fig.10 shows polynomial regression values

Conclusion

Since its inception in 2008, Bitcoin has gained widespread popularity among millions of users all around the world. Since then, academics has created a huge body of work on bitcoin transactions. Bitcoin, Ethereum, and Ripple are the most well studied cryptocurrencies. We conducted statistical analysis focusing on the BTC time-series data in this study between 2015 to 2021.applying various tools of data mining and present it visually majorly applying qualitative mean approach based on linear regression and the data can be predicted using the history and the bitcoin is been increasing due to the closeness values.

References:
[1] Coinmarketcap. Accessed: Nov. 1, 2020. [Online]. Available: https:// coinmarketcap.com/ [2] Blockchain Explorer. Accessed: Nov. 1, 2020. [Online]. Available: https://www.blockchain.com/explorer
[3] Ethereum Blockchain Explorer. Accessed: Nov. 1, 2020. [Online]. Available: https://etherscan.io/
[4] F. Reid and M. Harrigan, ''An analysis of anonymity in the bitcoin system,'' in Proc. IEEE 3rd Int. Conf. Privacy, Secur., Risk Trust IEEE 3rd Int. Conf. Social Comput., Oct. 2011, pp. 1318–1326.
[5] D. Ron and A. Shamir, ''How did dread pirate roberts acquire and protect his bitcoin wealth?'' in Proc. 18th Int. Conf. Financial Cryptogr. Data Secur. (FC). Berlin, Germany: Springer, 2014, pp. 3–15.
[6] K. Nilsson. (2017). Breaking Open the MtGox Case, Part 1. [Online]. Available: https://blog.wizsec.jp/2017/07/breaking-open-mtgox-1.html
[7] H. Kuzuno and C. Karam, ''Blockchain explorer: An analytical process and investigation environment for bitcoin,'' in Proc. APWG Symp. Electron. Crime Res. (eCrime), 2017, pp. 9–16.
[8] S. Naqvi, ''Challenges of cryptocurrencies forensics—A case study of investigating, evidencing and prosecuting organised cybercriminals,'' in Proc. 13th Int. Conf. Availability, Rel. Secur. (ARES), 2018, pp. 1–5.
[9] A. S. M. Irwin and G. Milad, ''The use of crypto-currencies in funding violent jihad,'' J. Money Laundering Control, vol. 19, no. 4, pp. 407–425, Oct. 2016
[10] Nakamoto, S. Bitcoin: A Peer-to-Peer Electronic Cash System. Available online: https://bitcoin.org/bitcoin.pdf (accessed on 13 April 2022).

ONLINE SOURCES
[O1] Aleth.io, EthStats.io, Accessed: April, 2021. [Online]. Available: https://ethstats.io/. [O2] Alethio, Accessed: April, 2021. [Online]. Available: https://aleth.io/.
[O3] BitBonkers, BitBonkers- A Bitcoin Blockchain Transaction Visualisation, Accessed: April, 2021. [Online]. Available: https://bitbonkers.com/.
[O4] Bitcoin Globe, Bitcoin Globe, Accessed: April, 2021. [Online]. Available: http://bitcoinglobe.com/.
[O5] BitcoinWisdom.com, BitcoinWisdom, Accessed: April, 2021. [Online]. Available: https://bitcoinwisdom.com/.
[O6] Bitfly gmbh, Etherchain, Accessed: April, 2021. [Online]. Available: https://www.etherchain.org/.
[O7] bitFlyer, chainFlyer, Accessed: April, 2021. [Online]. Available: https://chainflyer.bitflyer.jp/. [O8] BitForce5, BitForce5, Accessed: April, 2021. [Online]. Available: http://bitforce5.com/.