**Predicting the price of Bitcoin**

Data Analytics in Practice

Case Study 4

Group 8

Contents

[Executive Summary 1](#_Toc159591891)

[1.Introduction: 1](#_Toc159591892)

[1.1 Background 1](#_Toc159591893)

[1.2 Objectives and Hypothesis 2](#_Toc159591894)

[2.Data and Methodology 2](#_Toc159591895)

[2.1 Correlation Analysis 3](#_Toc159591896)

[3.Timeseries Analysis of Bitcoin and Bitcoin Halving: 3](#_Toc159591897)

[4.Estimation and Further Discussion 3](#_Toc159591898)

[4.1 Regression Analysis 3](#_Toc159591899)

[4.2 Random forest estimation 4](#_Toc159591900)

[4.3 LSTM estimation 4](#_Toc159591901)

[5.Limitation 4](#_Toc159591902)

[6.Conclusion 5](#_Toc159591903)

[References: i](#_Toc159591904)

[Appendices: A](#_Toc159591905)

[Appendix 1. Histograms A](#_Toc159591906)

[Appendix 2. Correlation Heatmap A](#_Toc159591907)

[Appendix 3. Timeseries Analysis of Bitcoin Price B](#_Toc159591908)

[Appendix 4. Regression Analysis test result B](#_Toc159591909)

[Appendix 5. Random Forest B](#_Toc159591910)

[Appendix 6. RSME score of LSTM Model 2 C](#_Toc159591911)

[Appendix 7. Predicted VS Actual Price of Bitcoin in LSTM Model 2 C](#_Toc159591912)

[Appendix 8. Graphical representation of Prediction and Actual Price of Bitcoin C](#_Toc159591913)

# Executive Summary

Bitcoin, once dismissed as a potential Ponzi scheme, has emerged as a frontrunner in response to the shortcomings of the traditional financial system. The recent recognition of Bitcoin and other cryptocurrencies as legal tender by countries such as El Salvador and the Central African Republic, along with surprising endorsements from institutional giants like BlackRock and Fidelity, has bolstered support for the decentralized finance sector. Although widespread adoption is still a work in progress, the rapid pace of acceptance and continual innovations underscore its significance.

Despite these advancements, the technology faces numerous challenges, especially concerning macroeconomic conditions and government embargoes. Notably, China and Central Asian States, previously pivotal for Bitcoin mining, imposed a ban, resulting in a significant decline in Bitcoin prices. However, other factors contribute to this fluctuation, as explored in this report. The study aims to forecast Bitcoin prices and identifies the impact of other cryptocurrencies like Ethereum and Litecoin, as well as traditional indices such as the NYSE and NASDAQ Composite.

Additionally, the report delves into the influence of mining difficulty on Bitcoin prices. Our analysis aligns with the observed trend where institutional investors, traditionally heavily invested in NYSE and NASDAQ, are turning their attention to Bitcoin. The high correlation between the two markets is reflected in inferential results. Lastly, mining difficulties, exemplified by recent events in China and Russia, further highlight their profound impact on Bitcoin prices.

# 1.Introduction:

## 1.1 Background

Money is characterised by three features namely store of value, unit of account and medium of exchange. However, the 2008 financial crisis revealed the inherent fragility of the traditional financial system and people raised doubts on the current form of fiat money which has no gold or any commodity backing. Hence in response to that bitcoin was created to enable a secure and borderless transactions without the need for banks or any other intermediaries. Nonetheless, people still raise doubts on bitcoin being a store of value, unit of account and medium of exchange primarily because of its volatile nature. The narrative however is changing and after El Salvador and Central African Republic many countries are in line to accept bitcoin as a legal tender. A wider adoption of such nature will decrease the volatility of bitcoin and other major cryptocurrencies. Moreover, the increased institutional embrace of Bitcoin has significantly contributed to the enhanced price stability of the cryptocurrency. Institutions such as BlackRock and Fidelity, along with other major investors, adopting BTC have exerted influence on significant financial markets. The establishment of Exchange Traded Funds (ETFs) for Bitcoin has further played a pivotal role in revolutionizing this trend. The motive of this report is to dig deeper in the price of bitcoin and predict what the future price would be. By combining a set of bitcoin specific, crypto specific, commodity, security and google trends variables we aim to predict the price of bitcoin.

After this introduction and a hypothesis, we will introduce the data and methodology in the next section. Following that we will move to our estimation and the final chapter will give a conclusion.

## 1.2 Objectives and Hypothesis

The report aims to predict the price of bitcoin by keeping commodity assets, traditional financial instrument indices, and some bitcoin related variables which can affect the price of bitcoin. Translating this into hypothesis we have formed a hypothesis below:

**H0:** None of the variables affect the price of bitcoin.

**H1:** At least one of the variables affect the price of bitcoin.

# **2.Data and Methodology**

The data of traditional financial instrument indices and gold is taken from Yahoo Finance whereas the data pertaining to bitcoin features and other cryptocurrencies is taken from Quandl. Google trends provided the data for bitcoin trends.

In addition, we have used daily data from January 2015 to 15th Feb 2024. Where the data was null we have used interpolation and forward fill especially in the case of NASDAQ, NYSE and gold data as bitcoin is traded 24/7 while the traditional commodity and financial instrument indices are traded on working days only. In the case of other crypto currencies like Litecoin and Ethereum the coins started trading from 2017 onwards so the prices for the year 2015 and 2016 of these has been kept null. Such data cleaning approaches might not be the best and is a limitation of our work.

Our variables are **date**, **BTC** (in USD), **Gold** (in USD), **Litecoin** (in USD), **Ethereum** (in USD), **BTC Volume**, difficulty of mining BTC (denoted by **difficulty\_BTC**), **BTC Popularity**, **BTC\_CASH** (in USD), **NYSE** (index value in USD), **NASDAQ** (Index value in USD).

Now referring to Appendix 1, we see the histograms of our variables and see some interesting observations. All the cryptocurrencies in our dataset have right skewed distribution. This is because of volatility, speculation and FOMO (fear of missing out). With such trading instruments especially the DeFi (decentralized finance) assets price spikes are common leading to a rightly skewed distribution. Added to this investor psychology play an important role and be it some news or development FOMO leads to price increases. If we compare this to traditional finance instruments, be it stocks or commodities the distribution is relatively normal. With this it is pertinent to assess what the correlation between these instruments would be.

## 2.1 Correlation Analysis

Appendix 2 sheds light on correlation between our variables and reveals a strong positive relationship between bitcoin, Litecoin, Ethereum and the stock indices of both NASDAQ and NYSE. This confirms our earlier observation in the introduction where we asserted that more institutional investors are attracted towards investing in bitcoin and have created exchange traded funds. The positive correlation of BTC and other crypto assets is natural as bitcoin is a leader in this arena and other crypto assets follows the investing trend of BTC. Interestingly there is a positive correlation in value of bitcoin and in difficulty in mining bitcoin and this adds to the notion that not just speculation but scarcity of BTC adds value to its price. Likewise, we see a positive correlation between BTC and its volume which indicates that when price of the coin increases so does its volume, again confirming our observation on FOMO.

# 3.Timeseries Analysis of Bitcoin and Bitcoin Halving:

Finally, before concluding this section see refer to Appendix 3 where we graphically analyse the price of bitcoin over time. We can see three spikes in the price of bitcoin over time and then a sudden downfall. Various factors contribute to this particularly global macroeconomic condition as bitcoin came into existence in response to the failure of global financial institutions in 2008. In addition, government interventions have a keen role in the price of bitcoin as post 2022 the prices dramatically fell as China and Central Asian states which were the hubs of bitcoin mining announced a complete ban on crypto mining. However, an event known as **bitcoin halving** alsoplaysa vital role in the sudden spikes of bitcoin. Halving is an event that occurs every four years or after 210,000 blocks are mined. During halving the reward that the miners get is cut in half so as to ensure scarcity and mimic the diminishing rate of production just like in gold mining. Till date there have been three bitcoin halvings andthe first in November 2012 led the price of BTC from $12 to $1000 a year later but since we are covering data from 2015 onwards this isn’t visible in our graph. The next event happened on July 9, 2016 and though the price dipped to $670 initially, it skyrocketed to $19,700 by the end of next year. Similarly, the halving of May 11, 2020 increased the price of bitcoin from $ 8787 to at an all time high of $69,000 by next year end.

Nevertheless, it is important to reassert that halving alone did not cause the price increase or decrease but market sentiments, government regulations, macroeconomic environment and institutional interest have a likewise important hand at play.

# 4.Estimation and Further Discussion

## 4.1 Regression Analysis

We had divided our data into test and train datasets and initially ran a regression analysis. Appendix 4 reveals the result of regression analysis where our R-Squared is 0.999 indicating an ideal result. However, our data was not normally distributed as we discussed earlier. Plus in time series a normal regression analysis may not give a true picture and R-squared result can be misleading unless the assumptions of regression analysis are met which pertains to unbiased estimators. Hence, we move to random forest estimation.

## 4.2 Random forest estimation

The rationale for using random forest estimation in a time series setting is to introduce lagged versions of our target variable BTC and capture the temporal dependencies. Although it requires adaptation to cater to autocorrelation and temporal dependencies that violate the independence assumption, feature engineering fixes these limitations and hence we see a lower RMSE than regression. This along with the coefficients is laid down in appendix 5. However, unlike a linear regression, random forest regression coefficients are non-linear, and we cannot ascertain that a $1 increase in price of bitcoin leads to a $2 increase in price of Ethereum or anything. Hence, we sum this with the result of a lower RMSE than regression and hop on to LSTM estimation which is by far the best estimation method for time series data and for stock price predictions.

## 4.3 LSTM estimation

Our data is sequential, i.e. time series and hence LSTM models are designed for capturing patterns and dependencies in sequential data. In our LSTM-based crypto price prediction, historical prices, trading volumes, and relevant indicators are organized sequentially for input. The model learns patterns during training to predict future price movements. After training, it uses new data to generate forecasts, refined based on actual market results.

We made different combinations of parameters and the model with lowest RMSE was 100.21(as shown in Appendix 6).

We compared the actual and the predicted price of Bitcoin after training the model and observed that there is slight difference (Appendix 7). The graphical representation helped in better visual depiction of our actual and predicted values (Appendix 8).

# 5.Limitation

The study aimed to predict Bitcoin prices using various factors from different areas. However, it's important to note that external factors, particularly those related to the broader economy, could also influence prices. Additionally, certain global events, such as the halving phenomenon, were found to have a significant impact, as highlighted in our analysis. Unfortunately, these factors couldn't be included in our models, representing a limitation in our study.

Furthermore, this was our first venture into working with time series data as part of our data analytics program. While some classmates had prior experience, the diverse elective courses each of us took meant that some individuals might have been more adept at modelling this report, while others may have faced challenges. The varying technical expertise and the use of different data types posed both advantages and limitations. Despite putting considerable effort into refining our data models, it's worth acknowledging that there's always room for improvement, and not employing a more advanced statistical model stands as a constraint in our approach.

# 6.Conclusion

Summing up, we conclude that LSTM model better predicted the price of bitcoin. The variables at play that better captured this was crypto assets like Ethereum and Litecoin and traditional asset class indices of NYSE and NASDAQ and further the difficulty in mining bitcoin. Future prediction studies should aim at modelling halving exercises and government interventions like bans on mining in China and Russia or in places where bitcoin mining is on a bigger scale. In addition, the adoption of bitcoin as legal tenders by some countries and the introduction of BTC exchange traded funds would also have had a profound impact on its price and such variables needs to be modelled for a better price prediction.

# References:

Nicholson-Messmer, E. and Greifeld, K. (2024) *Blackrock’s bitcoin- spot ETF (IBIT) is first to cross $1 billion threshold*, *Bloomberg.com*. Available at: https://www.bloomberg.com/news/articles/2024-01-18/blackrock-s-bitcoin-etf-is-first-to-cross-1-billion-threshold?leadSource=uverify+wall (Accessed: 20 February 2024).

*Bitcoin cash USD (BCH-USD) price history & historical data* (2024) *Yahoo! Finance*. Available at: https://finance.yahoo.com/quote/BCH-USD/history (Accessed: 17 February 2024).

*Ethereum USD (ETH-USD) price, value, news & history* (2024) *Yahoo! Finance*. Available at: https://finance.yahoo.com/quote/ETH-USD (Accessed: 17 February 2024).

*Litecoin USD (LTC-USD) price, value, news & history* (2024) *Yahoo! Finance*. Available at: https://finance.yahoo.com/quote/LTC-USD (Accessed: 17 February 2024).

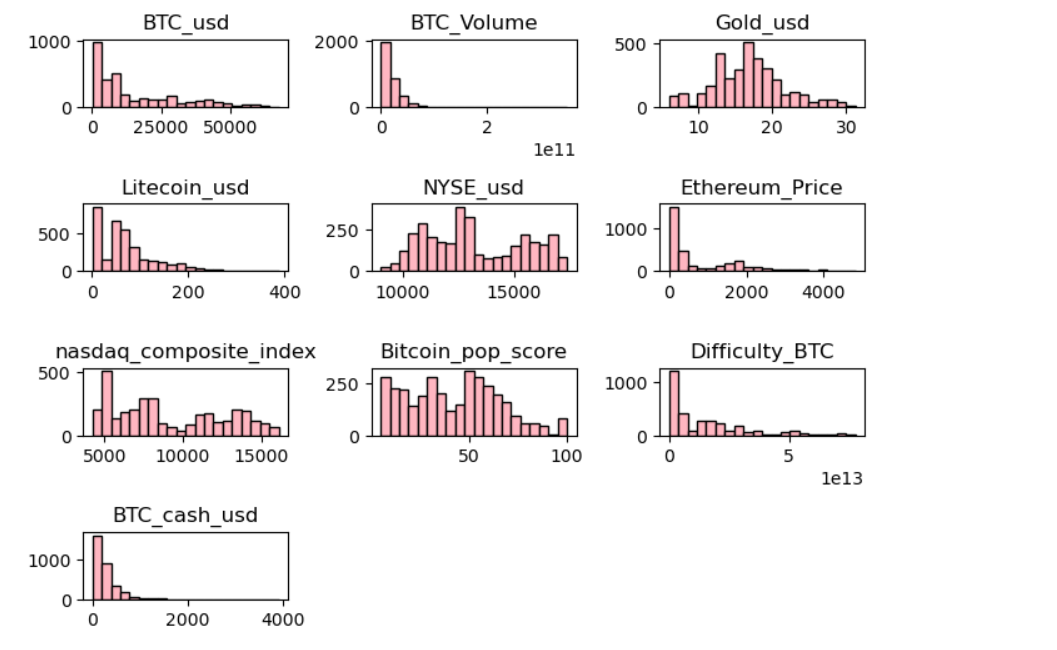
*Gold Apr 24 (GC=F) stock price, news, Quote & History* (2024) *Yahoo! Finance*. Available at: https://finance.yahoo.com/quote/GC%3DF (Accessed: 17 February 2024).

*Nasdaq composite (^IXIC) historical data* (2024) *Yahoo! Finance*. Available at: https://finance.yahoo.com/quote/%5EIXIC/history (Accessed: 17 February 2024).

*Nasdaq Data Link* (2024) *Financial, Economic and Alternative Data*. Available at: https://data.nasdaq.com/data/BCHAIN (Accessed: 17 February 2024).

# Appendices:

## Appendix 1. Histograms

****

## Appendix 2. Correlation Heatmap

**A blue and white grid with black text

Description automatically generated with medium confidence**

## Appendix 3. Timeseries Analysis of Bitcoin Price

**A graph of a graph

Description automatically generated with medium confidence**

## Appendix 4. Regression Analysis test result

**A number on a white background

Description automatically generated**

## 

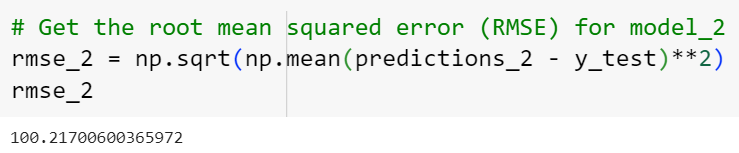
## Appendix 5. Random Forest

**A screenshot of a computer

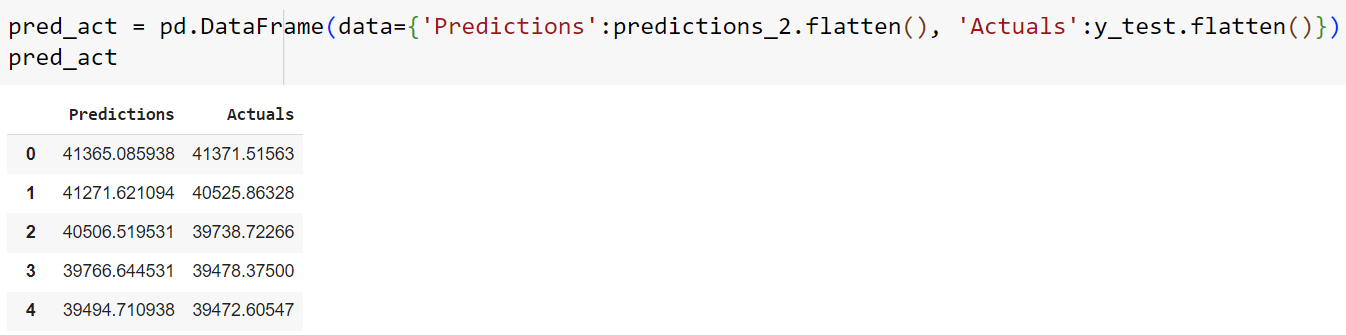
Description automatically generatedA black text on a white background

Description automatically generated**

## Appendix 6. RSME score of LSTM Model 2



## Appendix 7. Predicted VS Actual Price of Bitcoin in LSTM Model 2



## Appendix 8. Graphical representation of Prediction and Actual Price of Bitcoin

