Forecasting Bitcoin and Ethereum Crypto currencies

Using Multi-linear Regression, QDA, KNN, K fold, and LSTM models

Roja Reddy Sareddy and Manizheh Zand

Computer and Electrical Engineering Department
Santa Clara University

March 15, 2022

Summary

1. Why we should be concerned about crypto-currency
2. Research
3. Our Model Prediction
4. Conclusion

Why we should be concerned about crypto-currency

Santa Clara University

Bitcoin and Ethereum Cyptocurrencies

- Advantages
- Disadvantages
- Different types of Digital Currency
- Who should invest
- See the "EVIL"
- Models
- Data Source

Bitcoin

How and who started?

In 2008, Satoshi Nakamoto, Born: 5 April 1975 (age 46) (claimed); Japan (claimed), wrote an 8 pages white paper, A Peer-to-Peer Electronic Cash System [1]. In 2010, Bitcoin value was $0.0008 and by April 2013, it jumped to $250

Who is accepting bitcoin as a payment[2]?

- Microsoft, Wikipedia, Paypal, Starbucks
- AT & T, Overstock,Twitch, Amazon
- Home Depot, Whole Food, CheapAir,
- Newegg, Namecheap,Rakuten, KFC, Burger King

Values

If you invested $100 in 2010, by today, you would have $ 4.9 billions

Ethereum

Founder

Vitalik Buterin , 27 year old Russian, came up with the idea when he was 19.

Who is accepting Ethereum as a payment[3]?

- OverStock, Travala.com, Snel
- OpenBazaar, Peddler.com, Galaxus
- Ethlance, Sirin Labs, Mobisun
- TapJets

Volumes and Values

Start value in Aug 2015 $.66, and $2,611.43 Mar 11,2022. Bitcoin correlation for Ethereum is 0.916[4]. Your investment of $100 back in 2015, would worth $400,000 today.

Is Bitcoin or Ethereum correlated to the Stock market?

Historical correlations with Stock market

While bitcoin is often described as an alternative to gold, its historical price action suggests it’s more closely related to stocks[**5**].

How long does it take to mine one Ethererum and one Bitcoin?

- It takes around 7.5 days to mine Ethereum as of September 13, 2021 [**6**]
- Depending on the cost of electricity in a miner’s area, it could potentially cost $73,000 to process one bitcoin in a month’s time[**7**]

Is it too late to buy Bitcoin now?

- The market cap for Bitcoin is $1 Trillion now, the 10th most valuable asset in the world [**8**]
- Bitcoin supply is capped at 21 million by design[**8**]

Is bitcoin a ”gold safe-haven”[5]?

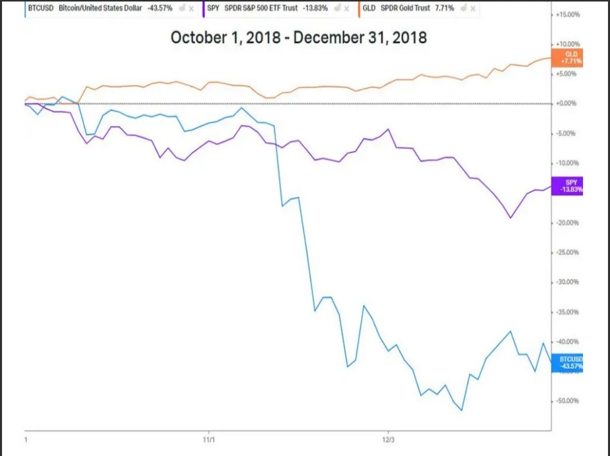
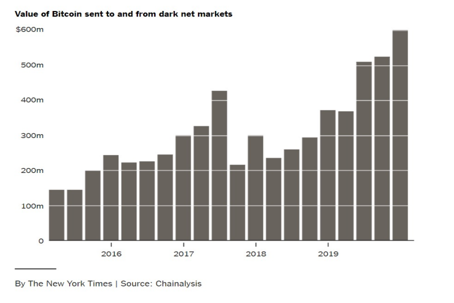


Figure: Gold in 2018 vs SP500 and Bitcoin[5]

See the Evil

”Crypto money laundering rises 30%, report finds”[11]


Value of Bitcoin sent to and from dark net markets

By The New York Times | Source: Chainalysis

Figure: Bitcoin is still popular among currency speculators, and illicit activity accounts for only 1 percent of all Bitcoin transactions. But that nearly doubled from the previous year. Illegal activity appeared to be one of the few parts of the Bitcoin economy impervious to changes in price, according to Chainalysis’s new Crypto Crime Report[10]

Models

Deep Learning

LSTM model

Machine learning

Multi-Linear Regression
Quadratic Discriminant Analysis (QDA)
K Nearest Neighborhood (KNN)
K-fold Cross validation

Data Source

NASDAQ

Gold
SP 500
Tesla
Bitcoin

Investing.com

Ethereum
Bitcoin

Research

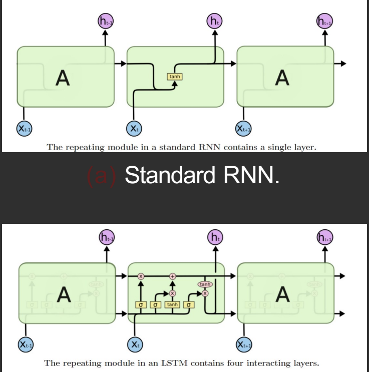
Santa Clara University

Long Term Short Term Memory, LSTM model

How important it is to remember the past

Humans don't start their thinking from scratch every second. As you read this essay, you understand each word based on your understanding of previous words. You don't throw everything away and start thinking from scratch again. Your thoughts have persistence[**9**].

LSTM model vs Standard RNN model



The repeating module in a standard RNN contains a single layer.

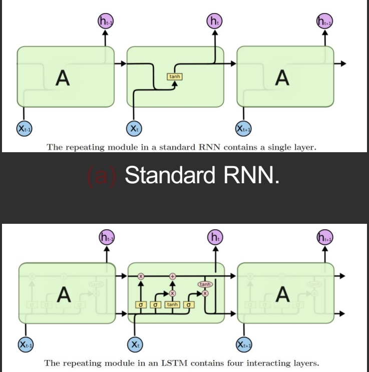
(a) Standard RNN.

The repeating module in an LSTM contains four interacting layers.

(b) LSTM.

Figure: One Layer LSTM vs One Layer Standard RNN[9].

LSTM model



The repeating module in a standard RNN contains a single layer.

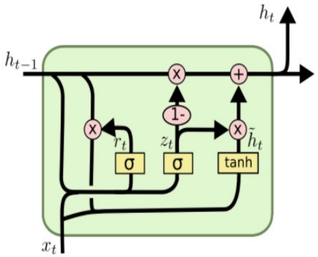
(a) Standard RNN.

The repeating module in an LSTM contains four interacting layers.

(b) LSTM.

Figure: One Layer LSTM vs One Layer Standard RNN[9].

LSTM inner layer



$$z_t = \sigma \left(W_z \cdot [h_{t-1}, x_t] \right)$$

$$r_t = \sigma \left(W_r \cdot [h_{t-1}, x_t] \right)$$

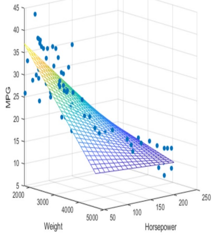
$$\tilde{h}_t = \tanh \left(W \cdot [r_t * h_{t-1}, x_t] \right)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

Figure: popular LSTM variant, introduced by Gers Schmidhuber (2000)[9]

Multi Linear Regression Model[12]

- A statistical technique that uses several explanatory variables to predict the outcome of a response variable.
- Multiple regression is an extension of linear (OLS) regression that uses just one explanatory variable.
- MLR is used extensively in econometrics and financial inference.



Resource: https://medium.com/analytics-vidhya/new-aspects-to-consider-while-moving-from-simple-linear-regression-to-multiple-linear-regression-dad06b3449ff

Figure: MLR Weight and Horse Power to predict MPG

Multi Linear Regression Model Equations and Definitions, Lecture

- $\hat{y}_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip}$
- Residual= $r_i = e_i = y_i - \hat{y}_i$
- Standard Residual= $r_i = e_i / stdDev(e_i) = \frac{e_i}{\sqrt{MSE(1-p_{ii})}}$
- Outlier: A data point whose response does not follow the general data trend (that is, an extreme y value)
- High Leverage: A data point with extreme predictor x values
- Influential: Observations that unduly influence regression, Cook’s distance
- Quantile-Quantile plot will tell the Residuals are normally distributed results.

When we need to use Ridge Regression

- In ordinary least squares fitting, estimates regression coefficients $\beta_0, \beta_1, ..., \beta_p$ by minimizing RSS

$$RSS = \sum_{i=1}^{n} (y_i - \hat{y_i})^2$$

$$RSS = \sum_{i=1}^{n} (y_i \hat{\beta}_0 - \hat{\beta}_1 x_{i1}, ..., \hat{\beta}_p x_{ip})^2$$

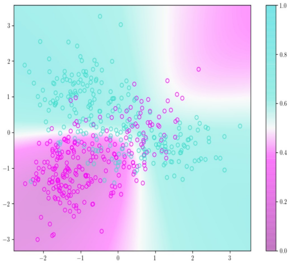
- Ridge regression minimized :

$$RSS + \lambda \sum_{j=1}^{p} \hat{\beta}_j^2$$

- Cross validation is used to estimate λ

Quadratic Discriminant Analysis

Quadratic Discriminant Analysis (QDA) is a generative model. QDA assumes that each class follow a Gaussian distribution. The class-specific prior is simply the proportion of data points that belong to the class. The class-specific mean vector is the average of the input variables that belong to the class.

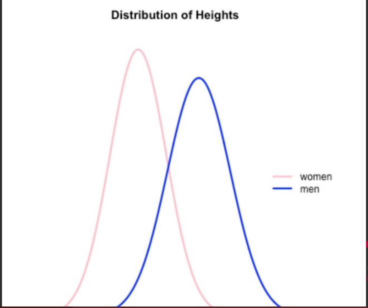


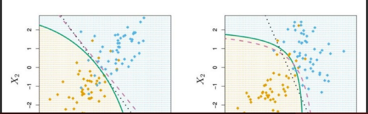
Decision boundaries of the fitted QDA along with the original classes of the data points. The colorbar illustrates the probability of belonging to class 1. Image by author.

Figure: Data over the decision boundary

Quadratic Discriminant Analysis equations and definitions

- The probability density function for a normal distribution $N(\mu,\sigma^2)$ is: $f(x|\mu,\sigma^2) = \frac{\sqrt{}1}{2\pi\sigma^2}\exp^{\frac{-(x-\mu)^2}{\sigma 2}}$
- For a given distribution the likelihood of the distribution parameters being $\mu,\ \sigma^2$ given the observation x is: $L(\mu,\ \sigma^2|x) = f\ (x|\mu,\ \sigma^2)$
- Classification Rule assign the observation x to the class with the greatest likelihood.$\hat{y} = argmaxL(\mu,\sigma^2|x)$


Distribution of Heights



K-Nearest Neighborhood Algorithm (KNN)

In statistics, the k-nearest neighbors algorithm (k-NN) is a non-parametric supervised learning method first developed by Evelyn Fix and Joseph Hodges in 1951,[1] and later expanded by Thomas Cover.[2] It is used for classification and regression. In both cases, the input consists of the k closest training examples in a data set. The output depends on whether k-NN is used for classification or regression.

KNN equations and definition

- KNN classifier identifies the K points in the training data that are closest to x_0 represented by N_0 the conditional probability for class j as the fraction of points in N_0 whose response values equal j:
  $Pr = (Y = j|x = x_0) = \frac{1}{K} \sum (I(y_i = j)$
- Classifies the test observation x_0 to the class with the largest probability.

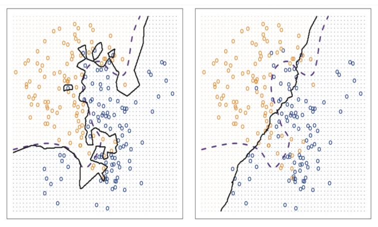


Figure: Boundary decision KNN

K-fold Cross Validation

- It is a data partitioning strategy using data to build a more generalized model
- The intention is to train data to predict the ”unseen” data avoiding over-fit
- It used to evaluate a model’s performance
- It is used for hyper-parameter tuning
- LOOCV is a special case, k=n; useful working with small dataset

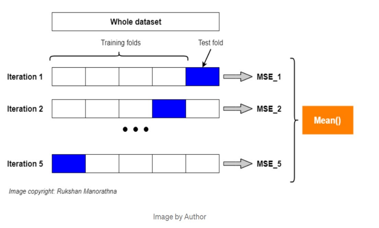


Image copyright: Rukshan Manorathna

Image by Author

Figure: 5-fold cross-validation for evaluating a model’s performance

Our Model Prediction

Hypothesis

There exists a positive correlation between the Price of Bitcoin(ETH) for a day and Price of Tesla,gold and S&P Stocks.

Relevant Null Hypothesis:
H0: There is no relationship between the Price of Bitcoin(ETH) and Price of Tesla,gold and S&P Stocks.

Alternative Hypothesis:
H1: There exists a relationship between the Price of Bitcoin(ETH) and Price of Tesla,gold and S&P Stocks.

How Bitcoin and Ethereum did in 2018

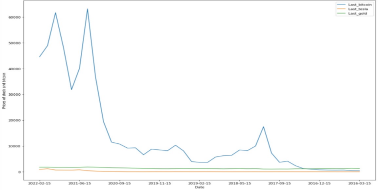


Figure: Bitcoin crash and recovery vs Gold and Tesla

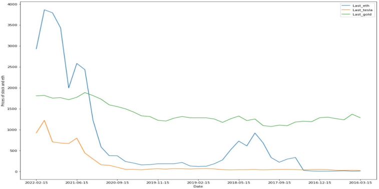
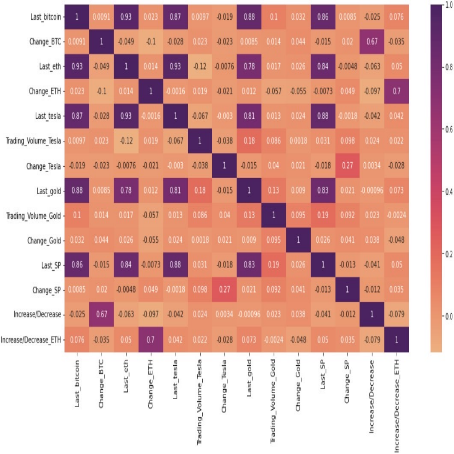


Figure: Ethereum crash and recovery vs Gold and Tesla

How Bitcoin and Ethereum are correleted to the stock market?



	Last_bitcoin	Change_BTC	Last_eth	Change_ETH	Last_tesla	Trading_Volume_Tesla	Change_Tesla	Last_gold	Trading_Volume_Gold	Change_Gold	Last_SP	Change_SP	Increase/Decrease	Increase/Decrease_ETH
Last_bitcoin	1	0.0091	0.93	0.023	0.87	0.0097	-0.019	0.88	0.1	0.032	0.86	0.0085	-0.025	0.076
Change_BTC	0.0091	1	-0.049	-0.1	-0.028	0.023	-0.023	0.0085	0.014	0.044	-0.015	0.02	0.67	-0.035
Last_eth	0.93	-0.049	1	0.014	0.93	-0.12	-0.0076	0.78	0.017	0.026	0.84	-0.0048	-0.063	0.05
Change_ETH	0.023	-0.1	0.014	1	-0.0016	0.019	-0.021	0.012	-0.057	-0.055	-0.0073	0.049	-0.097	0.7
Last_tesla	0.87	-0.028	0.93	-0.0016	1	-0.067	-0.003	0.81	0.013	0.024	0.88	-0.0018	-0.042	0.042
Trading_Volume_Tesla	0.0097	0.023	-0.12	0.019	-0.067	1	-0.038	0.18	0.086	0.0018	0.031	0.098	0.024	0.022
Change_Tesla	-0.019	-0.023	-0.0076	-0.021	-0.003	-0.038	1	-0.015	0.04	0.021	-0.018	0.27	0.0034	-0.028
Last_gold	0.88	0.0085	0.78	0.012	0.81	0.18	-0.015	1	0.13	0.009	0.83	0.021	-0.00096	0.073
Trading_Volume_Gold	0.1	0.014	0.017	-0.057	0.013	0.086	0.04	0.13	1	0.095	0.19	0.092	0.023	-0.0024
Change_Gold	0.032	0.044	0.026	-0.055	0.024	0.0018	0.021	0.009	0.095	1	0.026	0.041	0.038	-0.048
Last_SP	0.86	-0.015	0.84	-0.0073	0.88	0.031	-0.018	0.83	0.19	0.026	1	-0.013	-0.041	0.05
Change_SP	0.0085	0.02	-0.0048	0.049	-0.0018	0.098	0.27	0.021	0.092	0.041	-0.013	1	-0.012	0.035
Increase/Decrease	-0.025	0.67	-0.063	-0.097	-0.042	0.024	0.0034	-0.00096	0.023	0.038	-0.041	-0.012	1	-0.079
Increase/Decrease_ETH	0.076	-0.035	0.05	0.7	0.042	0.022	-0.028	0.073	-0.0024	-0.048	0.05	0.035	-0.079	1

Figure: Correlation

Conclusion

What we learned

Models pros and cons?

We used the following models:

- Multi-Linear Regression
- Quadratic Discriminant Analysis (QDA)
- K Nearest Neighborhood (KNN)
- K-fold Cross validation
- Long Term Short Term Memory (LSTM)

References

[1]"Using machine learning to predict future bitcoin prices" https://towardsdatascience.com/using-machine-learning-to-predict-future-bitcoin-prices-6637e7bfa58f

[2]"Time-Series Forecasting: Predicting Stock Prices Using An LSTM Model" https://towardsdatascience.com/lstm-time-series-forecasting-predicting-stock-prices-using-an-lstm-model-6223e9644a2f

[3]" There are 23 bitcoin datasets available on data.world.” https://data.world/datasets/bitcoin

[4]"Ethereum Historical Dataset" https://www.kaggle.com/prasoonkottarathil/ethereum-historical-dataset

[5]" Is bitcoin an uncorrelated asset?" https://www.marketwatch.com/story/is-bitcoin-an-uncorrelated-asset-these-stocks-and-funds-boast-correlations-higherand-lowerthan-coinbase-11622826320

Santa Clara University

The End

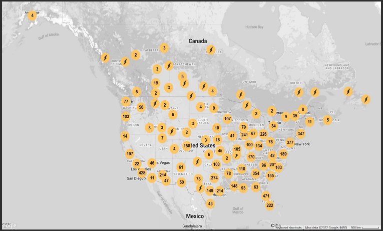


Figure: Some of Bitcoin ATM locations! Are we late?