Social Cryptolytics

Bitcoin and Social Media Data Investigation

Michael Runyan†, Michael E. Cotterell
Department of Computer Science
University of Georgia
Athens GA, USA
{msr88812, mepcott}@uga.edu

ABSTRACT

This investigation furthers the field of social cryptolytics, by proposing a relationship between the frequency of "trigger words" in social media outlets such as forums, and fluctuation in the value of cryptocurrency. Bitcoin was chosen as the currency of investigation for the large quantity and long history of posts and discussions, and the long history of price data. The social media outlets include a subset of bitcointalk.org posts and comments and the Bitcoin subreddit. These sources combined totaled over 7 million posts or 1.6 GB of data. These two outlets have data dating back to June 2010 and are some of the leading medium of Bitcoin pertaining social media posts. Including two outlets help to diversify the data and avoid the influence of forum specific vernacular skewing the data. We proposed that the frequency of words (hits) used which are most correlated to the price of cryptocurrency, once reduced to account for similar variation, can be used to create a model to predict the future value of bitcoin through the use of Tensor Flow, an open source machine learning framework. This study attempts to find a group of word frequencies which defines Bitcoin, a decentralized virtual currency.

Problem Statement

Although the potential for cryptocurrency is through the roof, it is still a rather new form of payment. The uncertainties of any new currency make it very volatile. Unlike traditional currencies, cryptocurrencies are not dependent on the state of a country's economy, they do not resemble gold or silver which have a physical counterpart, or like stock which tie into the prosperity of the company. Although cryptocurrency has been a part of mainstream financial media, it has little research on its drivers, as Kristoufek states in his 2015 paper "the research community is still primarily focused on the currency's technical, safety and legal issues, but discussion about the economic and financial aspects remains relatively sparse" [1]. This paper attempts to find a counterpart in social media to this new form of investment.

Hypothesis

• We proposed that frequency of words used which are most correlated to the price of cryptocurrency, once reduced to account for similar variation, can be used to create a model to predict the future value of Bitcoin.

KEYWORDS

Bitcoin, Social Media, Data Analysis, Machine Learning

INTRODUCTION

1.1 Evolution of Cryptocurrencies

Cryptocurrency is a new emerging phenomenal which seems to hold a valuable place in the future. Although bitcoin is the first currency to secure a place as a reliable form of digital currency, it is not the first to attempt at the market. The first digital cash was a project by David Chaum to provide one of the most important services of cryptocurrency, to make blind and untraceable payments online. Digicash accomplished this through RSA encryption, an algorithm still used today in web encryptions [2]. Although an ingenious inventor, the revolutionary Chaum saw too much value in his work, and turned down multiple offers for his currency, including a $180 million offer from Bill Gates, and eventually went bankrupt in 1998 [2].

Following Digicash, a wave of digital currency attempted to corner the new market in the US. PayPal found a spot in the market by allowing person to person interactions, a home in the user-base eBay community, and a merging with X.com, Elon Musk's online banking company [3]. Other currencies such as e-gold ensured the value of their currency with the price of gold and grew a strong customer base, but eventually fell victim to regulation busts, and federal investigation [2].

After these failed attempts (not including Paypal which is currently valued by some to be worth $49.6 billion) [4]. Satoshi Nakomoto, the anonymous founder of Bitcoin, created a digital currency in 2008 which revolutionized the market [4]. The differentiating feature of Bitcoin was its decentralized nature, which took advantage of blockchain technology to authenticate transactions between users in a peer-to-peer system.

The Blockchain is the foundation of all current relevant cryptocurrencies. It is the key that ensures the reliability and authenticity of the currency. Each transaction that takes place over a specific interval of time is bundled into a 'block". These blocks are then sent to 'miners' who solves complex algorithms to confirm the block and add it to the chain of blocks or otherwise referred to as a ledger. Once a block is added, it is timestamped and becomes unchangeable. This ledger is stored on a series of private nodes, and when changes are made, they are broadcasted and stored on every node.

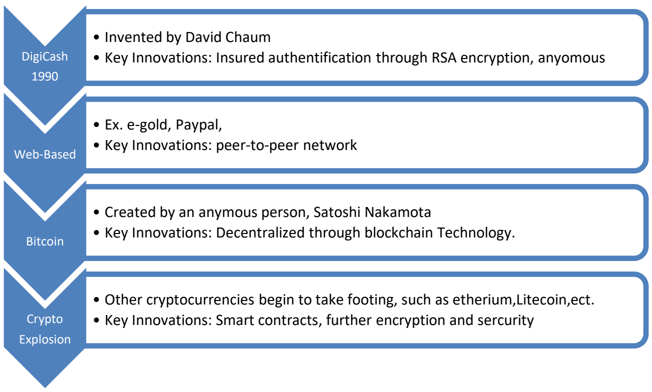


Figure 1: A flowchart of the evolution of cryptocurrency and the key features associated with each step [2], [3], [4].

After the success of Bitcoin, many other cryptocurrencies gained a footing in the market. Ethereum was launched in 2015 and furthered the usefulness of cryptocurrencies by allowing smart contracts and streamlining the mining process. Litecoin quickens the transaction time and lowered the transaction fee compared to bitcoin [5]. Other currencies such as Ripple, target financial institutions and strive to create a high-speed low-cost payment framework [5]. Our study focuses on Bitcoin as a representative of most cryptocurrencies, because it is the most first decentralized and most well know cryptocurrency.

1.2 Possible Price Drivers of Bitcoin

There's been much speculation on the price of bitcoin, and its decentralized nature leaves endless possibilities for underlying factors behind the price. A few of these possible factors which are worth mentioning are the mining community, online interest, and legal regulations.

Although decentralized, the software behind its platform is prone to change. If more then 50% of the global network of miners agree, then a change can take place [6]. Bitcoin has gone through two hard forks in its lifetime to create Bitcoin cash and Bitcoin gold. Both of these forks have led to large fluctuations in price.

Interest has been found to be a driver in the price of Bitcoin and also vice versa. Price has been found to be a driving factor in the interest of Bitcoin, as found in Kristoufek 2015 study who compared the search frequency of the word "Bitcoin" on Google and Wikipedia, and the price of Bitcoin. He found a strong correlation between the two.

Major political events pertaining to bitcoin have also been observed to have a significant influence on the price. Baidu's, an important Chinese online shopping player, announcement to accept bitcoin showed a huge surge in value, but China's ban of the coin led to a sharp drop in price [6]. Japans Virtual Currency Act in March 2017, made bitcoin a legal tender and Bitcoin showed another surge in price [6].

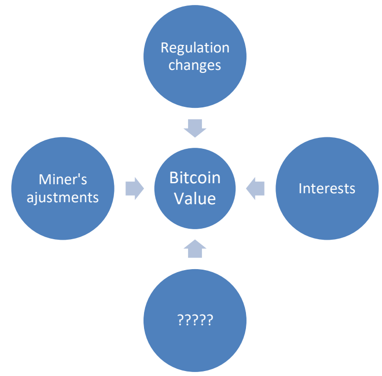


Figure 2: Describes possible underlying drivers behind the value of Bitcoin.

1.3 Related Work

Previous studies around social media and cryptocurrency price suggest there is a strong correlation between the two. Feng Mia found in "How Does Social Media Impact Bitcoin Value? A Test of the Silent Majority Hypothesis" that with a test of Twitter and bitcointalk.org (a bitcoin forum), the forum had a stronger correlation with bitcoin price [7]. He found "More bullish (or bearish) forum posts are significantly associated with higher (or lower) next-day bitcoin market price" [7]. Feng more specifically focused on the silent-majority theory, and which type of user correlated best with the price fluctuation.

In "Network Structure and Predictive Power of Social Media in the Bitcoin Market" by Peng Xie, data was again collected from bitcointalk.org, however, the researchers focused on acquiring valuable data from the comments and topics. They filtered out authors quoting each other, and other comments which add hardly any new information. They found "when the percentage of negative words in discussion networks with an average degree of 0.553 (sample mean) is 1% higher, the next-day Bitcoin return is 0.076% lower [8]. In contrast, the same 1% increase in the percentage of negative words for discussion networks with an average degree of 0.342 (one standard deviation lower from the sample mean) is associated with a doubled decrease of 0.153% in the next-day Bitcoin return." [8]. This study relates similarly to ours in the fact they analyze the post by the frequency of words, however, they do not diversify their data. It does suggest that a next day model does suggest a correlation.

Young Kim performed a similar study but measured each comment and topic on the sentiment of Bitcoin, Ethereum, and Ripple. Each post was given a tag from very positive to very negative with a classifier based on machine learning. The data was run through the Granger test and it concluded a strong relationship. For Bitcoin "The prediction result proved to be the highest when the time lag was six days with an accuracy of

79.57%", Ethereum "The predicted result proved to be highest when the time lag was six days with an accuracy of 71.823%", and "predicted fluctuation in the price of Ripple proved to be highest when the time lag was seven days with an accuracy of 71.756%" [9].

MATERIALS AND METHODS

2.1 Data Collection

Data was gathered from two different sources in order to diversify the data and avoid the influence of forum's tendency to adopt unique vernacular and to form alike opinions. Bitcointalk.org was chosen because it contains one of the largest user bases of any cryptocurrency specific network. As of November 2018, Bitcoin Talk had over two million members. Bitcoin talk also has one of the largest archives of Bitcoin-related posts and discussions dating back to 2010. Reddit as of 2018 is the 18th most popular site worldwide, and the Reddit bitcoin subreddit also carries the advantage of having a large community of 982k subscribers. The community provides hundreds of post a day ranging from Bitcoin memes to heated debates on the direction of the currency. These two forums are based heavily in well informed tech-based users and provide an excellent source of posts and discussions on everything pertaining to Bitcoin. The posts collected from these two sites are available here: https://www.kaggle.com/underdog7890/bitcointalk-and-reddit-rbitcoin-post-data

Key Python Tools: Requests, Beautifulsoup, pushshift.io

2.2 Word Determination

A sample of 25,000 posts which ranged over a month was taken from the data set mentioned above and was then parsed for the frequency of specific trigger words and was congregated with a time stamp into CVS files. The set consisted of Reddit submission posts, which introduce topic discussions and best represent the direction of the comments on that day. Most of the words to be searched for were chosen from the AFINN-111 dictionary [10], a public dictionary of words which suggest sentiment on a scale ranging from -5 to 5 (-5 the most negative sentiment, and 5 the most positive). This dictionary has been used in multiple studies successfully to estimate sentiment in blog posts and tweets. The dictionary was reduced from 3,000 words to 2,000 of the words which suggest the strongest sentiment positive or negative (all words associated with a -1 or 1 sentiment were removed). This was done to more efficiently parse the post, but still retain the words which may be weighted the most in the analysis. A smaller set of words were chosen for their connection to bitcoin discussions. Words such as blockchain, legality, and cryptocurrency, etc. These words are not in a sentiment dictionary but are used often to describe confidence or doubt in bitcoin.

To perform a later reasonable analysis of the total data set and to reduce possible noise in the model later, the Kendall's Tau Coefficients were found for the hits in the sample data set. Kendall Tau Coefficients are used to determine a relationship between two sets of data by comparing the coordinate or discordant pairs in the data set. Coordinate refers to whether both price and hits are increasing or both decreasing between two adjacent days. This test was chosen for its advantage of looking at data a step at a time instead of inspecting overall trends in the data.

$$\tau = \frac{(\#\ of\ concordant\ pairs) - (\#\ of\ discordant\ pairs)}{n(n-1)/2}$$

Equation 1: [11] Equation of Kendall coefficient (τ) where n is the number of pairs.

Kendall test was performed on the sample set of data which was compiled from 25,000 posts. Those words in which the absolute value of their Kendall coefficients > .1 were not used in the later analysis of the data. .1 was chosen because it is large enough to assume a connection of the two data sets, and it reduced the words to a number which was reasonable to search for with provided resources.

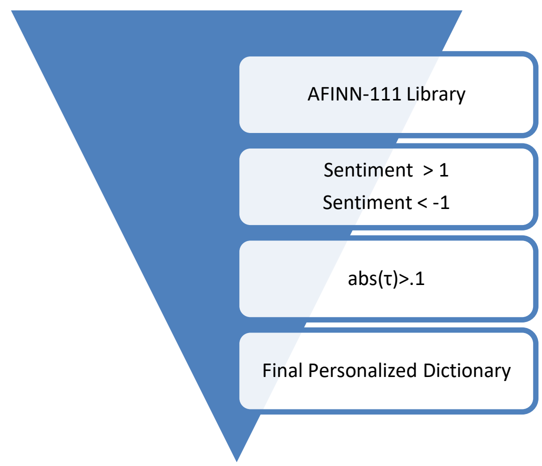


Figure 3: Visual of the reduction of the word selection. The Kendall Tau coefficients were derived from counting a sample set of posts of 25,000 from Reddit submissions.

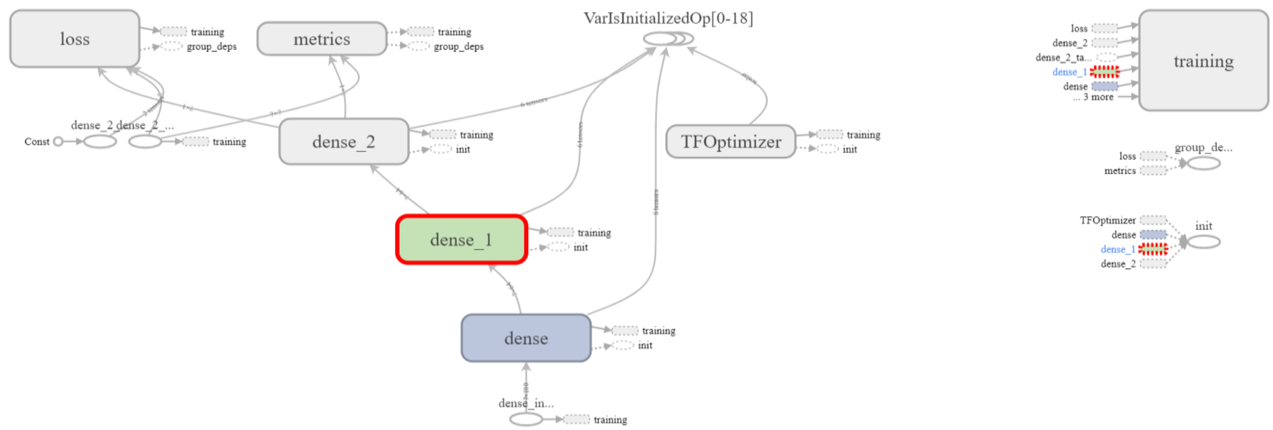
With this reduced word set, it was efficient to find the frequency of each word in the complete dataset of over 7 million posts (1.62 GB).

2.3 Hit Distribution

The distribution of the data was then analyzed by plotting the total number of hits for each specific day divided by the total hits. This data is used to account for the differing number of posts analyzed each day. Because lack of posts can skew the relationship between work frequency and price, the data was reduced to the days which contained 98% of the total number of hits (word frequencies).

2.4 Data Reduction

To further reduce the data set to more efficiently perform machine learning, it was shrunk to the words which describe the most amount of data. This was done with the implementation of singular value decomposition on the matrix. The matrix is decomposed to three separate matrices.



$$A = U\Sigma V^T$$

Equation 2: [12] Equation of singular value decomposition where A represents the original array, U are the gen coefficient vectors, Sigma is the mode amplitudes, and V^T expresses level vectors.

The elements of V^T the array are then compared to determine the similarity of the variance in each column, and then only those columns which explain the most variance are kept. 100 hundred variables were kept because it reduces the noise of extra variables down to an improved amount and increases the efficiency of running the models.

2.5 Multiple Linear Regression

To acquire an overview of the data's predictive capability, we applied multiple linear regression. The data was mixed into a random order and split into a training set of 75% and a testing set of 25%. The training set was then used in multiple linear regression to create a multiple linear regression model (MLR) and then evaluated with the testing set. The coefficients were then used to indicate the strength of each of the words in predicting the value of bitcoin.

$$\widehat{P}_t = m_{w1}f_{w1,t} + m_{w2}f_{w2,t} + \cdots + m_{wn}f_{wn,t}$$

Equation 3: $\widehat{P}_t$ represents the price of the currency at that specific day t. $f_{wn,t}$ is the frequency of a specific word and is a function of time t. m_{wn} represents the coefficient determined through MLR.

2.6 Machine Learning

The reduced data set is again mixed into a random order and split into a training set which consists of 75% of the data and a testing set of 25% of the data. The data was then normalized into a range of -1 to 1 to help ensure the data converges in the model. A keras Sequential model was created consisting of two Dense layers of 64 nodes each, and an output layer of a single node. Only two layers are chosen to avoid overfitting of the data by the network, and the output consists of continuous value which represents the price prediction. The number of nodes was decided on graphical visual of the le because it was used as the example number in tensorflow's regression demo. A graphical visualization of the learning process is provided in figure 5.

The network is then trained using the training data. The model is set to stop training after the Mean Absolute Error stops decreasing in order to avoid overfitting to the training set. An example of this in a model predicting the price from hits 12 days before the price is shown in Figure 4. This network was then used to test for the mean absolute error of the price from words frequency of 1 day before to 30 days before, to determine the lag to produce the most accurate results. Each lag was tested 10 times with the data randomly sorted into the training and testing set, and then the average of the Mean Absolute error was recorded.

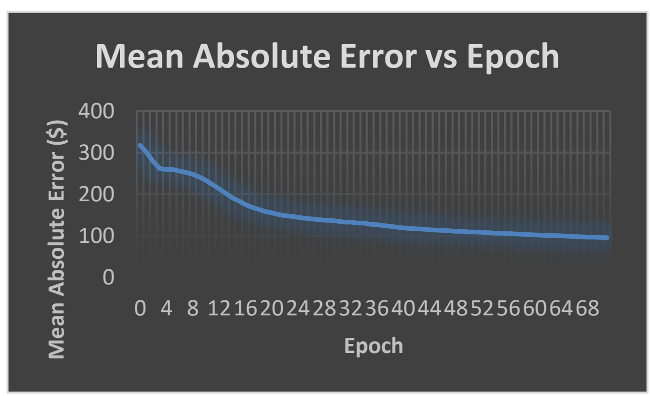


Figure 4: Models the improvement of the keras model throughout the fitting process. The keras callback model stopped the epoch after 71 iterations because the model stopped improving significantly at this point.

Figure 5: Graphical visualization of the keras model used to predict the price. This model is simplified, and the interactive model is provided in the GitHub and can be viewed using Tensor Board.

For further Investigation, the data from multiple days before were all used to create a model with keras. We performed an analysis on data including 1 day prior to the data for up to 90 days prior.

RESULTS

The sample set of 25,000 posts chosen to represent the data consisted of posts from the 2/22/2013 to 11/19/2013, a month which well reflects the fluctuation of Bitcoins price throughout the timeframe of the total data. The Kendall Coefficients of this data ranged from .411 (Ban) to .504 (Lawsuit). The 20 words which have the highest absolute value of their Kendall Tau coefficient are graphed below in figure 6.

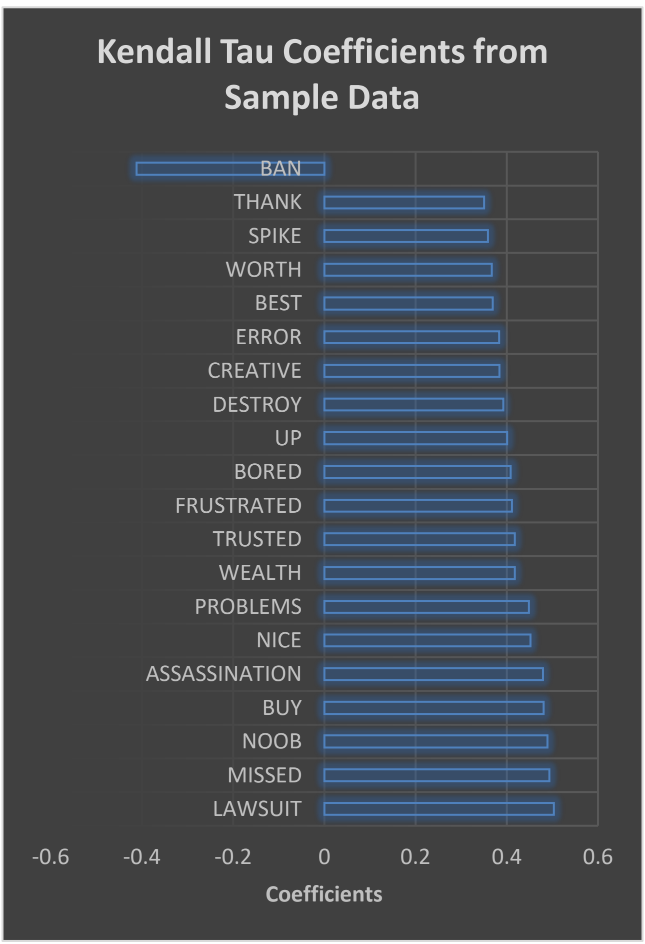


Figure 6: The twenty Word Frequencies which had the highest Kendall Coefficients once the absolute value was taken.

The complete dataset data consisted from hits from June 2010 to September 2018, but the beginning and end of the data set analyzed very few posts in comparison to the rest of the data. It was found that September 2010 to March 2017 contained 98% of the total hits, and the data was reduced to this range.

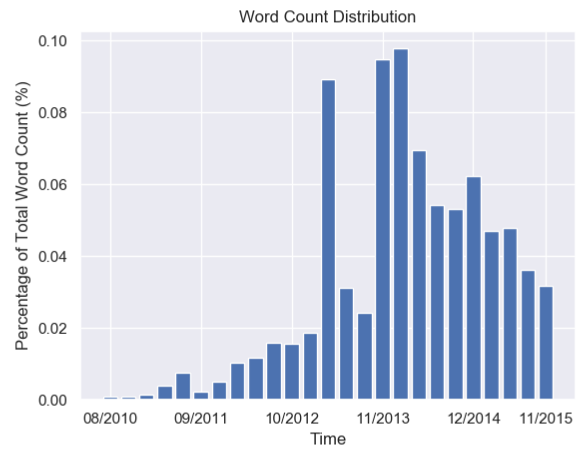


Figure 7: Describes the distribution of posts analyzed by dividing the number of hits per unit of time by the total number of hits. This data was then arranged into bins as seen above.

Reducing the excess noise in the data set with SVD reduced the mean absolute error of next day predictions from the by $5.91.

The predictive power of word frequency diminishes in a linear fashion and had a trend line with an R squared value of .571. The first 12 days before have similar predictive power with 12 days before having the lowest error of any lag time.

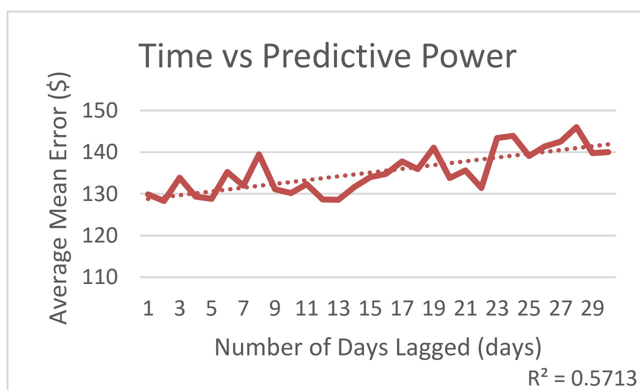


Figure 8: Average Mean Error is taken as the average of 10 trials. The values are acquired from the predictions of the 2 layered keras model

When multiple linear regression was performed to predict the price 12 days before, the average of the mean absolute error of 100 trials was $163.07 and had the average R squared value was .742.

Figure 9: The average linear coefficient assigned to the 25 words with the largest absolute value of the coefficient. These coefficients are an average of 100 trials.

The Average mean absolute error of a 12-day prior prediction was \$128.65. The example model visualized below in figure 10 had a mean absolute error of \$128.19 when compared against its test set. The average R squared value of this data using a uniform average was .723, slightly higher than the one produced by multiple linear regression.

The mean absolute error when plotted vs the price, indicates they share a positive relation. This relation can be modeled by polynomial regression and has an R squared value of .6669.

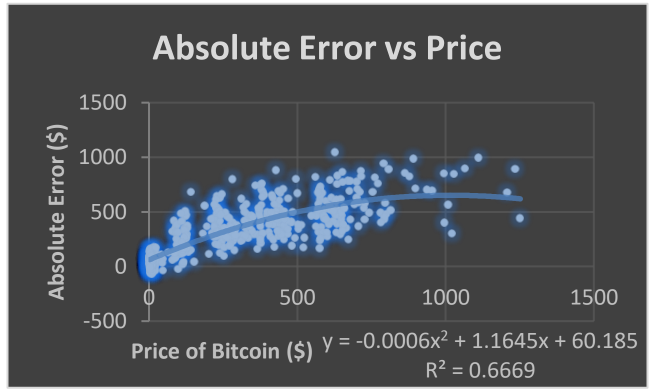


Figure 10: Describes the distribution of the error across the value of bitcoin at the time of prediction. This is done by plotting the absolute error of the prediction vs the price of bitcoin.

Training the model with more than just one day's worth of data provided a much more accurate model. The accuracy increased in a logarithmic fashion with an R squared value of .9606. It seemed after 90 days the improvement slowed to an amount that was insignificant. Training with 90 days of data created a model with an average mean absolute error of \$53.23. This model produced an R squared value of .878.

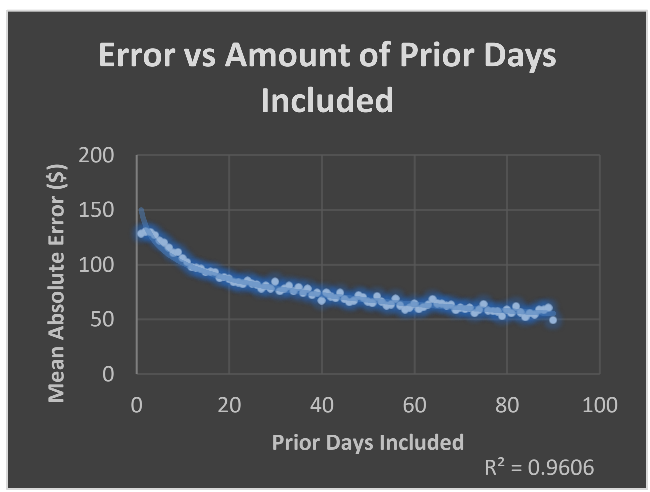


Figure 11: The amount of data used to train the keras model vs the mean absolute error.

DISCUSSION

The strong correlations described by the Kendall Tau coefficients in the sample set of data is the first indication that word frequency, and price share a relationship. The coefficients were skewed more positive than negative, and this is to be expected with a majority of the post falling in a time when bitcoin was gaining most of its popularity and the price increasing with seemingly no ceiling.

As seen in figure 8, the accuracy of the keras model stays relatively even until predictions made of price 13 days later, and at this point the accuracy begins to diminish. The frequency of words 12 days former are the best indicators of price.

The coefficients in the multiple linear regression model, indicate the weight of each word in the model and their predictive power. A few of the top contributors carry a positive sentiment and indicated an increase in price. Unexpectedly, words which carry a negative sentiment have the largest coefficients negatively and positively. For example, greenwashing, defined by Merriam-Webster as “expressions of environmentalist concerns especially as a cover for products, policies, or activities” [13], carries a strong negative sentiment, however, has the highest direct correlation to the price. Other words such as menace, disgrace, and enslaves carry a direct correlation also. The average R squared value for this model of .742 showed there is a strong correlation between the two variables. The average mean absolute error of $163.07 makes this model capable of predicting the value of Bitcoin, however, there is a lot of room for mistakes.

The keras model based on hits 12 days before, achieved a mean absolute error of $128.65, a value $34.42 less than multiple linear regression. This error is unevenly distributed, straying more on the larger predictions. It resembles a second-degree polynomial trend line which suggests the error will lie between a minimum of $60.18 when the value of bitcoin is $0 and $625.21 when the value is $970.41. This is to be expected as the percentage change of larger prices is a larger dollar amount than lower prices. The R squared value of .6669 indicate this is a very strong relationship and the mean absolute error makes this model of similar usefulness in multiple linear regression for predicting future value.

Including multiple days of data dramatically improved the model. The logarithmic shape of the number of days included vs error resembles the fact that further and further prior days correlate less and less with the future price. The model trained with 90 days’ worth of prior data made a clear improvement in predictive power. Models depending on multiple days of prior data creates a model the ensures a relationship between the price of Bitcoin and social media comments with its R squared value of .878 but also can predict the value with a mean absolute error of only $53.23

CONCLUSION

The value of bitcoin has a clear correlation with word frequencies of specific words used in social media. The correlation in our model, however, does not rely on the positive or negative sentiment of the words, and we urge further investigation of word features which influence their predictive.

Using a keras sequential model, the best predictions with a single day of data are performed with word frequencies 12 days prior to the predicted price. This correlation diminishes linearly when tested more than 12 days prior to the desired prediction.

Both of these models indicate a very strong relationship between the value of Bitcoin and the frequency of words used in social media. The model created is of limited usefulness because of the range of error.

Including multiple days prior greatly increased the usefulness and minimized the error to a useful amount. The value of bitcoin may fluctuate more the $60 dollars in one day. With this model, the price of bitcoin can be predicted with a reasonable degree of accuracy. This can be used to make assessments of bitcoins next day value.

APPENDIX

In order to recreate these results, and or improve the process used here, the code has been provided on GitHub and the data on.

GitHub: https://github.com/Michael-Runyan/Social_Cryptolytics

Data: https://www.kaggle.com/underdog7890/bitcointalk-and-reddit-rbitcoin-post-data

To recreate these results, the set of data scraped is provided. The collected posts are already provided in …, so the Data_Collection folder is unneeded unless a person would desire to completely scrape these two sources. In that case, Run the file bit_driver.py located in ~/Data_Colletion/Bitcoin_Talk/ and change topic variable to the topic that is desired to start from. In ~/Data_Collection/Reddit run Red_comment.py for submission data and then change URL to gather comment data.

Using the post reader class, the posts can be parsed through and searched for the words in a text file that is found in ~/Post_Reader/input/words. A sample set of our personalized dictionary and a csv file of the times are provided but can be changed. The posts to be parsed through should be arranged into the subfolders of ~/Post_Reader/input/posts/. Following this step, the File_Iter.py file should be run with the parent variable labeled as the parent file is in the input. Once all the parent directories have been parsed, the created csv files should be combined with combine_all_files.py. The outputted csv file should be moved to ~/Word_Determination/input

Within ~/Word_Determination Kendall_driver.py should be run, and the words with a Kendall coefficient above .1 should be moved to a new text file and inserted to ~/Post_Reader/input/words and the steps in the previous paragraph should be repeated except the outputted file should be moved to ~/Evaluation/input where Distribution.py should be run and the best range of dates should be included for further analysis and be moved to ~/Data_Reduction/input

From ~/Data_Reduction, SVD_Driver.py should be run and the outputted file moved to ~/Prediction/input.

Within ~/Prediction/, edit Lag.py to name the input and output file correctly and then run to determine the best number of days to lag.

With this lag, ~/Prediction/Mul_Lin_Red.py should be run with the determined lag rate. The coefficients will be saved to a csv file.

Inside ~/Predictions, Lag.py should be run with the portion of code uncommented to include multiple days prior and determine the best number of days to use.

To determine the R squared values for predictive data, uncomment out the line to save prediction values in ~/Predictions/Neural_Net.py and input the saved data into ~/Evaluation/R_Squared.py/input and run R_Squared.py.

Further research into the subject could not reduce the word count as much and create a model which take into account more words.

The posts are stamped down to the minute, and this dataset could be used to compare posts divided into categories between hours of the day. These different times a day may correlate to price at differing times the following day.

ACKNOWLEDGMENTS

This research was supported in part by the Department of Computer Science at the University of Georgia through the CSCI 4960 course.

REFERENCES

[1] Kristoufek L (2015) What Are the Main Drivers of the Bitcoin Price? Evidence from Wavelet Coherence Analysis. PLoS ONE 10(4): e0123923. https://doi.org/10.1371/journal.pone.0123923

[2] Ken Griffith. 2014. A Quick History of Cryptocurrencies BBTC - Before Bitcoin. (April 2014). Retrieved December14, 2018 from https://bitcoinmagazine.com/articles/quick-history-cryptocurrencies-bbtc-bitcoin-1397682630/

[3] Christina Mercer. 2015. History of PayPal: 1998 to now. (November 2015). Retrieved December 14, 2018 from https://www.techworld.com/picture-gallery/business/history-of-paypal-1998-now-3630386/

[4] Rebecca Borison. 2015. Just How Much Is PayPal Worth as an Independent Company? (February 2015). Retrieved December 14, 2018 from https://www.thestreet.com/story/13041918/1/just-how-much-is-paypal-worth-as-an-independent-company.html

[5] Anon. 2018. Bitcoin vs. Litecoin, Ethereum, Ripple, and Dash. (February 2018). Retrieved December 14, 2018 from https://www.bitcoinmarketjournal.com/bitcoin-vs/

[6] Alicia, Lucy Cameron, Lucy, Cameron, and Kelly Trinh. 2018. Four factors driving the price of Bitcoin. (August 2018). Retrieved December 14, 2018 from http://theconversation.com/four-factors-driving-the-price-of-bitcoin-87244

[7] Feng Mai, Zhe Shan, Qing Bai, Xin (Shane) Wang, and Roger H.l. Chiang. 2018. How Does Social Media Impact Bitcoin Value? A Test of the Silent Majority Hypothesis. Journal of Management Information Systems 35, 1 (February 2018), 19–52. DOI:http://dx.doi.org/10.1080/07421222.2018.1440774

[8] Peng Xie, Hailiang Chen, and Yu Jeffrey Hu. 2017. Network Structure and Predictive Power of Social Media for the Bitcoin Market. SSRN Electronic Journal (2017). DOI:http://dx.doi.org/10.2139/ssrn.2894089

[9] Young Bin Kim et al. 2016. Predicting Fluctuations in Cryptocurrency Transactions Based on User Comments and Replies. Plos One 11, 8 (2016). DOI:http://dx.doi.org/10.1371/journal.pone.0161197

[10] "A new ANEW: Evaluation of a word list for sentiment analysis in microblogs", Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big things come in small packages 718 in CEUR Workshop Proceeding : 93-98. 2011 May. http://arxiv.org/abs/1103.2903

[11] Kendall, Maurice; Gibbons, Jean Dickinson (1990) [First published 1948]. *Rank Correlation Methods. Charles Griffin Book Series (5th ed.)*. Oxford: Oxford University Press. ISBN 978-0195208375.

[12] Golub, Gene H.; Kahan, William (1965). "Calculating the singular values and pseudo-inverse of a matrix". *Journal of the Society for Industrial and Applied Mathematics*: Series B, Numerical Analysis. 2 (2): 205–224. doi:10.1137/0702016. JSTOR 2949777.

[13] Anon. Greenwashing. Retrieved December 14, 2018 from https://www.merriam-webster.com/dictionary/greenwashing?src=search-dict-hed