

Computational Approaches to Predicting Cryptocurrency Prices

Chris Pelletier '20
AI for the Humanities

Introduction

Background

My project idea was originally intended to look at the sentiment of Bitcoin news over the past year, but the Aylien API I used only went back sixty days which still ended up yielding some interesting results. Bitcoin experienced one of its longest plateaus since it gained mainstream popularity between October 12th and November 14th 2018. The plateau in pricing is interesting to look at because the sentiment and quantity of news articles from that time period did not remain static and fluctuated constantly, unlike Bitcoin's price. This data is particularly interesting in the context of Ben Hunter-Craig's article *Sentiment Analysis on Cryptocurrency News* (Fig. 1). His scraping and sentiment analysis were accomplished using different methods than mine. He used NLTK (Natural Language Tool Kit) paired with Vader for sentiment analysis from articles obtained using Cryptocompare's news API. His graph indicated Bitcoin news influenced Bitcoin price, but his data was collected during the time when Bitcoin hit its peak value and mainstream relevance. My data in the context of his findings shows that the media more easily influences mainstream Bitcoin owners than tech enthusiasts that have continued to use Bitcoin for years.

Hypothesis

People in Cryptocurrency forums will often call Bitcoin Hypecoin due to the prevalence of messages saying to hold with posters citing reasons it will recover during times of declining value. The mainstream media gave the opposite message during Bitcoin's most successful year comparing it to the tulip bubble, with outspoken finance experts such as Jamie Diamond calling it fraud. The SEC's ruling that Bitcoin was a taxable security also garnered negative press and predictions of its future value. My experiment was carried out under the assumption that most buyers of Bitcoin were people who got their trading advice from mainstream news sources and that the prevalence of negative or positive articles would correlate with Bitcoin's price.

Methods

Aylien API was used for this project due to its simplicity for the end user. Aylien is a natural language processing GUI that automates the scraping and collection of data from various news sources. It also automatically assigns these articles a sentiment, and classifies the articles based on their content. I only used articles that were tagged as financial because articles about the social implications of Bitcoin aren't as pertinent to its pricing. I did this under the assumption most buyers don't care about the energy waste or criminal uses of Bitcoin. Average buyers care only about turning a profit.

Sentiment Analysis vs. Technical Analysis

Using Sentiment Analysis to Predict Stock Prices

In Dev Shah, Haruna Isa, and Farhana Zulkernine's essay *Predicting the Effects of News Sentiments on the Stock Market* they "...achieved a directional accuracy of 70.59% in predicting the trends in short-term stock price movement." They used natural language processing programs to create sentiment scores for news articles from the moneycontrol.com news API using a preexisting corpus to find sentiment. They looked at pharmaceutical stocks and the news sentiment surrounding them to predict price and found a stronger correlation than my data was able to with Bitcoin. Their program made highly accurate buy and sell predictions, but their limited data and lack of explanation as to how their model was prevented from over-fitting makes their data questionable. They used stemming, a technique that removes redundant content from articles. The usefulness of their model without extreme supervision is currently untested. Their model works a lot like counting cards where points are given to words based on the probability they indicate negative or positive sentiment and total scores are used for predictions. Regardless it is an interesting case study in the possibilities of sentiment analysis in financial markets. The technology and algorithms used for sentiment analysis are still in their infancy and if this program incorporated the methodology of the *Currency exchange prediction using machine learning, genetic algorithms and technical analysis* it could better avoid over-fitting and have more general usefulness.

Acknowledgments

Thanks are due Professor Chun for his help in normalizing the data.

Sentiment Analysis on Cryptocurrency News

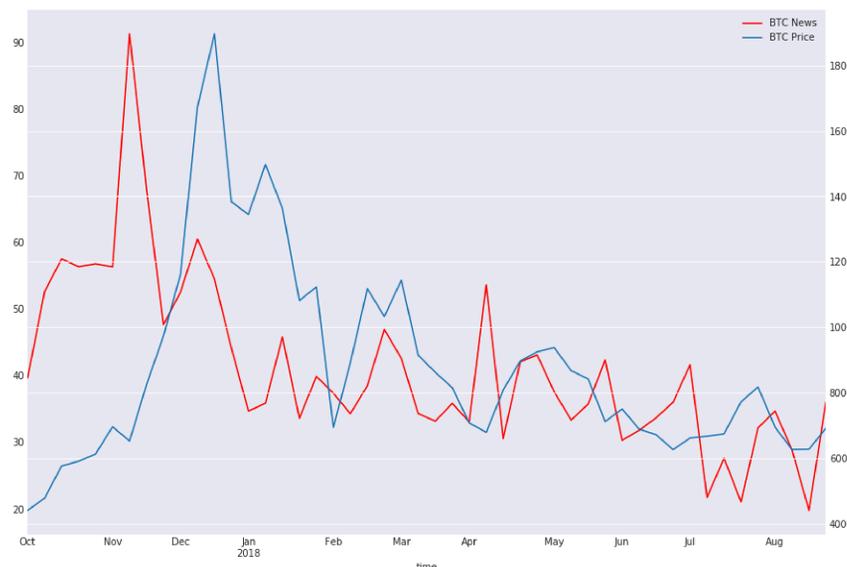
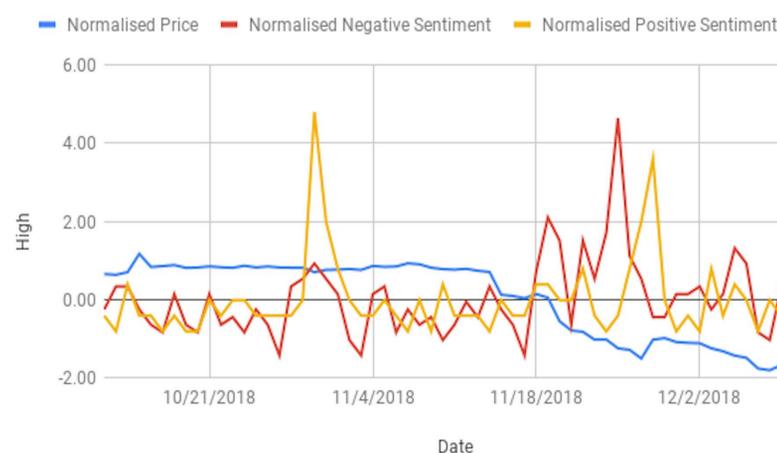


Figure 1. Ben Hunter-Craig's *Sentiment Analysis on Cryptocurrency*

The Aylien data didn't end up supporting my hypothesis. During the plateau in price the number of negative and positive sentiment stories still fluctuated. The highs on the negative sentiment graph correlated with sharper decreases in price. Between November 18th and 25th Bitcoin experienced a steady decline in price and the negativity surrounding the coin increased. There was also a staggered increase in positive sentiment articles during this time with the peak correlating with Bitcoin's first increase in price after a steep decline on November 28th. The correlation of news sentiment to price isn't constant and October 30th is a clear anomaly having the highest number of positive articles of any day despite the continuing pricing plateau.

Aylien Sentiment vs. BTC Price

High vs. Date



Neutral Sentiment from Aylien

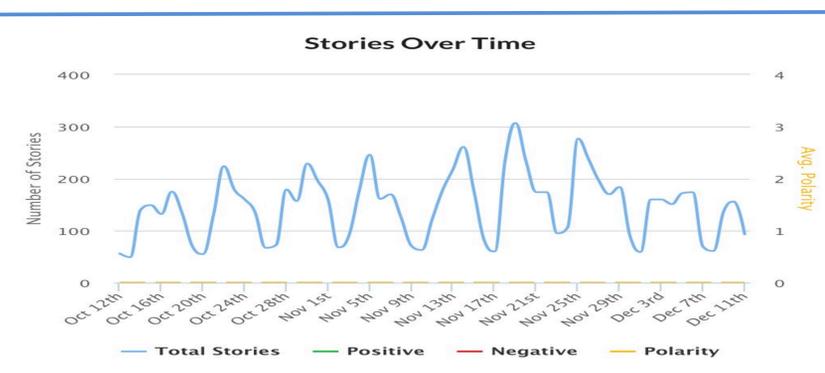


Fig 2. Neutral Sentiment Stories from Aylien

Conclusions and Further Research

My data was collected using the Aylien Api and covers a shorter time frame than I would have liked. Looking at Bitcoin's recent plateau compared to media sentiment suggests that after Bitcoin left mainstream relevance in the year following its peak remaining investors were not influenced by news. A limitation of this thesis is that my data was collected from news sources in English despite Bitcoin being an international currency predominantly held and traded in China and South Korea. Unfortunately I did not have the resources to process data from these countries and given more time and the ability to buy a full license for Aylien API or another news API I would incorporate that data as well. Having a larger and more complete data set would yield more complete results. There are many ways to analyze market performance and find price indicators and media sentiment is definitely a useful tool, but better results could be achieved by combining it with machine learning and genetic algorithms trained on technical analysis of cryptocurrency's performance. With more technical expertise a hybrid model incorporating both technical analysis and sentiment analysis could be created, and this would likely yield better results than either methodology could on its own given that both have been successful predictors of market performance.

Other Contributing Factors to BTC Price

Neutral Sentiment (Fig 2.)

When looking at a graph of articles with a neutral sentiment published over the past sixty days the pattern remains unchanging with periodic dips on the weekends and peaks during the weekdays. This pattern does not correlate with Bitcoin's plateau pricing suggesting that media sentiment is not the best indicator of market fluctuations. Nino Antulov-Fantulin, Dijana Tolic, Matija Piskorec, Zhang Ce, Irena Vodenska's essay *Inferring short-term volatility indicators from the Bitcoin blockchain* used methodology suggesting that Bitcoin pricing is not influenced by the media due to the prevalence of tech enthusiasts that use the currency. Mainstream consumers using Bitcoin as an investment were the target of their research so they filtered out early adopters and holders of the technology by removing people who'd held for over 600 days from their research.

The Effect of Internet Discourse on Bitcoin Pricing
Marvin Aron Kennis' essay *Multi-channel online discourse as an indicator for Bitcoin price and volume* looks at the effect of discussion on internet forums on Bitcoin's price. He found that, "...price and volume movements lead sentiment on forum and Reddit channels at significance levels of 0.05 or below." His findings are interesting because it has the opposite relationship found in *Sentiment Analysis on Cryptocurrency News*. That essay posited that the media influences market movements while Internet forums tend to be influenced by market movements. Reddit comments and posts tended to be negative after a dip in price and positive after an increase.

References

1. News API | Analyze News Content at Scale,
2. Abreu, et al. "Currency Exchange Prediction Using Machine Learning, Genetic Algorithms and Technical Analysis." [Astro-Ph/0005112] A Determination of the Hubble Constant from Cepheid Distances and a Model of the Local Peculiar Velocity Field, American Physical Society, 29 May 2018, arxiv.org/abs/1805.11232v1.
3. Hunter-Craig, Ben. "Sentiment Analysis on Cryptocurrency News - CryptoCompare Research." CryptoCompare Research, CryptoCompare Research, 17 Sept. 2018, blog.cryptocompare.com/sentiment-analysis-on-cryptocurrency-news-d28923b356be.
4. Isah, et al. "Predicting the Effects of News Sentiments on the Stock Market." [Astro-Ph/0005112] A Determination of the Hubble Constant from Cepheid Distances and a Model of the Local Peculiar Velocity Field, American Physical Society, 11 Dec. 2018, arxiv.org/abs/1812.04199v1.
5. Kennis, and Marvin Aron. "Multi-Channel Discourse as an Indicator for Bitcoin Price and Volume Movements." [Astro-Ph/0005112] A Determination of the Hubble Constant from Cepheid Distances and a Model of the Local Peculiar Velocity Field, American Physical Society, 6 Nov. 2018, arxiv.org/abs/1811.03146v1.
6. Nino, et al. "Inferring Short-Term Volatility Indicators from Bitcoin Blockchain." [Astro-Ph/0005112] A Determination of the Hubble Constant from Cepheid Distances and a Model of the Local Peculiar Velocity Field, American Physical Society, 19 Sept. 2018, arxiv.org/abs/1809.07856v1.