Reddit, cryptocurrencies, human irrationality: Did r/cryptocurrency replicate the wisdom of the crowds during the 2017 crypto bubble?

1. Introduction: Contextualization and Motivation

During 2017, Bitcoin became a household name. The cryptocurrency market saw incredible growth and Bitcoin was spearheading this phenomenon, price reaching highs of $19,500, from just $1,000 at the start of the year. As bitcoin grew, it attracted the attention of the world, news stories began appearing as people remortgaged their houses (1) and google trends interest in bitcoin exploded (2). The speculative bubble kept growing until mid-December until it finally popped, leaving many private individuals broke, with some individual traders sadly committing suicide (3). After billions flowed into crypto, billions flowed out just as fast, and many investors were left destitute.

One is compelled to give the famous quote from Joe Kennedy during the 1929 crash: "Taxi drivers told you what to buy. The shoeshine boy could give you a summary of the day's financial news…" (4) this was the great sell signal for him as mania took over people in 1929, and 2017 was no different.

With the speculative nature of 2017 unfolding, are we able to parse through the data to find a group of investors who felt they were 'in the know'? The phenomena of wisdom of the crowds (5) has been discussed at length in many circles, but the essence of the idea is the following:

A. A large group of individuals is likely to outperform an individual expert in estimating things such as weight of a cow, land surface area of a country and perhaps, market movements of a financial instrument.
B. For this to be true, one needs a crowd of diverse opinions, therefore echo-chambers are expected to have no better probability of being right than simple chance.

Previous work has taken place in attempting to use sentiment analysis to directly such as Wooley et al (6). In this paper we see a that the authors succeed in showing that there exists a relationship between reddit community sentiment and price predictions. The success rate was in the range of 73-74% for sentiment tracking price, although the authors used additional methods, we do not cover here such as network building between communities and introducing PageRank to this network. It is encouraging to see that statistically significant results were obtained by other researchers. Our approach will be different and described below, relying more on sentiment alone compared to Wooley but the results presented will show price and sentiment giving a Spearman correlation of 68.2% with a p-value of 5.1×10^{-22}. Our 'naive' sentiment only method gives us similar result for the period of the 2017 bubble.

2. Description of methods

We are selecting two datasets, firstly we focus on extracting sentiment from comments on the r/cryptocurrency forum during the period of November 1st 2017 to March 31st 2018. We conducted the experiments on a small sample size of around 1% of the data which gave good results. After checking all methods worked, running the code on the entire dataset gave even better results for sentiment of the comments. We use the VADER (7) package for our sentiment analysis. VADER provides us with a reliable way to analyse social texts with a rich lexicon for sentiment scoring. We mainly focus on using the compound sentiment score as suggested by the authors due to being the most reliable metric. The results display between -1 to +1, from extremely negative to extremely positive, with the compound metric giving us the best overall interpretation of the text.

As sentiment can be extremely variable over short spans of time, to increate the likelihood we are capturing the current mood of our crowd, we group each sentiment category by day. This given us the sentiment for an entire 24 hour period therefore any small noise on shorter timespans dissipates and we do not need to worry about it.

Below we can look at a graph also available in the notebook:

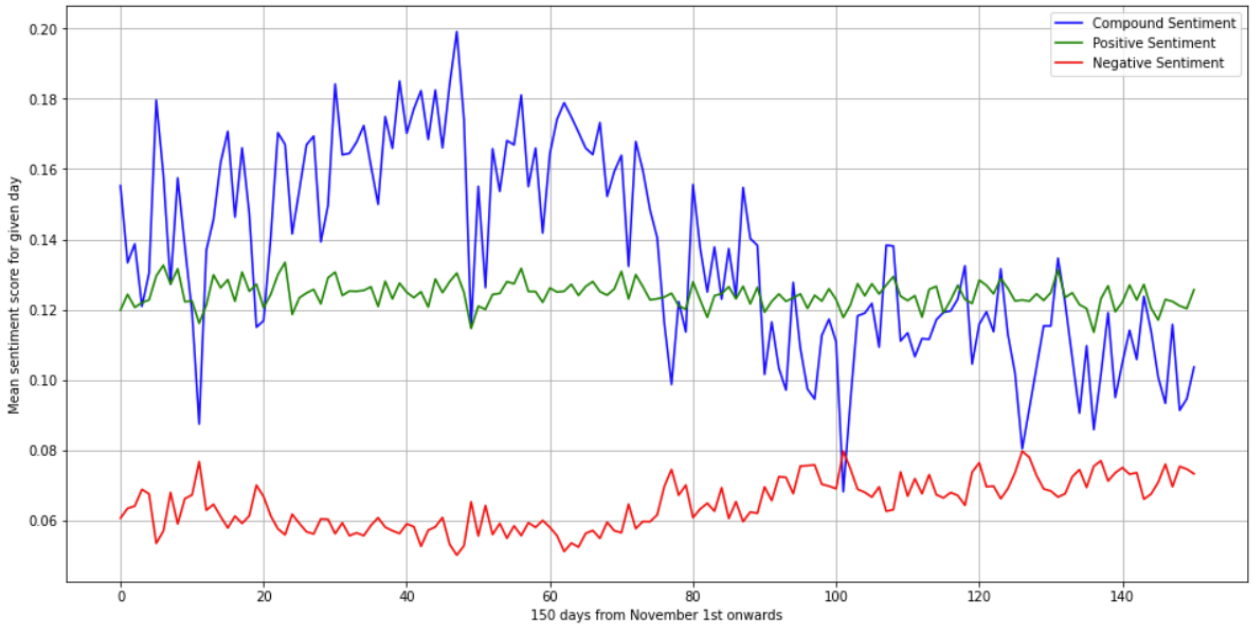


Figure 1: Mean sentiment by day, note change of sentiment at around day 90

We then move onto selecting a 'crowd within a crowd'. As stated by our description of the wisdom of the crowds in A and B, we have a large group of comments in our dataset, this being around 2 million. By selecting only comments which mention the word 'Bitcoin' we can create a second comparison crowd from the first, with around 56,000 comments. We can treat each of these comments as a stand-alone member in our theoretical crowd representing the voice of an individual. In reality, many members comment hundreds of times therefore some voices are stronger than others in our dataset, which is a drawback. By comparing these two crowds we could see their respective 'wisdoms' and explore their predictive power. By using r/cryptocurrency we are attempting to get a diverse range of views to satisfy criteria B, while using other forums such as r/Bitcoin or r/etherium would mean we work on 'echo-chamber' like crowds.

Finally, we focus on financial data from the period in question. Calculating normalised price to compare with normalised compound sentiment to explore the possibility of predictive power and look into the possibility of sentiment reacting to price action.

3. Discussion about experiments

Entire Crowd Result

We start with the overall compound sentiment and compare it to the normalised price of bitcoin. After computing the sentiment, we can graph it over the period in question in figure 2. From the graph we can roughly see the sentiment has a very noisy gaussian structure, while the price is skewed greatly. From December 1st to 31st we can see a huge discrepancy between price and the sentiment on the r/cryptocurrency forum. Roughly speaking we could state that the sentiment is reacting to the new all time high for bitcoin. Another explanation could be that the sentiment is representing the bullish opinion on the prospect of other cryptocurrencies rising soon, as it could be noted that the second biggest cryptocurrency, Ethereum, not reaching it's all-time high for this bull run until middle of January.

Computing the Spearman correlation for these two sets of data gives us a result of 0.68. It should be noted that this is a decent result but does not seem to have much predictive power when focusing on bitcoin reaching it's

all-time high of almost $20,000 in the 2017 bull run. Therefore, it is likely this result is capturing information not discussed at length here. We can see the result in Figure 3.

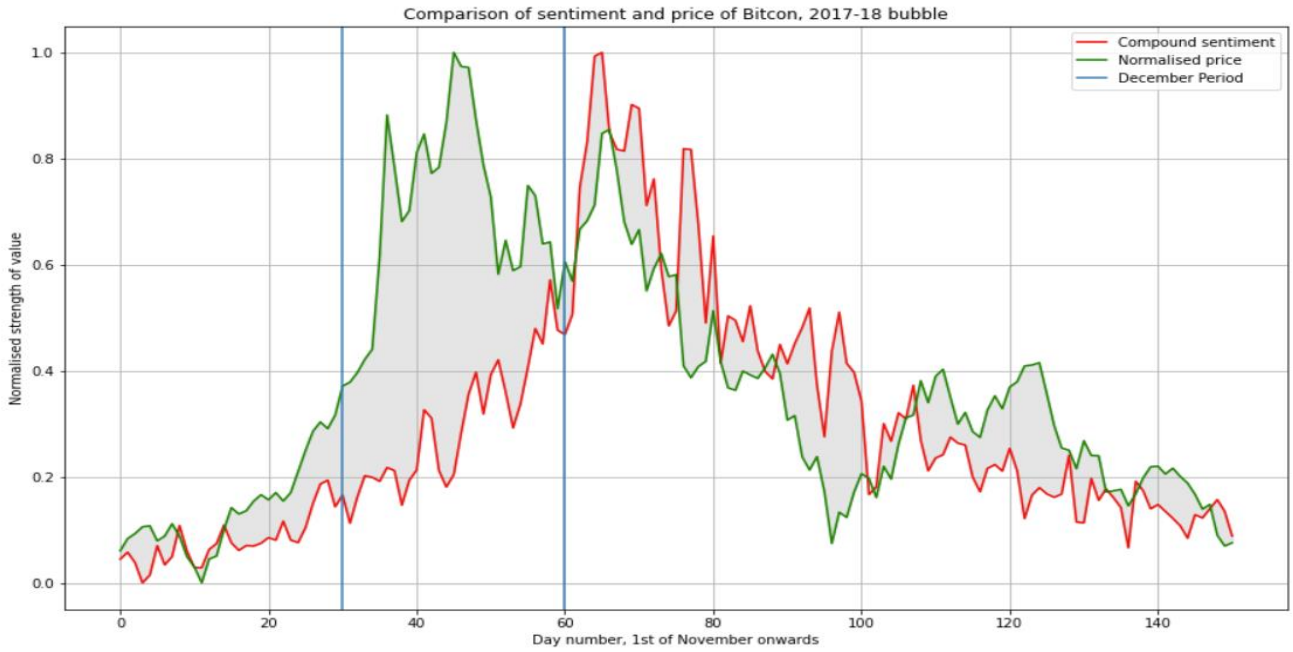
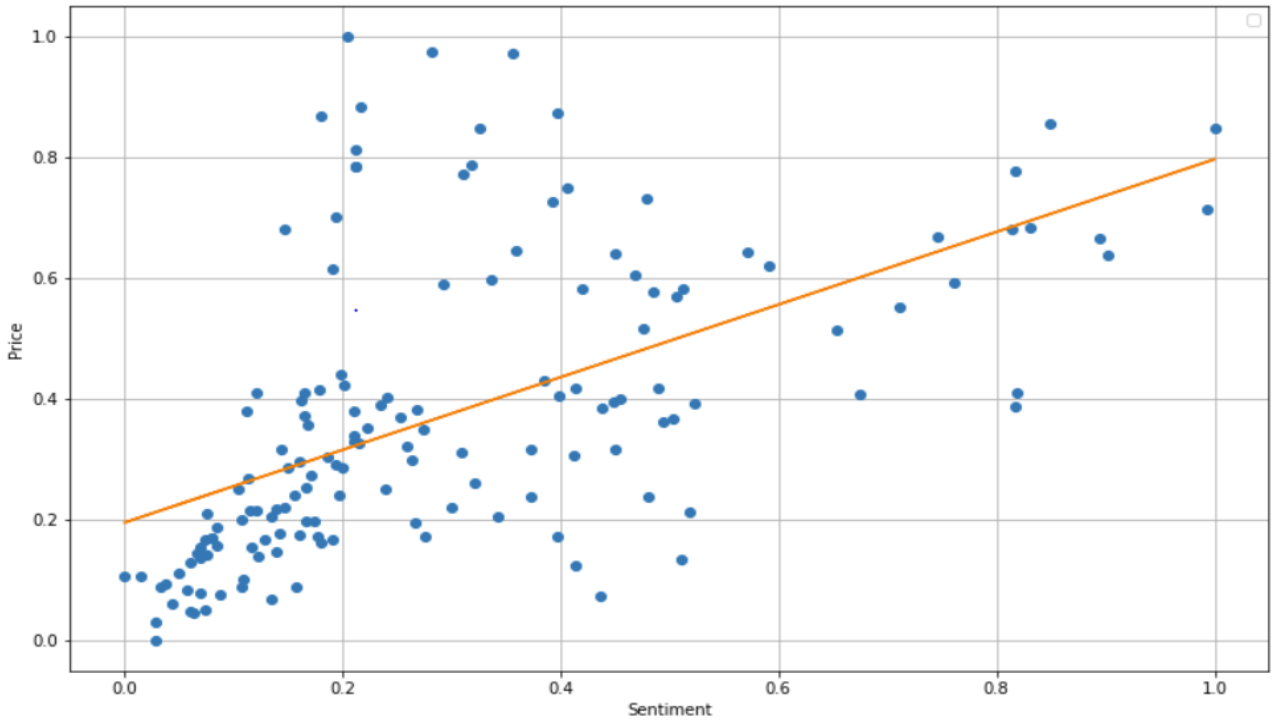


Figure 2: Normalised price with sentiment, shaded area represents difference between values.




```
SpearmanrResult(correlation=0.6822655977692574, pvalue=5.132780458940153e-22)
```

Figure 3: Spearman Correlation result with line of best fit for price and overall crowd sentiment

Bitcoin crowd result

After selecting only comments containing the word Bitcoin we see a different distribution of sentiment, with compound sentiment being far higher for November, before the rapid price increase, following by closely

matching the price. It may appear that the high sentiment is predictive of price action in this sub-crowd as figure 4 displays:

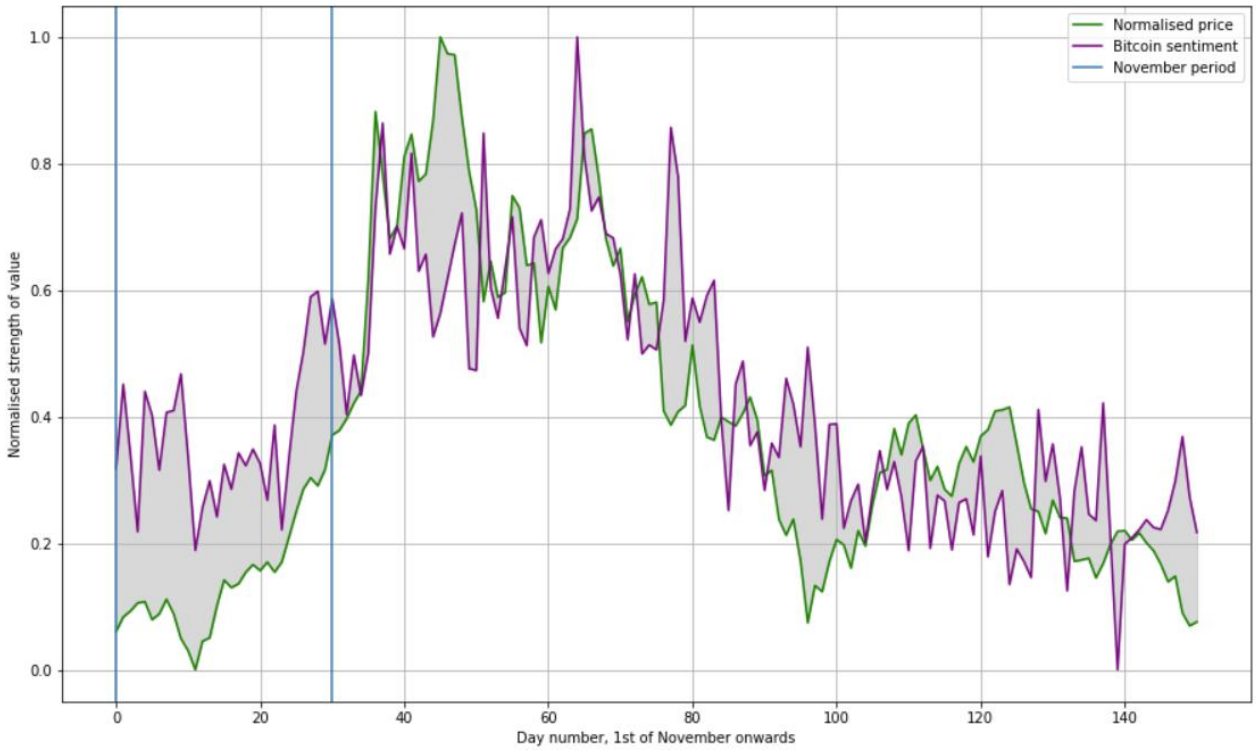
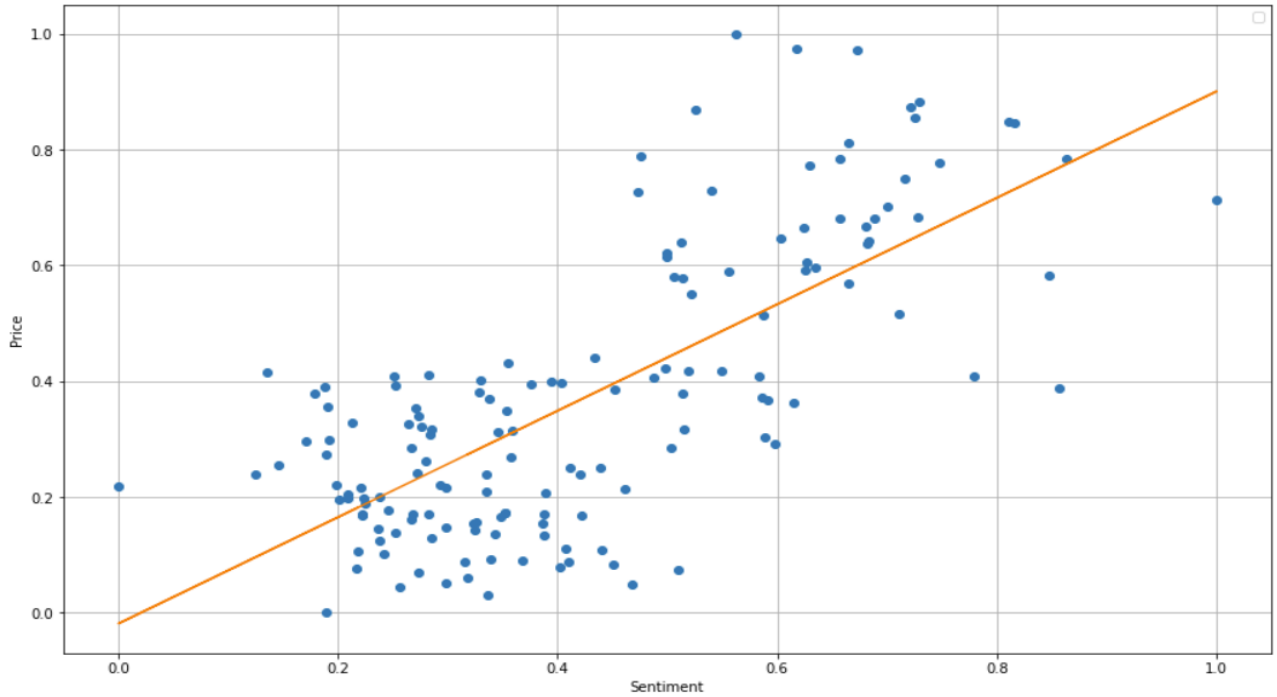


Figure 4: Sentiment when selecting for mentions of bitcoin, note the first 30 days discrepancy.

Let us explore the Spearman correlation in figure 5 before going onto discuss the results:




```
SpearmanrResult(correlation=0.6477134890205646, pvalue=2.516410081181927e-19)
```

Figure 5: Spearman Correlation result with line of best fit for price and bitcoin crowd sentiment

As we can see the Bitcoin crowd's sentiment is far higher for November than the overall crowd. This may be because we are capturing the bullish sentiment from the October price increase, which was from around $4,000 to $6,500 at the end of October. But this may not quite explain it as the November price increase was around $6,500 to $11,000 which is far greater in absolute terms, therefore it is likely that in November we are capturing November's sentiment of the Bitcoin crowd. This higher positive sentiment could be a contributing factor in the December's all time high as this momentum could have pushed the price up.

We can look at figure 6 to see a comparison of the difference in sentiment of both crowds.

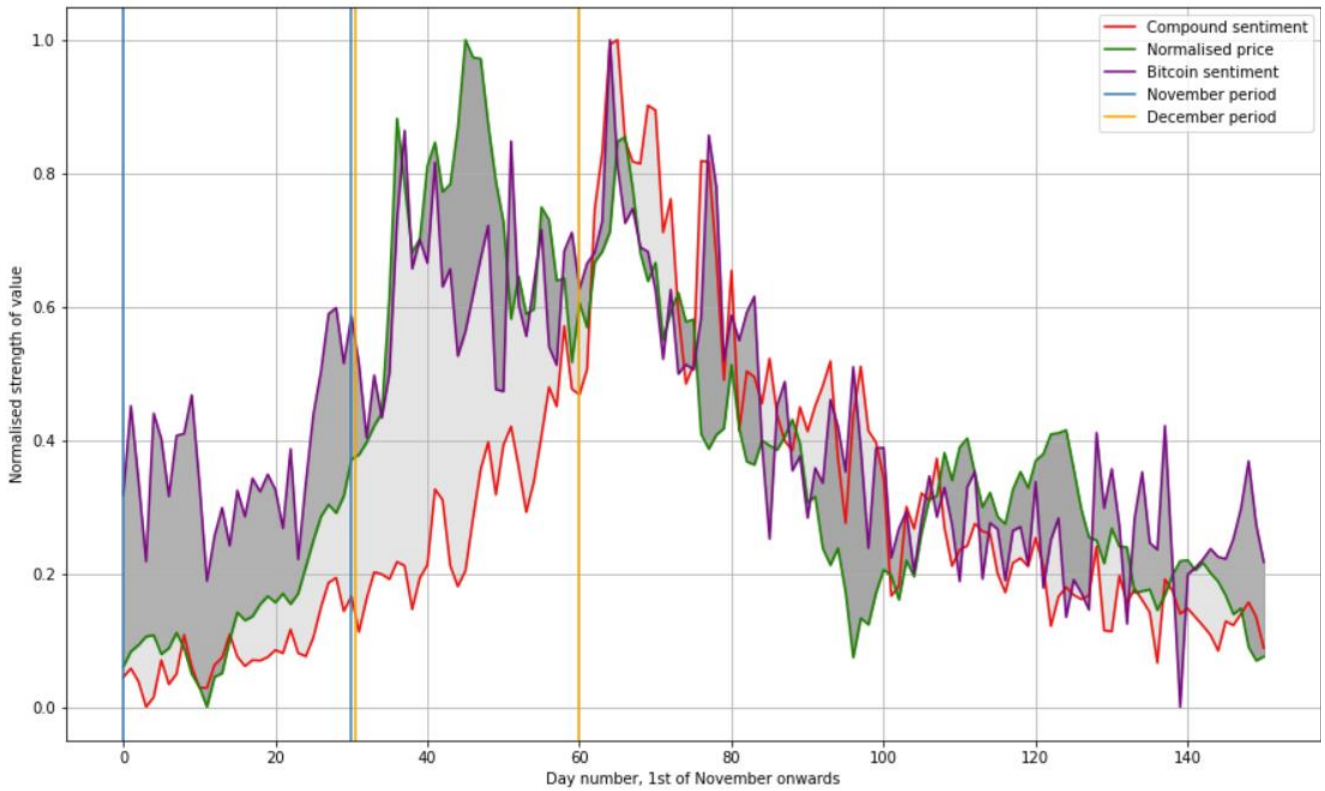


Figure 6: Comparison of sentiments of both crowds, with normalised price

Dark Gray: Area of discrepancy between Bitcoin sentiment and normalised price (mostly overestimate)

Light Gray: Area of discrepancy between overall sentiment and normalised price (mostly underestimate)

In figure 6 we can see this 'predictive versus reactive' dynamic very clearly which can be summarised as the following table when we compare to price action visually:

	Bitcoin Crowd Sentiment	Overall Crowd Sentiment
November Period	Predictive of December price rise	Mostly Tracks or Slightly Underestimates
December Period	Mostly Tracks Or blind to January price drop	Predictive of January price drop Or blind to December price rise
January & after	Mostly Tracks	Mostly Tracks

It should be noted that we could also interpret the overall crowd sentiment in the December period as underestimating the price movement, there are multiple interpretations of this graph given in the table. This is a

multidimensional problem therefore it is difficult to state 'x will happen' by taking the sentiment into account alone. An investor has many options where we can discuss two of them.

Let us suppose a hypothetical investor stands before our two crowds and decides to use a dual strategy, to maximise wisdom of the crowd but also minimise risk. We would want to follow the Kelly criterion (8) and consider the new information the crowd is telling us. How could these hypothetical events unfold?

November events

1. **Information**: Overall crowd tells us the price should be a little lower or is roughly right for them.
2. **Information**: Bitcoin crowd tells us the price is far too low, at no point does the price make them happy.

Two actions available for our investor:

A. **Invest**. Overall crowd is roughly neutral, maybe tilting negative, bitcoin crowd is very positive.
B. **Do not invest**. Null action, terminate this branch of thought experiment.

December events

1. **Information**: Overall crowd really does not like this price; at no point would they say to buy.
2. **Information**: Bitcoin crowd is mostly ok with price, except with deviation at all-time high in December

Three actions available for our investor:

A. **Invest more**: This action goes against overall crowd and is at best neutral with respect to bitcoin crowd, with a slight tilt to not being supported.
B. **Keep Investment as is**: not supported by overall crowd, bitcoin crowd mostly neutral to this action.
C. **Sell**: Highly supported by overall crowd, bitcoin crowd neutral to this action, possibly slightly tilting in support. Results in biggest gain, action terminates actions of our investor.

January and onwards events

1. **Information**: sentiment of both crowds roughly follows the price action, cannot easily act on this information or lack of it.

Three actions available for our investor:

A. **Invest more**: This investment would be a gamble, no better than coin flip, not informed by crowds. Results in bigger loss.
B. **Keep investment as is**: This is a position that is roughly neutral in terms of information from both crowds, but this action results in loss
C. **Sell**: Mostly neutral with respect to crowds.

The actions taken by our hypothetical investor if using both Kelly criterion and sentiment analysis would be buy, sell, do nothing with respect to the actions outlined. But how many investors are rational, and this hypothetical assumes countless things, but it is interesting to consider, nonetheless.

4. Conclusions

In this report we consider how to analyse the historical cryptocurrency bubble of 2017. Using sentiment analysis we show that separating our dataset in two and considering the wisdom of the crowds phenomena, an investor could be better informed of the risks involved in these speculative assets. Further research could include larger dataset for sentiment and separation into more crowds. Further we could consider how to implement the hypothetical trader and the actions taken by writing a mathematical model to see how different strategies preform. This could potentially help inform retail traders and stop them from making the same mistaken in the 2021 bitcoin bull run by learning from 2017.

Wordcount – 2017 words

Bibliography

1. **Fox, Michelle.** People are taking out mortgages to buy bitcoin, says securities regulator. *CNBC.* [Online] 12 11, 2017. https://www.cnbc.com/2017/12/11/people-are-taking-out-mortgages-to-buy-bitcoin-says-joseph-borg.html.

2. **Google.** Google.com. [Online] https://trends.google.com/trends/explore?date=today%205-y&q=bitcoin.

3. **Voon, College Student Commits Suicide After Losing Money On Cryptocurrency - Sarah.** *https://www.asianmoneyguide.com/.* [Online] [Cited: 05 10, 2021.] https://www.asianmoneyguide.com/student-commits-suicide-cryptocurrency.

4. **Rothchild, John.** WHEN THE SHOESHINE BOYS TALK STOCKS IT WAS A GREAT SELL SIGNAL IN 1929. SO WHAT ARE THE SHOESHINE BOYS TALKING ABOUT NOW? *Fortune.* 1996.

5. **Surowiecki, James.** *The Wisdom of Crowds.* New York : Anchor Books, 2005.

6. *Extracting Cryptocurrency Price Movements From the Reddit Network Sentiment.* **Stephen Wooley, Andrew Edmonds , Arunkumar Bagavathi , Siddharth Krishnan.** Charlotte : UCNN, 2019.

7. **Hutto, C.J. & Gilbert, E.E.** VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. [Online] Eighth International Conference on Weblogs and Social Media , 06 2014. https://github.com/cjhutto/vaderSentiment#citation-information.

8. *A New Interpretation of Information Rate.* **Kelly, J.L.** s.l. : AT&T Reproduction of 1956 paper, 1956.