# Social Media Research

There are a number of ways in which social media data can be analysed. Various factors such as volume of posts, sentiment and subjectivity of posts could all be measured for a potential correlation with the price of Bitcoin.

## Twitter

Upon researching the topic of cryptocurrency prices relative to social media activity, I quickly became aware that numerous studies show that ‘the number of tweets is a significant driver of next day trading volume and realized volatility’ (Shen, Urquhart and Wang, 2019).

One significant research article (Burnie, A. and Yilmaz, E, 2019) considered various studies on the relationship between social media posts and Bitcoin metrics. One such study assessed the sentiment and uncertainty of posts on Twitter and found a ‘significant positive correlation between emotional tweets and the close price’ (Kaminski, 2014).

Other studies found ‘a positive mood on Twitter predicted a bitcoin’s price rise 3–4 days later [[4](https://royalsocietypublishing.org/doi/full/10.1098/rsos.191068#RSOS191068C4)], and a positive Twitter sentiment ratio had a positive short-run impact on bitcoin prices [[5](https://royalsocietypublishing.org/doi/full/10.1098/rsos.191068#RSOS191068C5)]. Garcia & Schweitzer [[7](https://royalsocietypublishing.org/doi/full/10.1098/rsos.191068#RSOS191068C7)] found that a higher valence in Tweets preceded increased opinion polarization and exchange volume, which occurred before rises in price. The higher the valence, the greater the degree of pleasure over displeasure in an emotional experience. Polarization measures the extent to which both positive and negative sentiment tweets occurred together. Price drops further led to increased polarization [[7](https://royalsocietypublishing.org/doi/full/10.1098/rsos.191068#RSOS191068C7)].’ (Burnie, A. and Yilmaz, E, 2019). Another source states that ‘a significant relationship with future Bitcoin’s price and volume of tweets exists on a daily level.’ (Matta, Marchesi and Ilaria Lunesu, 2015). Because of this, I will also be taking volume of posts into account.

## Assessing Sentiment & Volume

Many of the existing studies do not describe exactly how they measure the sentiment and subjectivity of posts. There are a number of libraries available for Python which will analyse the sentiment of text using a dictionary of words and associated scores. TextBlob is one of the most popular sentiment analysis libraries for Python.

I will be using TextBlob because it works similarly to Kaminski’s method of deriving two separate scores: sentiment and subjectivity, and Kaminski was successful in finding a significant correlation. As well as TextBlob I will include my own library of cryptocurrency buzzwords for the app to check for, as these words would not be given appropriate sentiment scores by the TextBlob library (in the context of cryptocurrencies).

<https://textblob.readthedocs.io/en/dev/>

The Twitter bot will continuously stream Tweets under the hashtag: #Bitcoin. The application will calculate the sentiment and subjectivity for every Tweet and display this as an average at a set interval – currently every 5 minutes. It will also calculate the volume of Tweets and display the percentage change in volume compared to the previous update. The user could then choose whether or not to consider this information when making investment decisions. This method is beneficial as it allows every new Tweet to be assessed while the program is running, rather than a small sample.

## The Twitter API

Twitter API documentation - <https://developer.twitter.com/en/docs/basics/getting-started>

Twitter requires you to fill out an application in order to use their API and explain how and why you are using the data.A screenshot of a cell phone

Description automatically generated

Once the application was approved, I created a twitter application using twitter’s website:

A screenshot of a computer screen

Description automatically generated

I created a basic python application to interact with the Twitter app and store the posts in Azure table storage. This was done using a Python framework called ‘Tweepy’ to iteract with Twitter’s API:

A screenshot of a computer screen

Description automatically generated

Twitter’s API allows you to search for the last X number of posts for a given hashtag.

The API has a feature for iterating through posts for specific users. This could allow me to choose a range of influential cryptocurrency accounts and iterate through their tweets until I find the required date.

There is a streaming feature which allows a bot to stream all Tweets for a given hashtag. This would be useful for harvesting live data but would require a long-running application.

The API can also be used to post Tweets using the bot account.

Unfortunately, Twitter’s API is quite restrictive, and there is no option to search for specific dates. This makes getting older Tweets very difficult, as the only way to do so involves getting all Tweets for a specific user and iterate through them. There is no option to do this for hashtags though. Twitter also maintain a rate limiting policy, meaning if their API is used too frequently, Twitter limits its service to your account.

I believe that the most beneficial use of Twitter’s API is to provide users with live updates on user activity on Twitter. Twitter is well suited to this sort of data gathering as it offers a streaming function in their API. The streaming feature is not perfect as it is subhect to rate limiting, and Twitter may choose to lock the account, but if this happens it can easily be solved by logging back into the account. The bot can also be used to post live update Tweets which will allow for displaying information on a Twitter feed.

## The Reddit API

[*http://www.reddit.com/dev/api*](http://www.reddit.com/dev/api) - link to Reddit’s API

<https://www.reddit.com/prefs/apps/> - link to create app

*A screenshot of a computer

Description automatically generated*

IMPORTANT: Reddit allows a maximum of one request every 2 seconds – or 30 per minute.

<https://docs.microsoft.com/en-gb/azure/app-service/containers/quickstart-python?tabs=bash>

A screenshot of a computer screen

Description automatically generated

PRAW offers the ability to search for top posts in the past day/week/month/year etc. This means that it would still be effective for providing recent updates (such as top post from the past week). PRAW could also be used to iterate through posts in a subreddit and using Python to check the dates of these posts, only the necessary posts would be stored. There is another API for Reddit called Pushshift, which allows for more advanced requests and potentially even searching dates.

|  |
| --- |
| BIBLIOGRAPHY |
| Burnie, A. and Yilmaz, E. (2019). Social media and bitcoin metrics: which words matter. *Royal Society Open Science*, [online] 6(10), p.191068. Available at: https://royalsocietypublishing.org/doi/10.1098/rsos.191068 [Accessed 1 Feb. 2020]. |
| Kaminski, J. (2014). *Nowcasting the Bitcoin Market with Twitter Signals*. [online] arXiv.org. Available at: https://arxiv.org/abs/1406.7577 [Accessed 1 Feb. 2020]. |
| Matta, M., Marchesi, M. and Ilaria Lunesu, M. (2015). Bitcoin Spread Prediction Using Social And Web Search Media. In: *Workshop Deep Content Analytics Techniques for Personalized & Intelligent Services*. pp.1-5. |
| Shen, D., Urquhart, A. and Wang, P., 2019. Does twitter predict Bitcoin?. *Economics Letters*, 174, pp.118-122. |