Bitcoin price fluctuation prediction using Twitter Sentiment Analysis

Jeevan Reddy Rachepalli

Bitcoin

Bitcoin is a decentralized electronic currency system, which brought an enormous change in the financial system.

Bitcoin has grown tremendously and has managed to attract large number of users and has gained huge popularity due to its frequent mention and propagation in the media.

Due to its popularity, the Bitcoin price, which fluctuates constantly on real-time like a stock exchange, it is very curious to build a model that can predict of the price of Bitcoin on real-time using the social media data from the internet.

Figure: Bitcoin price change/trend between January 2012- March 2018:
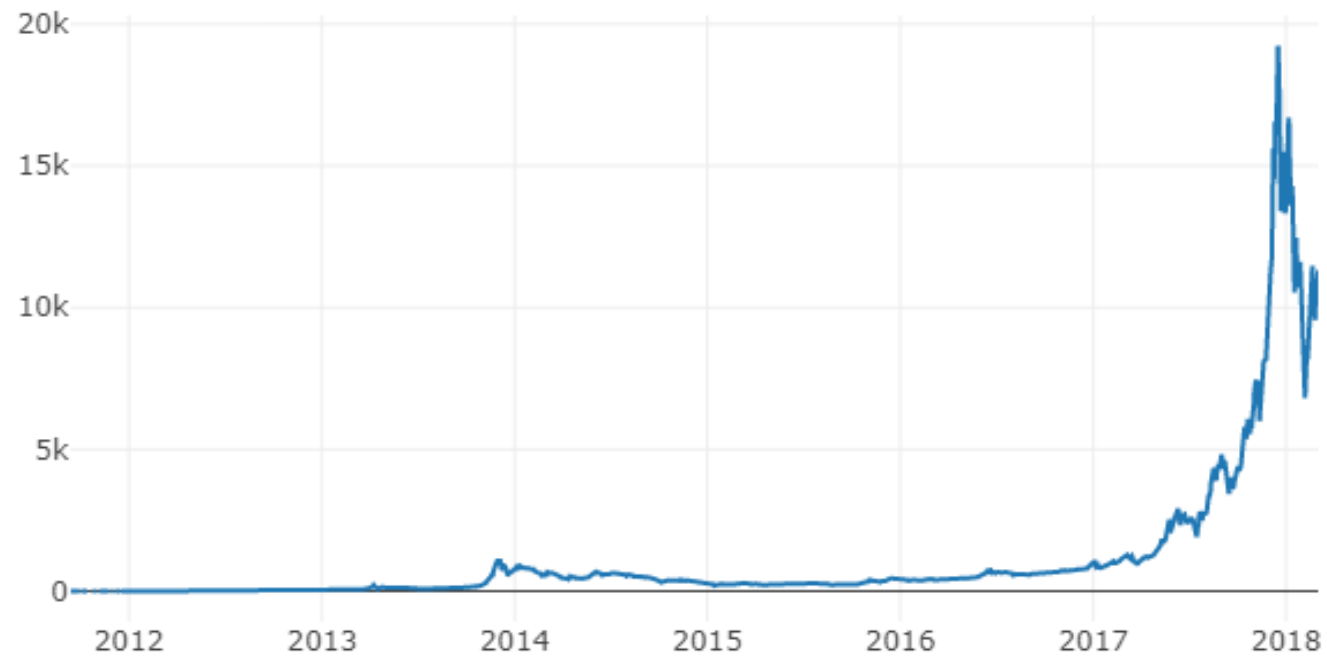
My research question

Hundreds of thousands of Twitter users generate huge volumes of tweets data every day related to Bitcoin.

This huge data can be helpful to study the trends in Bitcoin using technologies like Machine Learning, Natural Language Processing, Time Series Analysis etc.

Our study might fill the gap by building Emoji based Sentiment analysis for effective capture of Twitter Sentiment.

Bitcoin Price Data Collection

We have used Quandl module API of Python for collecting the Bitcoin Price on that particular day.

We have collected the Bitcoin price data from 4 major websites- "BITSTAMP", "COINBASE", "ITBIT", "KRAKEN" that tracks the Bitcoin Price.

Out of the "Opening Price", "Closing Price", "High Price", "Low Price" of Bitcoin in a day we have selected only the "Close Price" in a day of Bitcoin.

Models used for the sentiment analysis:

AFFIN Sentiment Classifier

TextBlob Sentiment Classifier

VADER Sentiment Classifier

Emoji Classifier

AFFIN Sentiment Classifier

In AFFIN sentiment analysis, we use AFFIN, which is a list of English words, which are scored with an integer ranging between -5 and +5. In the twitter sentiment analysis, using AFFIN we would be evaluating the overall average sentiment/polarity score (positive or negative) for the extracted text from the tweets data.
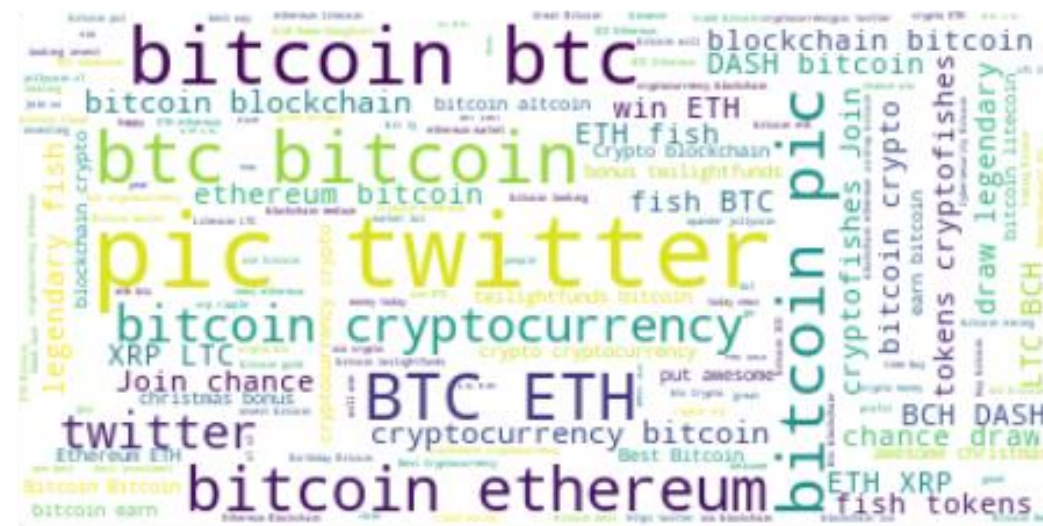
Positive Vs Negative Word Cloud

TextBlob Sentiment Classifier

In the TextBlob sentiment analysis, by feeding the unique tweets, we obtain polarity as the output that ranges between -1 to +1.

So, a tweet has Positive sentiment when it's polarity is greater than 0 and negative sentiment when it's polarity is lesser than 0.

When the sentiment polarity is exactly 0 the tweet is said to have neutral polarity. The TextBlob also gives the Subjectivity of a tweet

Positive Vs Negative Word Cloud

VADER Sentiment Classifier

VADER (Valence Aware Dictionary and Sentiment Reasoner) Sentiment Classifier.

This works just like the AFFIN as it also has word weights ranging from positive to negative but this VADER is designed in such a way that it is aware of the Social media jargon used by its' users.

Positive Vs Negative Word Cloud

Emoji Classifier

Happy Emojis

🙂 😁 😄

happy_set = set([":)",":-)","=)"])

Sad Emojis

☹️ 🙁 😠

sad_set = set([":(",":-(","=("])

Positive Vs Negative Word Cloud

Getting Effective Sentiment of a tweet

We have classified the effective tweet polarity as +1 if a happy emoji is present in it and -1 if a sad emoji is present in it.

Out of the 3 sentiment scores we have defined a function that returns 1 if a Happy emoji is present in a tweet and -1 if a sad emoji is present in a tweet.

It also checks the sum of all sentiment scores of 3 classifiers. If it is greater than or equal to 1 then the effective polarity is given 1 and if it is less than or equal to -1 then the tweet is given an effective polarity of -1.

For all other cases, it is given a 0.

Resampled the time series data with their mean on a particular day. So, on a day we have the Bitcoin average price and average Sentiment polarity on Twitter. As it is a time series data we can calculate the percentage change in the price of Bitcoin data with respect to the previous day. It can be Positive or Negative

RESULTS AND FUTURE WORK

Observed that the Bitcoin percentage change and the Twitter sentiment are positively correlated. The Orange color graph is the scaled percentage change of Bitcoin over a day. The Blue color graph is the effective sentiment polarity of Twitter on Bitcoin within a given day. We can also say that there is a time shift between the Bitcoin Price fluctuation and Twitter Sentiment on Bitcoin. The Twitter sentiment lags in the Bitcoin Price fluctuation within a time frame. That's why our correlation between these two is around 0.17.
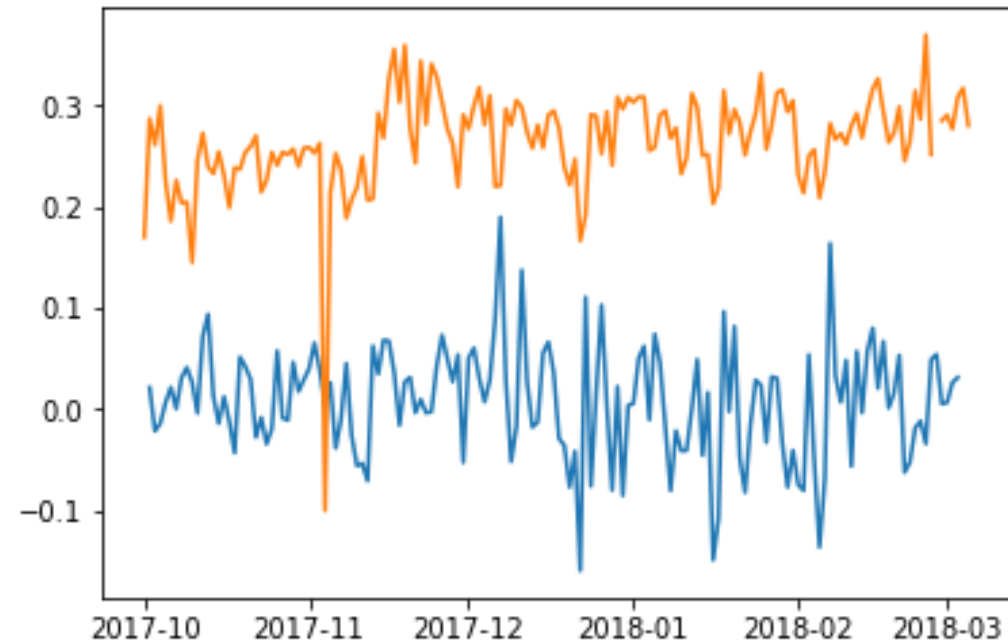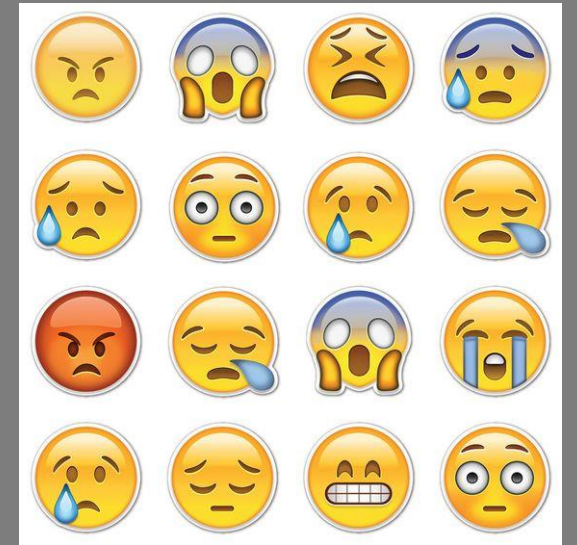
If we have taken these emojis into account our emoji contained tweets count would have been increased significantly from mere 11,000 (out of 2.5 Million tweets).

Thank You!