

Bitcoin price prediction using sentiment analysis on social media

Team:

Harish Puvvada - hp1047

Rohit Suvarna - rus209

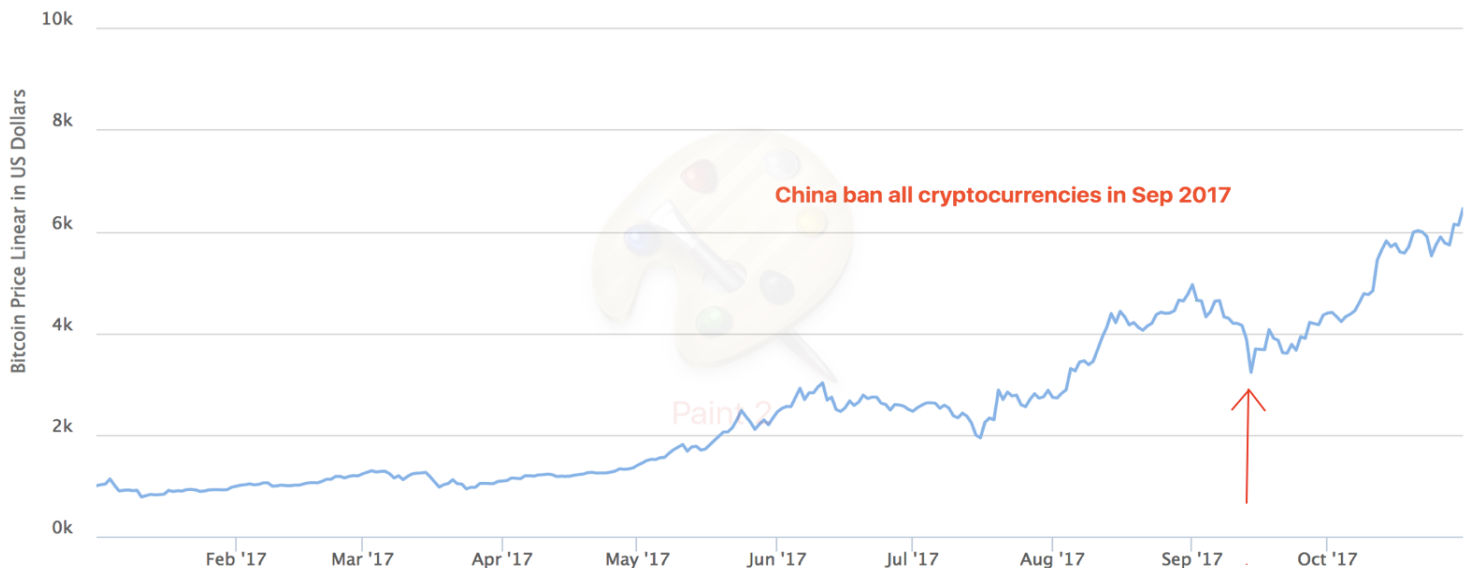
Manoj Sundar Balaji - msb662

Janeil Atul Patel - jap813

Motivation:

Cryptocurrency is a decentralised and secure form of digital cash transactions which has taken the world by storm. Cryptocurrency is largely driven by young tech savvy millennials who are extremely active on social media. Bitcoin just like any other traded currency relies heavily on market speculation. Apart from the traditional sources of media namely news articles and television with the rise of Bitcoin we see a new source of information that is social media.

Bitcoin being extremely volatile makes it harder to predict. We aim to use the Efficient Market Hypothesis (EMH) which states that stock market prices are largely driven by new information and follow a random walk pattern. By incorporating social media as an additional channel of information we can better emulate the speculative patterns of the traders. For example the fall in prices of bitcoin with the Chinese regulator's decision to ban all virtual currency and its subsequent rise when the government decided to soften its stands.



Abstract:

The project attempts to predict the future value of Bitcoins by identifying the correlation between social media sentiment and market sentiment. We will achieve this by collecting user feeds from social media such as twitter, facebook and linkedin. Once we have our corpus we will map their associated sentiments using IBM Watson's Natural Language Understanding API. While mapping sentiments to our corpus we attempt to capture granular level categories namely joy, anger, happiness, etc. We use these as feature vectors to our ML/DL algorithms. Then we compare the results of the different algorithms and choose the one with the best accuracy score.

Technologies:

Spark ML

Pig Latin

Hadoop Mapreduce

MongoDB(optional)

IBM Watson Natural Language Understanding Api

Libraries: Pandas, Matplotlib, Scikit learn, TensorFlow

Programming Languages: Python, Java, Ruby

Data Sources:

Twitter Api to get the tweets about BitCoins/Cryptocurrencies.

LinkedIn Api to get the corpus on blogs.

Web Scraping to get data from News articles.

References:

J. Bollen and H. Mao. Twitter mood as a stock market predictor. IEEE Computer, 44(10):91–94.

<http://dataconomy.com/2014/07/bitcoin-big-data-can-predict-future-value-virtual-currency>

Mittall et Goel. Stock Prediction Using Twitter Sentiment Analysis