Proposal

Members

- Haruki Moriguchi (# 260665818) : *haruki.moriguchi@mail.mcgill.ca*

- Edwin H. Ng (# 260732345) : *edwin.ng@mail.mcgill.ca*

Predicting Bitcoin Price based on Sentiment Analysis on News and Social Media

In recent years, Bitcoin has gained enormous popularity. After receiving many media coverages in 2017, the price went up drastically from USD 1,000 to USD 20,000, from which it has since gone down. In fact, this sudden increase in price is not surprising since behavioral economics states that there are correlation between the public sentiment and the financial market. Fortunately, with the advent of social media, the information about public feelings has become abundant, where Twitter has received a lot of attention from researchers. The aim of this project is thus to test the hypothesis that, in addition to past historical prices, the public sentiment also does play an influence to the stock market. We will look particularly for the case of the Bitcoin prices. The plan of the project is then as follows. First, public sentiment will be gathered and pre-processed from tweets on Twitter, where instead of using the N-gram model for their textual representation, we are going to train a word2vec representation on them since the latter allows for sustainability in word meaning across different contexts. Then, we are going to train different classifiers and choose the best sentiment analysis model to apply on tweets related to Bitcoin or cryptocurrencies. Once sentiment analysis on Bitcoin tweets has been assessed, we are then going to extract its past historical price movements. The idea is that the future movement of a stock price should mainly be able to reflect its past tendencies, but that the public opinion will certainly also influence its future paths. Therefore, we will then train a final machine learning model that will classify the Bitcoin price movement as being "up" or "down" at a certain time based on both its past historical movements, and on its public opinion at that time, which we would have obtained based on the sentiment analysis as described above.

Public Sentiment

- word2vec representation for tweets

- to determine the sentiment of a tweet, list out positive and negative keywords and see if the tweet contains those keywords.

- LDA for finding topics in news articles

- If the main topic of a news article is cryptocurrency/Bitcoin, include it into our corpus

- for each news artcle, determine the sentiment by using the positive and negative keywords approach

Predicting Bitcoin Price Movement

1. Choose a time frame (minutes, hours, days, etc)

2. For each time frame, determine the public sentiment.

3. To do this, take all the tweets and articles within that time frame, and the public sentiment will be the majority of sentiments.

4. See if the Bitcoin price increased or decreased within that time frame

5. Use (time, sentiment, price movement) as input for different models (SVM, neural networks, etc) to predict the rise and fall of the price.

6. Compare outcomes of the different models