Unit 12—Tales from the Crypto

Stock Sentiment

Background

There's been a lot of hype in the news lately about cryptocurrency, so you want to take stock, so to speak, of the latest news headlines regarding Bitcoin and Ethereum to get a better feel for the current public sentiment around each coin.

In this assignment, you will apply natural language processing to understand the sentiment in the latest news articles featuring Bitcoin and Ethereum. You will also apply fundamental NLP techniques to better understand the other factors involved with the coin prices such as common words and phrases and organizations and entities mentioned in the articles.

Complete the following tasks:

1. Sentiment Analysis
2. Natural Language Processing
3. Named Entity Recognition

Files

Starter Notebook

Instructions

1 - Sentiment Analysis

Use the newsapi to pull the latest news articles for Bitcoin and Ethereum and create a DataFrame of sentiment scores for each coin.

Use descriptive statistics to answer the following questions:

> Which coin had the highest mean positive score?
>
> Which coin had the highest negative score?
>
> Which coin had the highest positive score?

2 - Natural Language Processing

In this section, you will use NLTK and Python to tokenize text, find n-gram counts, and create word clouds for both coins.

Tokenize

Be sure to:

1. Lowercase each word.
2. Remove punctuation.
3. Remove stop words.

N-grams

Next, look at the ngrams and word frequency for each coin.

1. Use NLTK to produce the ngrams for N = 2.
2. List the top 10 words for each coin.

Word Clouds

Finally, generate word clouds for each coin to summarize the news for each coin.

btc-word-cloud.png

eth-word-cloud.png

3 - Named Entity Recognition

In this section, you will build a named entity recognition model for both coins and visualize the tags using SpaCy.

btc-ner.png

eth-ner.png

Resources

Vader Sentiment Analysis

Hints and Considerations

The free developer version of the News API limits the total daily requests, so be careful not to exceed the free limits.

Submission

- Use the starter Jupyter Notebook provided to conduct the NLP analysis and host the notebook in a GitHub repository.

- In your GitHub repository, include a ReadMe file that uses Markdown to summarize your homework.

- Submit the link to your GitHub project to Bootcamp Spot.