

Crypto Sentiment Analysis with Twitter

Garret Tullio

Sentiment analysis for the cryptocurrency market with Twitter.

This project is to establish a deep learning model that can predict the sentiment of the cryptocurrency market from Twitter. Since Twitter is a congregation for opinionated traders and investors, it is the perfect platform to gauge the attitude of the market. The deep learning model uses a neural bag-of-words for vectorization of data and a convolutional neural network and Multilayer perceptron for the structure of the models.

Introduction

The cryptocurrency market has been gaining deep traction in the last year due to several factors such as; institutional adoption, scarcity, and as a hedge against inflation. In the short term market sentiment deeply effects the current prices of an asset. If there is positive sentiment in the market there is higher probability of the price rising and negative sentiment would drive prices down. Twitter is the main hub for individuals expressing the condition of the market, which makes it a great choice for predicting the sentiment of a market as well as gathering large quantities of data for an accurate model.

Results

The results of running through both models with the test and train data set went well. For the Multilayer perceptron model it scored a 98.81% accuracy on the training data set and a 95.00% accuracy on the test data set. For the single channel convolutional neural network it scored a 90.74% accuracy on the training data set and a 90.00% accuracy on the test data set. When generating a few sample tweets I used a positive sentiment tweet: "crypto will definitley go up. It has a bright future! #bitcoin". Both models predicted positive which was correct. The other sample tweet I used that had a negative sentiment was; "crypto will die soon!!! Please dont buy". Both models accurately classified them as a negative sentiment.

Conclusion

When finishing both models and evaluating them it was clear that the Multilayer perceptron model performed better that the convolutional neural network. This may be due to the structures of both models. In this certain circumstance the Multiplayer perceptron model would be a better choice for this particular problem. When trying both models with different filtering data sets it made a significant difference when gathering good quality tweets. A lot of the times when I used hashtags for filtering I would get a lot of crypto startup promotion tweets which are essentially useless for this model. When using filtering such as "crypto will", I ended up getting a lot better tweets that contained actual opinions of the market.

Data

To gather the data I used Twitters Tweepy API, which is an API used for positing/gathering tweets. What I used for filtering the tweets was I looked for tweets that contained the sequence of words "crypto will", "crypto wont", "crypto can", "crypto cant", "crypto is", "crypto isnt", and "#crypto". After a lot of different filtering I found these keywords to be the most effective in accessing opinions about the Crypto market. I collected 20,000 tweets for the model which took about roughly 2 hours to collect. Once the tweets were collected I cleaned each tweet by removing punctuations, stop words, non-vocab words, and rare words. I split the data into two sets. A training data set and a test data set. The training data set consisted of 19,900 tweets and the test data set consisted of the last 100 tweets. I used VADER to classify the training and test data set as either positive or negative and saved the classification in a text file called classify.txt which was later used for fitting the model.

Application

When building this project I had time to think of an actual application of this project. I believe the best project would be to use sentiment analysis for the market that can be used for investing and trading. When individuals want to invest or trade they look at market indicators such as price and indexes. They could also use this as an indicator for contributing to their decision to buy or sell.

The actual application could be to gather all of the tweets from the beginning of the day to the current moment and clean then feed them into the models. Then get the total percentage of positive and negative sentiments. If there are is a larger percentage of positive sentiments that the application would determine an overall positive sentiment in the market and vice versa.

Resources

<https://catanacapital.com/blog/sentiment-analysis-stock-market-sentiment/>
<https://blog.coinpriceinc.com/2021/08/15/4-reasons-why-bitcoin-cryptocurrency-prices-are-on-the-rise/>



Implementation

Once all the tweets were collected I used the neural bag-of-words for vectorization and created two models. The first is a multilayer perceptron model and the second model is a single channel convolutional neural network.

