Grant McAllister
Taylor Adams
Gil Kiesler

The correlation between Cryptocurrency market trends and the corresponding Twitter sentiment analysis

Introduction

Cryptocurrencies have been serving as a new global way of investment, with incredibly high returns that exceed the returns of assets such as stocks or commodities. However, Cryptocurrencies are also associated with potential financial instability and considerable risk.

This paper discusses the correlation between the value of Cryptocurrencies (Bitcoin, Ethereum and Shiba Inu) and the sentiment of their appearances in 'Tweets' on the social media platform Twitter. Twitter was found to be the most reliable social media platform for this research with the ease of filtering comments to provide the most relevant data for our analysis. Moreover, the "hashtag" function for Twitter posts makes it easier to identify the topic of each Tweet, which eases the burden of data filters to only information relevant to our research. Therefore, we will be looking into if a correlation exists between the price of cryptocurrency and the sentiment analysis, if so, what type of correlation exists between the sentiment of Tweets and the value of said Cryptocurrency along with the number of tweets.

The main three Cryptocurrencies that we will be examining in our research are Bitcoin, Ethereum, and Shiba Inu. Ethereum and Bitcoin are the top two cryptocurrencies (in terms of market capitalization) at this time. Shiba Inu is a relatively new cryptocurrency released in 2018, but has gained significant market traction since its inception. We will check if there is a positive, negative, or no correlation at all between the price, sentiment analysis and tweet count.

Twitter is considered as one of the largest social media platforms with: 1.3 billion accounts, 83% of the world's leaders having a Twitter account, and 500 millions tweets being posted every day. This makes Twitter a powerful factor for changes in many trends.
Moreover, Bitcoin and Ethereum have a combined value of 160.9 billion dollars, making them the two largest Cryptocurrencies to exist, with Shiva Inu being one of the highest trending Cryptocurrencies today.

Examining what effect a powerful platform like Twitter has on these three currencies is crucial for investors to understand and predict trends of growth or decline in the values of those currencies. Since Cryptocurrencies tend to experience significant swings both in their daily and their long term values, being able to predict their value's trends can provide a huge advantage to their users.
In addition, Twitter has been increasingly used as a source of increasing currency purchases by informing users of the perceived market conditions.

Literature review

A study like ours has already been done in the past by a group of Master's students at Southern Methodist University in 2018. In their research, called "Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis", the group presented a method for predicting changes in Cryptocurrencies prices utilizing Twitter and google trends data. However, they only studied Bitcoin and Ethereum, without Shiba Inu.

In their research, Abraham, Hidgon, Nelson, and Ibarra collected data from Twitter's API, as well as collecting data from Google Trends, and divided the data between neutral and objective.
They found that sentiment of Tweets tend to stay positive regardless of price changes, which indicates that Tweets volume may be a better indication than sentiment. They found that there is a strong positive correlation, utilizing PearsonR, between Tweets volumes and Bitcoin prices, with an average correlation of 0.841 and P-value of 0.000.

Data Collection

Tweet Collection

In order to collect sentiment data on our different crypto currencies we utilized the Twitter API. Twitter allows access to their API for limited pulls for different purposes. In order to be able to collect all of the data that we needed we had to make several requests to the API and run a web scraper for a couple of days. The program worked by making a request around every fifteen minutes for 48 hours. Our only reason for not collecting more data was due to the fact that we were suspended from using API due to us maxing out the number of calls that a user can make in a month. Everytime we made a call we took in the json object and appended it to the CSV file that we were using for data storage. Over our data collection period we were able to collect 154,702 unique tweets. All of these tweets contained a hashtag that related to one of the three crypto currencies we were interested in or a combination of them.

```
language = 'en'
querry = '#BTC OR #btc OR #Bitcoin OR #bitcoin OR #ETH OR #eth OR
#Ethereum OR #ethereum OR #shib OR #SHIB OR #ShibainuCoin OR
#shibainu -filter:retweets'

tweets_df = pd.DataFrame(columns=['id', 'Text', 'Date', 'User',
'Hashtags', 'RTCount', 'FavoriteCount', 'Lang'])
#tweets_df = tweets_df.set_index('id')
#Enter a for loop that loops through all of the items in
while True:
   try:
           print('Starting the while loop')
           tweets = tweepy.Cursor(api.search_tweets, q = querry,
lang=language).items()
           print('Data pulled at ' + time.asctime(time.localtime()))
```

```
            #Get a list of all the items we are interested in
            tweets_list = [[tweet.id_str, tweet.text,
tweet.created_at, tweet.user.screen_name, tweet.entities['hashtags'],
tweet.retweet_count, tweet.favorite_count, tweet.lang] for tweet in
tweets]
            #Add them to the data frame by converting and appending
            #print(tweets_list)
            print('Appending to dataframe')
            temp_df = pd.DataFrame(tweets_list,
columns=tweets_df.columns)
            #temp_df = temp_df.set_index('id')
            print(temp_df)
            tweets_df = tweets_df.append(temp_df, ignore_index=True)
            print(tweets_df)
            print('The current date-time for the last pulled tweet is
' + str(tweets_list[-1][2]))
            tweets_df.to_csv(path_or_buf='tweetdb.csv', sep=',',
mode='w+')
            print('Saved to File')
            #wait for 15 minutes

    except KeyboardInterrupt:
        print("Press Ctrl-C to terminate while statement")
        tweets_df.to_csv(path_or_buf='tweetdb.csv', sep=',',
mode='w+')
```

Sentiment Analysis

We used TextBlob to run sentiment analysis on our collected tweets. The body of each tweet was cleaned using natural language processing and passed through TextBlob. From here each tweet received a polarity score ranging from -1 to 1 using the sentiment analysis functionality of TextBlob.

```
def get_polarity(comment):
     comment = re.sub('[^a-zA-Z\']', " ", comment)
     comment = re.sub("\s+", " ", comment)
     comment = comment.strip()
     comment = comment.lower()
     return TextBlob(commnet).sentiment().polarity
```

Currency Pricing

For this research the yfinance (Yahoo Finance) python package was used. Provided below is a more basic version of the code used to retrieve and plot the data:

```
import finance as yf
import pandas as pd
import matplotlib.pyplot as plt

def main():
    btc = yf.Ticker("BTC-USD")
    eth = yf.Ticker("ETH-USD")
    shib = yf.Ticker("SHIB-USD")
    pandas_options()

# get stock info

# get historical market data
    btc_df = btc.history(period="7d", interval = "15m")
    eth_df = eth.history(period="7d", interval = "15m")
    shib_df = shib.history(period="7d", interval = "15m")
    print(btc_df)
    btc_df.to_csv("output.txt", sep='\t')
    createChart(btc_df, 'Bitcoin')
    createChart(eth_df, 'Ethereum')
    createChart(shib_df, 'Shiba Inu')

def pandas_options():
    pd.set_option('display.max_rows', None)
    pd.set_option('display.max_columns', None)
    pd.set_option('expand_frame_repr', False)

def createChart(df, coinName):
    plt.figure()
    plt.plot(df.index, df['Open'])

#plt.scatter(df['sentiment_score'],df['overall_pred'],color='red',
s=1)
    plt.title(f'{coinName} price to time')
    plt.xlabel('Date')
    plt.xticks(rotation = 45)
    plt.ylabel('Price')
    plt.show()

main()
```

Analysis of Bitcoin, Ethereum, and Shiba Inu : 11/11/2021-11/13/2021

(Unfortunately the Twitter API locked our research account out due to the number of tweets we were pulling. We were only able to obtain 2-3 days of twitter data, limiting our research.)

Average sentiment v. Price

	Bitcoin	Ethereum	Shiba Inu
Correlation	0.2687939812238277	0.052651687017889845	-0.23714730984029858
P-value	0.000521632048615217	0.5044517442853009	0.0023037402813186826

Tweet Count v. Price

	Bitcoin	Ethereum	Shiba Inu
Correlation	-0.5857279152117493	-0.4576728271842653	0.3538823782529401
P-value	0.0000000000000000021632	0.0000000008136449	0.0000035840917569

A note on the ols and R-values

For our project we did compute an ols model. However, because we are dealing with market data we had a very low R value. This is due to the fact that there are many different factors at play when it comes to market data. This makes it incredibly difficult to get a high R-value which accounts for most of the data.

For our sentiment analysis and tweet count on the twitter side we are only analysing tweets in English. The sentiment and tweet count analysis of tweets in another language could highly vary from our data sample. This is one of the many factors that would need to be considered when attempting to develop a full analysis of cryptocurrency sentiment analysis.

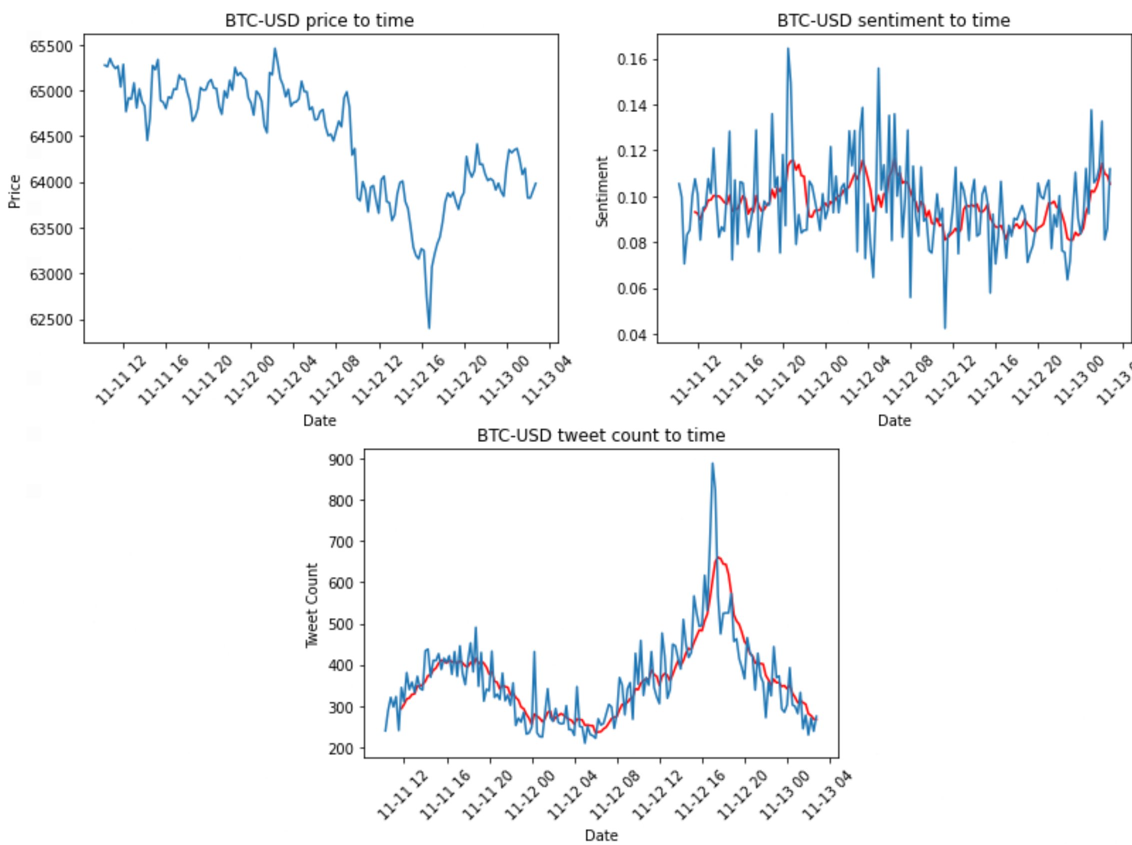
(Bitcoin v. Tweet Count - OLS regression model)

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            tweet_count   R-squared:                       0.343
Model:                            OLS   Adj. R-squared:                  0.339
Method:                 Least Squares   F-statistic:                     84.08
Date:                Thu, 02 Dec 2021   Prob (F-statistic):           2.16e-16
Time:                        23:30:30   Log-Likelihood:                -957.37
No. Observations:                 163   AIC:                             1919.
Df Residuals:                     161   BIC:                             1925.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept   6712.1271    692.649      9.691      0.000    5344.278    8079.976
Open          -0.0985      0.011     -9.170      0.000      -0.120      -0.077
==============================================================================
Omnibus:                       30.928   Durbin-Watson:                   0.733
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               64.263
Skew:                           0.849   Prob(JB):                     1.11e-14
Kurtosis:                       5.565   Cond. No.                     6.59e+06
==============================================================================
```

Bitcoin Analysis

In our analysis of Bitcoin we found there to be a negative correlation between the number of tweets about bitcoin and the price of Bitcoin. In the graph below when Bitcoin hits a peak loss around \$62,500 the number of tweets peeks at around 900. The P-value for this data was 0.0000 showing a high statistical accuracy for our model. The P-value matches the findings of Abraham, Hidgon, Nelson, and Ibarra in their paper, "*Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis*". However, we found a strong average negative correlation between the two at -0.586 whereas they found a strong positive correlation at 0.841. Our findings suggest, in addition to the findings by Abraham et al., there is a strong correlation in the number of tweets with any significant movement in the price of Bitcoin. The direct relationship between price and number of tweets is not linear.

In terms of sentiment analysis we found a slightly positive average correlation of 0.269 and a P-value of .0005 showing relatively high statistical significance.

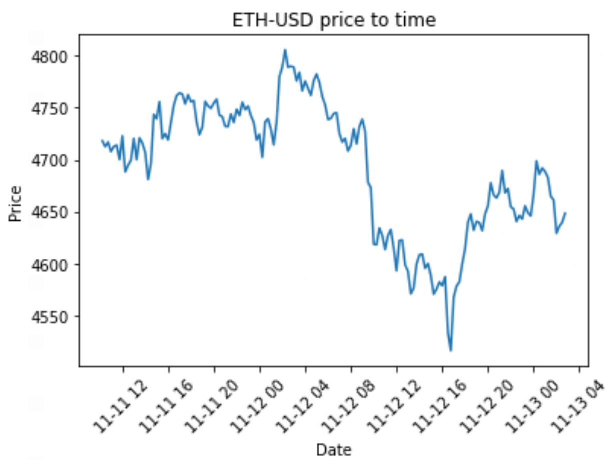


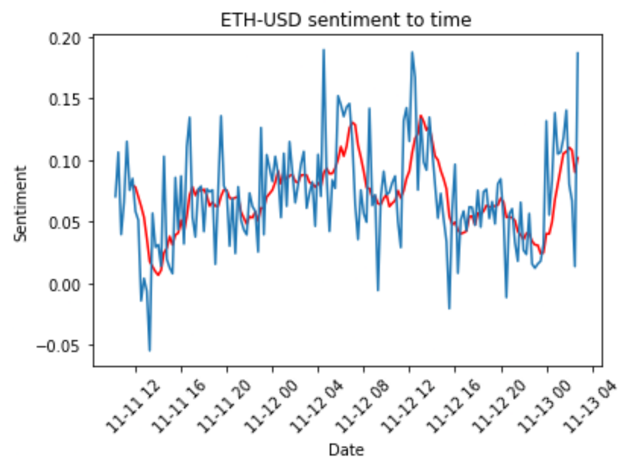
Ethereum Analysis

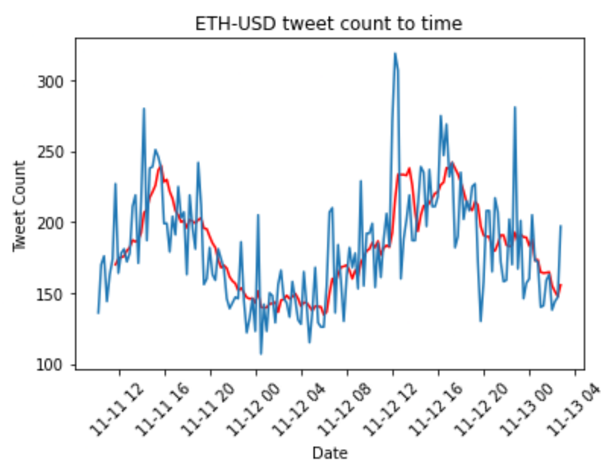
In our analysis of Ethereum, the results were similar to those of Bitcoin. Our graph shows a negative correlation between the number of tweets related to Ethereum and the price of it. In the graph below when Ethereum hits a peak loss of approximately $4500, the number of tweets peaks at over 300 Tweets. Looking at the accuracy of the data, a P-value of 0.0000 shows a high statistical accuracy for our model. The direct relationship between price and number of tweets is not linear.

In terms of sentiment analysis, we found a slightly positive average correlation of 0.0526 and a P-value of 0.50445 showing low statistical significance.

The paper written by Abraham, Hidgon, Nelson, and Ibarra, "*Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis*" does not talk about their findings regarding Ethereum, and therefore we were not able to compare our results to theirs.

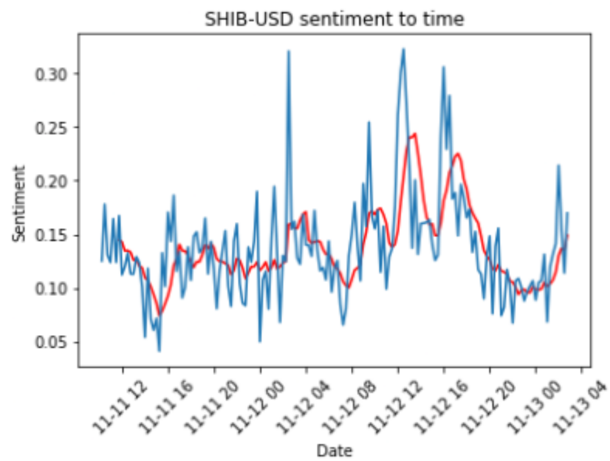


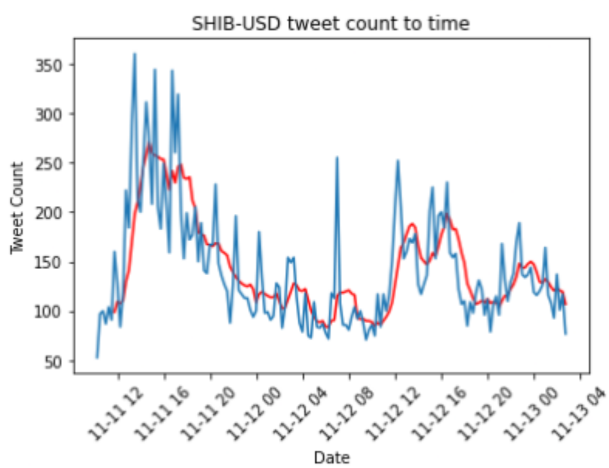




Shiba Inu Analysis

Shiba Inu was the third crypto currency that we analyzed for this project. With Shiba we saw another correlation between the number of tweets in the fifteen minute window and the price of Shiba. The correlation was 0.35388 which is positive but not as strong as the bitcoin correlation. The p-value is also well below 0.05 however it is higher than the ethereum and the bitcoin p-values. Shiba is a newer crypto currency and widely considered to be trendy and upstable. It is very likely that Shiba is not here to stay. We believed that because of the popularity of the coin on social media it would be a good subject for use to analyze. Unfortunately the opposite seems to be true. It seems that the peoplairty of the coin created a lot of noise in the data that made it more difficult to to check the correlation between the tweet volume, and the price. This is also clear in the average sentiment analysis of tweets on the Shiba coins. The relationship was actually negative meaning that as the positive sentiment increased the price of the coin went down. This was not predicted by us but it could fall into the old trading strategy of buy the rumor and selling the news. Overall we were able to find that there was a correlation between tweets and the price and also that all of the correlations were different for each of the coins.





Sources

Abraham, J., Higdon, D., Nelson, J., & Ibarra, J. (n.d.). *Cryptocurrency price prediction using tweet volumes and sentiment analysis*. SMU Scholar. Retrieved December 1, 2021, from https://scholar.smu.edu/datasciencereview/vol1/iss3/1/.

Lhessani, S. (2021, November 11). Python: How to get bitcoin data in real-time! (less than 1-second lag). Medium. Retrieved December 1, 2021, from https://medium.com/analytics-vidhya/python-how-to-get-bitcoin-data-in-real-time-less-than-1-second-lag-38772da43740.

Farell, R. (n.d.). *An analysis of the cryptocurrency industry*. ScholarlyCommons. Retrieved December 2, 2021, from https://repository.upenn.edu/wharton_research_scholars/130/.

Inamdar, A., Bhagtani, A., Bhatt, S., & Shetty, P. M. (2019). Predicting Cryptocurrency Value using Sentiment Analysis. *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, 932–934. https://doi.org/10.1109/ICCS45141.2019.9065838

Jain, A., Tripathi, S., Dwivedi, H. D., & Saxena, P. (2018). Forecasting Price of Cryptocurrencies Using Tweets Sentiment Analysis. *2018 Eleventh International Conference on Contemporary Computing (IC3)*, 1–7. https://doi.org/10.1109/IC3.2018.8530659

Karalevicius, V., Degrande, N., & De, W. J. (2018). Using sentiment analysis to predict interday Bitcoin price movements. *The Journal of Risk Finance*, *19*(1), 56–75. https://doi.org/10.1108/JRF-06-2017-0092

Kim, K., Lee, S.-Y. T., & Assar, S. (2021). The dynamics of cryptocurrency market behavior: Sentiment analysis using Markov chains. *Industrial Management & Data Systems*, *ahead-of-print*(ahead-of-print). https://doi.org/10.1108/IMDS-04-2021-0232

Kraaijeveld, O., & De Smedt, J. (2020). The predictive power of public Twitter sentiment for forecasting cryptocurrency prices. *Journal of International Financial Markets, Institutions and Money*, *65*, 101188. https://doi.org/10.1016/j.intfin.2020.101188

McAteer, C. (n.d.). *Twitter Sentiment Analysis to Predict Bitcoin Exchange Rate*. 86.

Mittal, A., & Goel, A. (n.d.). *Stock Prediction Using Twitter Sentiment Analysis*. 5.

Naeem, M. A., Mbarki, I., Suleman, M. T., Vo, X. V., & Shahzad, S. J. H. (2021). Does Twitter Happiness Sentiment predict cryptocurrency? *International Review of Finance*, *21*(4), 1529–1538. https://doi.org/10.1111/irfi.12339

Porshnev, A., Redkin, I., & Shevchenko, A. (2013). Machine Learning in Prediction of Stock Market Indicators Based on Historical Data and Data from Twitter Sentiment Analysis. *2013 IEEE 13th International Conference on Data Mining Workshops*, 440–444. https://doi.org/10.1109/ICDMW.2013.111

Rao, T., & Srivastava, S. (2012). *Analyzing stock market movements using Twitter sentiment analysis*. 119–123. https://doi.org/10.1109/ASONAM.2012.30

Valencia, F., Gómez-Espinosa, A., & Valdés-Aguirre, B. (2019). Price Movement Prediction of Cryptocurrencies Using Sentiment Analysis and Machine Learning. *Entropy*, *21*(6), 589. https://doi.org/10.3390/e21060589

Wołk, K. (2020). Advanced social media sentiment analysis for short-term cryptocurrency price prediction. *Expert Systems*, *37*(2), e12493. https://doi.org/10.1111/exsy.12493

Bing, L., Chan, K. C. C., & Ou, C. (2014). Public Sentiment Analysis in Twitter Data for Prediction of a Company's Stock Price Movements. *2014 IEEE 11th International Conference on E-Business Engineering*, 232–239. https://doi.org/10.1109/ICEBE.2014.47

Dickinson, B., & Hu, W. (2015). Sentiment Analysis of Investor Opinions on Twitter. *Social Networking*, *04*(03), 62. https://doi.org/10.4236/sn.2015.43008

Guo, X., & Li, J. (2019). A Novel Twitter Sentiment Analysis Model with Baseline Correlation for Financial Market Prediction with Improved Efficiency. *2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, 472–477. https://doi.org/10.1109/SNAMS.2019.8931720

Liew, J., Li, R., Budavári, T., & Sharma, A. (2019). Cryptocurrency Investing Examined. *The Journal of the British Blockchain Association*, *2*(2), 1–12. https://doi.org/10.31585/jbba-2-2-(2)2019

Loginova, E., Tsang, W. K., van Heijningen, G., Kerkhove, L.-P., & Benoit, D. F. (2021). Forecasting directional bitcoin price returns using aspect-based sentiment analysis on online text data. *Machine Learning*. https://doi.org/10.1007/s10994-021-06095-3