Sentiment-Based Prediction of crypto currency

Alay Amitkumar Patel

Data analytics for business

Email-AP185@myscc.ca   
St. Clair Collage Windsor, Canada  
Priyank Darji

Data analytics for business

Email-PD35@myscc.ca  
St. Clair CollageWindsor, Canada

*Abstract*—Recent work on crypto price predications shows that twitter and other social media somehow related to its price as the cryptocurrency is highly volatility of it. It’s very hard to predict the price of bitcoin, Ethereum, etc. which is highly correlated with the fluctuation of cryptocurrency price. As we did sentimental analysis on the twitter tweets regarding to the cryptocurrency, we found that the sentiment of tweets somehow effects the stock price of some cryptocurrency. Also, the related works shows that the sentiment of social media which affect the cryptocurrency price. Firstly we try to scrapy the data using twint, twitter api which us gave use data of week as per the limit of the data scraping through the api and twint library we cannot able to gather the data more than the a week so we used the public dataset which was available on the Kaggle and we also use the yfinance library to get the data for the price of the cryptocurrency to check the correlation of the price with the sentiment as we found that the price of bitcoin is highly correlated with the sentiment of the tweets. After that we applied some neural network models such as CNN, LSTM and Bi-Directional LSTM. The proposed experiment outperforms with the result 87% accuracy for Bi-directional LSTM.

Keywords—Twitter, crypto, bitcoin, sentimental analysis, cnn, lstm, bi-directional lstm.

# Introduction

Cryptocurrency is decentralized electronic system which change the financial system since the bitcoin was introduced in 2009 by Satoshi Nakamoto and created in 2008 [6] as it was bitcoin system which works on peer-to-peer network that makes it highly volatile and there is no physical existence of bitcoin or other cryptocurrency and there is no bank who is considered as central authority to this system the its basically works on the encryption the method for the cryptocurrency wallet which consist the set of encrypted keys like public address and private keys. Even the digital currency draws the attention of as it is highly volatility which also gives high returns. Market capitalization of crypto currency increased from approx. 1 billion U.S. dollar in 2013 which reached to 1000 billion U.S. dollar in 2021[8]. Public address is used to deposit the money to the wallet as the fund cannot be removed from the wallet. Private keys create on the seed phares which can be any kind of the text and then they apply the Elliptic Curve Digital Signature Algorithm (ECDSA) to create a new set of private key and corresponding public key[7].

Nowadays social media such as Twitter [9], reddit [10], and Facebook [11] on such social media platform we can able to see the latest news and social media post regarding to about easily accessible financial markets As a result, investors have been using a wide range of digital information to make trading judgments. Previous research has found evidence of a link between price movement and social media usage. Sentiment on currency social media posts with negative emotions, such as fear and grief, neutral emotions, such as calm and unsure, or good emotions, such as trust and enthusiasm, can be used to forecast cryptocurrency price swings and help with investment decisions. To apply sentimental analysis firstly we tried to gather the data using twitter API. We found that the social media somehow correlated to the price fluctuation of bitcoin.

Mostly Cryptocurrency price fluctuation gives the high returns as a result so many people want to invest in the cryptocurrency but as there is no method to forecast the price of the cryptocurrency and also the technical analysis works sometimes not mostly so there is a need to create a system which can help traders to trade there funds in crypto market as nowadays people aware about bitcoin and other cryptocurrency they interested in it more but due to less information is there they have fear of massive loss as in the past so many loss there funds due to this cryptocurrency in scams and in trading so due to this we want a way to predict the ups and downs of cryptocurrency. Related works gives promise that cross-correlation of the cryptocurrency with social media and google trends. The twitter affect the crypto market even tweets from elon musk recent raise the price of bitcoin that was impacted numbers of traders and which is quite unexpected as the cryptocurrency is decentralized that means there is no central authority and no physical existence of it. So price fluctuation is works differently as mining bitcoin is also exist which means that people provide there machines to increasing the hashing rate makes the bitcoin or other cryptocurrency hashing more powerful and also increase the security over each of the transactions. Social media posts regarding financial used by traders to put there money in cryptocurrency investment and due to so much of post, tweets its quite cross-correlated to the price.

# Related work

## Bitcoin Spread Prediction Using Social And Web Search Media [1]

**Authors:** Martina Matta, Ilaria Lunesu, Michele Marches

* **Problem Statement:** discover if the chatter of the community can be used to make qualitative predictions about Bitcoin market, attempting to establish whether there is any correlation between tweet’s sentiment and the Bitcoin’s price.

**Dataset:** They use twitter API and extracting the data using the hashtags. Twitter Streaming API that give update about the real time data.

**Approach:** Compare bitcoin price variation with tweets and google trends result using cross-correlation value and resulting relationships using vector autoregressive and vector error correction models. Because users' perspectives on many topics are occasionally expressed in tweets, we decided to assess user opinions on Bitcoin. We also looked into its ability to forecast real-world outcomes. We tried to forecast attitudes based on a set of tweets in order to see if a user truly appreciated the Bitcoin spread. In recent years, a large body of research has centered on machine learning algorithms for extracting and identifying subjective information in texts. Sentiment analysis or opinion mining is the term for this type of research. Sentiment analysis can derive signs of public mood from social media content directly.

**Results:** Report analysis about the relationship of twitter tweets and the cryptocurrency market. Bollen et al. found that tweets can accurately anticipate market trends 3-4 days ahead of time, with a high probability of success. We looked at the behavior of the Bitcoin price by comparing it to the number of tweets, the number of tweets with a good attitude, and Google Trends data. Cross-correlation calculation gave intriguing results. Our findings appear to support the theory that the number of tweets exchanged can anticipate Bitcoin price variations[12]. Furthermore, a comparison of tweets in a happy attitude and the price trajectory of Bitcoin appears to support this habit. The current data of 60 days looks promising if the analysis of more than 6 months might provide better quality result. The cross-correlation results of google trends and the bitcoin price that looks quite signification.

**Pros:** It works will with 60 days data and high cross correlation value with a zero lag.

**Cons:** They don’t able to gather more data but they want to do analysis on 6 months data.

**Conclusions:** Between January and March 2015, the corpus covers a 60-day period. We utilised automated Sentiment Analysis on these tweets to see if public sentiment could be used to predict the price of Bitcoin. We also looked at Bitcoin's popularity via the lens of Web search using Google Trends media. We compared the Bitcoin price's fluctuations to those of tweet volume, tweets with a positive mood volume, and Google Trends data in our preliminary study. We can confirm that favourable tweets may help anticipate the movement of Bitcoin's price in a few days based on the findings of a cross pearson correlation between both time series. By its high cross correlation, Google Trends might be considered a predictor.

## Advanced social media sentiment analysis for short‐term cryptocurrency price prediction [2]

**Authors:** [Krzysztof Wołk](https://www.researchgate.net/profile/Krzysztof-Wolk)

**Problem Statement:** Short-term cryptocurrency predication using the sentimental analysis.

**Dataset:** This research paper used the tweets from twitter as well as they us google trends of web search on price of bitcoin.

**Approach:** Ensemble method and linear regression model and the compare it. With the support of Twitter sentiments and Google Trends, the work was started utilising two projected models important for estimating the price of the coin. Least square regression analysis and Bayesian Ridge Regression Model are two of these models. These models are part of the SKLEARN Python language package. Bayesian Ridge Regression, as described by Nie and Ji (2014), is another essential technique used in analysis. Future learning, according to Nie and Ji (2014), refers to learning the transformation of raw data into useful and analytical data and other goals. Auto-encoders, dictionary learning, constrained Boltzmann machine, k-means clustering, and many other ways are examples of supervised and unsupervised feature learning techniques.

**Result:** Predicted price using hybrid model predicts better as empirical experiment on investing 100$ After a month the account balance showed 114.82$, what conforms method is profitable, especially crypto market is on its down currently. The hybrid models that were used, as well as their fit measures, are presented below. When predicting current price on a brand-new data point, each model was tested on testing data to obtain the RSI and ME values, where ME stands for mean error, R2 stands for regression analysis, and +-T stands for real error (the final interval). At first, sampling was done in 10- and 60-minute shifts. Overall, the five - minute shifts result in reduced inaccuracy in the hybrid model in this empirical investigation, hence it was decided to be employed in the studies.

**Pros:** we can conclude that cryptocurrency fluctuations depend heavily on social media sentiments and web data bases such as Google Trends.

**Cons:** problems associated with prediction of crypto is high level of flexibility of the currency due to volatility nature of cryptocurrency in the current market.

## LSTM Based Sentiment Analysis for Cryptocurrency Prediction [3]

**Author:** [Xin Huang](https://arxiv.org/search/cs?searchtype=author&query=Huang%2C+X), [Wenbin Zhang](https://arxiv.org/search/cs?searchtype=author&query=Zhang%2C+W), [Xuejiao Tang](https://arxiv.org/search/cs?searchtype=author&query=Tang%2C+X), [Mingli Zhang](https://arxiv.org/search/cs?searchtype=author&query=Zhang%2C+M), [Jayachander Surbiryala](https://arxiv.org/search/cs?searchtype=author&query=Surbiryala%2C+J), [Vasileios Iosifidis](https://arxiv.org/search/cs?searchtype=author&query=Iosifidis%2C+V), [Zhen Liu](https://arxiv.org/search/cs?searchtype=author&query=Liu%2C+Z), [Ji Zhang](https://arxiv.org/search/cs?searchtype=author&query=Zhang%2C+J)

**Problem Statement:** Bitcoin price volatility problem

**Dataset:** Tweets and the articles that talks about the cryptocurrency

**Approach:** LSTM model and sentimental Analysis is used to predict the price of bitcoin. They employed domain-expert knowledge to trawl user posts from China's most popular social media platforms, Sina-Weibo, and develop a crypto-specific sentiment lexicon. The LSTM back - propagation neural network was then used to model the sentiment polarity and provide real-time price trend predictions. The embedding layer converts this same word token into the crypto deep learning after tokenizing the social media post using crypto word vocabulary. The embedded feature vector sequence is used to train the LSTM based persistent network. The output of the LSTM is transformed subsequently using a fully connected layer, which is then activated with sigmoid to produce the prediction. The labels for the training posts were manually labelled and encoded with positive (1), neutral (0) and negative (-1).

**Result:** The Result shows that activity of influential personalities in the digital world does affect the price of this cryptocurrency. They used the previous seven days of Sina-Weibo posts from the top 100 crypto investor accounts as training data, and the next one day's postings as testing data. Our LSTM sentiment predictor's performance is measured using Precision and Recall. The model's precision evaluates its ability to return just relevant instances, while recall measures its ability to recognize all relevant instances.

**Pros:** 75% success rate

**Cons:** no public datasets available for this task

## Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis[4]

**Author:** Jethin Abraham(Southern Methodist University, jethina@smu.edu), Daniel Higdon (Southern Methodist University, [dhigdon@smu.edu](mailto:dhigdon@smu.edu)) John Nelson (Southern Methodist University, [nelsonjohn@smu.edu](mailto:nelsonjohn@smu.edu)) and Juan Ibarra (Southern Methodist University, jibarralopez@smu.edu)

**Problem Statement:** Predicting cryptocurrency price changes. To accomplish this, methods utilizing sentiment analysis of tweets are reviewed. This is involved utilizing Twitter’s API and a Python library called” Tweepy”2.

**Dataset**: tweet dataset was made up of 30,420,063 tweets.

**Approach:** Data collected, cleaned, and adjusted where needed the data was analyzed to use input for the final model and inputs into a linear model and summarize the results. Pearson R is indeed a measure of the correlation's strength. Its value is between -1 and 1. A positive number indicates that the individual variables are positively linked, or that an increase in the money supply produces an increase in the other (this is linkage, not connection, so we can't say that one variable causes the other to change, only that there is a correlation). A negative value, on the other hand, indicates that the two characteristics are negatively connected, or that a raise in one variable's value is associated with a reduction in the other. The p-value indicates how likely these relationship measures would also have been discovered by chance.

**Result:** Result of Statistics is that the twitter can be very rich source of data on how people feel about nearly topic. The model was created with LSTM and has a higher recall and precision than a standard auto regressive method. The current sentiment meter can be used to forecast bitcoin price fluctuations and linked into an automated trading system to aid in the purchase or sale of digital assets.

Pros/Cons: Pros: volume index is highly correlated with cryptocurrency prices both when prices rise and when they fall, as are tweet volumes.

**Cons:** inputs a multiple linear regression model, with the addition of lagged variables, accurately reflected future price changes.

## Sentiment-Based Prediction of Alternative Cryptocurrency Price Fluctuations Using Gradient Boosting Tree Model[5]

**Author:** Tianyu Ray Li , Anup S. Chamrajnagar , Xander R. Fong, Nicholas R. Rizik and Feng Fu

**Problem Statement:** Sentimental of social media affect the fluctuation of the bitcoin price

**Dataset:** In this they used the open-sourced rtweet package, which accesses Twitter’s REST and stream APIs.

**Approach:** Greedy algorithm for split finding used in research paper as their approach for price prediction model and also XGBoost method is used. They used 10-fold cross validation on 589 data points to determine between linear regression, logistic regression, polynomial regression, exponential regression, tree model, and support vector machine regression as the best model framework. The Xgboost Regression (commonly known as XGBoost) was chosen to train the model on our data because it had the smallest loss, or inaccuracy. For the following reasons, the XGBoost model, as well as many other tree-based models, is especially suited for use with their data: Tree topologies are unaffected by the information and feature arithmetic range. As a result, we don't need to normalize the data, perhaps avoiding data loss as a result of normalizing. Due to their building methods, tree models were the most extensible machine learning model—by merely adding more offspring nodes toward the pre-existing tree nodes, we can update the tree and continue to properly forecast price as our gathering of pricing and tweet data grows in the future. It also allows the model to be adjusted to accommodate currencies with higher daily tweet volumes. At a high level, the tree paradigm is a guideline learning method that, in contrast to standard regression learning methods, has a greater potential for revealing interesting correlations between information.

XGBoost is indeed a tree ensembles model that produces a weight value of multivariate regression tree forecasts by prioritising mislabeled samples.

**Result:** algorithm guessing the sentiment correctly is over 50%, 80% success rate in successfully classifying positive tweets, and correctly characterized 0% of positive tweets as negative in this sample.

**Pros/Cons:** Pros: Extreme Gradient Boosting Regression Tree Model serves as a viable means of predicting price fluctuations within the ZClassic cryptocurrency market.

**Cons:** The complete lack of research within this academic sphere, our model serves as a proof of concept that social media platforms such as twitter can be used to capture investor sentiment.

# Problem statement and Objectives

First of all, crypto market is not like stock market, so tracking the market becomes much more difficult. People generally have less knowledge about the crypto as there is no physical company that person can research on. People generally attracted to quick money-making stories and invest in crypto then mostly lose their money. People do not have any knowledge about how volatile the market is going to be. Not being able to comprehend the calculations and how quickly they are liquidated People generally fall into traps of market and end up loosing money because they have nothing to check. They do not have proper prediction strategy to invest in crypto market.

We use Twint[13] library and twitter api to scrapy the tweets from the twitter which allow us to gather almost 5 days of data due to the limit of scraping we cannot able to collect more data so we are used the dataset that was available on Kaggle public dataset of tweets[14] and we used the yfinance[15] to get the price of cryptocurrency over time period of at least 2 years as we have the tweets of 2 years.

Our objective is to help the traders which want to invest their funding in the cryptocurrency. There are two steps in making use of theoretical approach:

we have to construct a social network with all the participants that are directly or indirectly correlated with the cryptocurrency market. The aim is to identify specific nodes for sentiment analysis, which means that we can extract the sentiment from nodes through text mining. In our case, Bitcoin and other cryptocurrencies are the nodes, and then we search for sentiment-related texts from other relevant media and users' comments. Apply a machine learning model to analyse those texts in order to find out users' sentiments regarding a certain topic or cryptocurrency. Cryptocurrencies are notoriously difficult for sentiment analysis. The general sentiment dictionary created by natural language processing (NLP) is not applicable in the crypto domain. We introduce a novel way to build a crypto specific sentiment dictionary that can capture the unique characteristics of the crypto social communities.

We tried to check the correlation of the cryptocurrency price with the sentiment of the tweets which shows that the ups and downs in crypto stock price is quite predictable from the sentimental analysis of the tweets or other social media source can be use for traders even as technical analysis not works properly on crypto forecasting and other methods are not that much promising so this way its possible to get the basic information such as price will go up or down based on the sentiment the forecasting system just on price predication sometimes wrong about the price raise and fall but this twitter tweets method even include some of news as well which can also helps to understand ups and downs more clearly compare to the forecasting of price using few parameters or technical analysis of it.

# Proposed Solution/Method

Method, we used is to get the sentiment of tweets and compare it with the price of crypto currency. We used Twint[13] to scrapy the data and twitter api but due to limit of scraping the data we got is less which is not quite efficient to compare the it with stock price so then after we use the publicly available dataset from Kaggle bitcoin tweets[14] which consist of 13 features and the tweets it contains is from 6 may 2021 to 22 jan 2022 and the size of dataset is 1.27 gb it contains 2259788 rows and the yfinance data consist of 2691 rows and 7 columns as the tweets is for the whole day and the price is contains the price of Ethereum, bitcoin, Cardano, Litecoin, Polkadot, Ripple, Solana, USDtether, Dogecoin and Bitcoin cash in U.S. Dollar. So we did the cleaning the data from the tweets and that’s how and remove symbols and links from the tweets and apply regular expression rule to remove the links and useless text from the tweets text and then after the generated Polarity scores using NLTK[16] library and we check the correlation of the price and the sentiment which makes sense that the price is increased after the sentimental is positive which shows that crypto market somehow effected by social media posts.

Steps we did on the dataset for cleaning giving below:

* Step 1: Firstly, we sorted the data on the base of date time stamp. And remove unnecessary columns from the dataset.
* Step 2: We did cleaning using regular expression and it takes lot of time as we used for loop to apply multiple steps of cleaning together in all rows and in that we remove the hashtags, links, and usernames which was in the tweets as reply or referred accounts.
* Step 3: In this step we make column for polarity of it tweet text and even for this we used for loop to apply on all the rows.

As the data contains so many rows we use TQDM [17] library for this task as TQDM helps use to see the time of the process it takes with using progress bar.

We create the score using the polarity in that the equation is as follows:

(Polarity \* User follower \* User favourites) / (User follower \* is\_retweet)

Note: is\_retweet stands for did that user retweeted on the or not.

We grouped the data hourly so we can connect it with the price of the cryptocurrency after that we created some graphs to check the correlation of score and price raise. As graph was created for the top 10 cryptocurrency on the bases of dataset which includes BCH, DOGE, USDT, Solana, DOT, Ripple, Litecoin, Cardano, Bitcoin and Ethereum. In Bitcoin it works perfectly and also for Ethereum. Even in Solana, polkadot, Litecoin sentiment is little bit correlated.

Then we make prepare the data to create the model and that was did using the NLTK[16] library as they have so many functions and that helps to prepare the data for training the model. Functions like stop-words, tokenize and lemmatize (Lemmatization is the process of combining a word's several inflected forms into a single item that can be studied).

As we make dataset ready to feed in deep learning models which before that we also added two columns for Subjectivity and polarity of the tweets. This task was did using the help of TextBlob[18] library. TextBlob[18] is simplified text processing library which can be to process the textual data, polarity and subjectivity is use to sentiment is positive, negative or neutral. While split the dataset and target variable we just converted the target variable into the numeric using pandas[19] and also use tokenizer which convert streams of text into tokens.

Final step before creating a model was pad sequences ensures that all sequences in a list are of the same length. By default, this is accomplished by padding 0 at the start of each sequence until it equals the length of the longest sequence. This task did using keras and tensorflow library[20]. We created three deep learning model CNN(convolutional neural network), LSTM(Long Short-Term Memory) and Bi-directional LSTM(Long Short-Term Memory). CNN is a type of deep neural network that is frequently used to analyze visual imagery [21]. But it can also used for 1d array as our data is 1dimenisonal, so we applied this below is the image which shows the details of layer that we used to create the model. Table

Description automatically generated

fig 1.0 Cnn Layers

As the CNN model gives use the accuracy of 94% but there is no correlation between Training and validation data. We create model using LSTM which mostly used for text processing which gives the accuracy of 89% as other metrics such as confusion matrix shows there is so many of false positive value. So then after we tried the Bi-directional LSTM which gives accuracy of 92% on training data below images shows the layers of Bi-directional LSTM.

The target variable contains:

* 1 if the comment is positive
* -1 if the comment is negative
* 0 if the comment is neutral

Table

Description automatically generated

fig 1.1 Bi-directional LSTM Layers

As from the model summary in fig 1.0 and 1.1 we can see embedding layer which is used to give input to model for tokenized data.

# Results

We create plots which shows the correlation between the sentiment of tweets and the bitcoin, Ethereum price is shown below as green line shows the sentiment and blue line shows the price of the Bitcoin and Ethereum(eth).

Chart, histogram

Description automatically generated

Fig 2.1 Above graph is show the relation between the sentiment and the price of bitcoin

Chart, histogram

Description automatically generated

Fig 2.2 Above graph is show the relation between the sentiment and the price of Ethereum

So there is cross-correlation between sentiments and raise of price in Bitcoin(BTC) and Ethereum(ETH).

The relationship shows that tweets somehow affect the price raise and fall so then we created the models and to get sentiment of the given text as this method works well as compare to other financial techniques even as you saw in the related work so then after we create two deep learning model on base of sentiment.

CNN and Bi-directional LSTM results are given below:

1. Table Type Styles

| Models | Metrics | | |
| --- | --- | --- | --- |
| Accuracy | Recall | Precision |
| CNN | Train:98.03%  Test:94.30% | 90.28% | 92.61% |
| Bi-directional LSTM | Train:92.71%  Test:87.20% | 88..54% | 88.54% |

1. As the Recall and precision metrics was created on test data so there is only on test data.

Below is the graph which shows the accuracy on the particular interval of epochs

A picture containing graphical user interface

Description automatically generated

Fig 3.1 Above plot shows model accuracy and model loss of CNN

As we can see in above graph the correlation of the validation accuracy and train accuracy is not worth noting as they are very far from each other so here in this model optimizer tried but not gives us promising result, even in validation and training loss we can see that the relationship between is not some much they are very far from each other.

Chart, treemap chart

Description automatically generated

Fig 3.2 Confusion matrix for CNN

A picture containing whiteboard

Description automatically generated

Fig 3.3 Above plot shows model accuracy and model loss of Bi-directional LSTM

The correlation between to train and validation accuracy is so much as compare to CNN.

Chart, treemap chart

Description automatically generated

Fig 3.4 Confusion matrix for LSTM

# Discussion

We have build the CNN and LSTM model but before building the model we came up with scores based on various things like popularity of the user, polarity and subjectivity of particular tweet that describes the sentiments of people then we have build the combined graph of price and sentiment score. After doing this project one can get the deep understanding of people’s sentiments on twitter. After completion of this project we can say that both the models are working good but if we look at the epoch-accuracy and epoch-loss graph of both the models we can see that in CNN model lines of train and validation is far from each other and on the other hand in bi-directional LSTM model the epoch-accuracy and epoch-loss graph the lines does not have much gap in between so that could be the ideal situation for the model.

In the confusion matrix of both the model we can see both the models are performing really well as we have high true positive, negative and neutral values. Here, deciding which model works best was little difficult part but after careful observation of research papers and trying different models we come up with these two models and in those two models LSTM seems best fit for the objective that we want to archive.

# Conclusion

### In this report, we studied about the relationship between the tweets sentiment and price of cryptocurrency.As the traders nowadays like to trade in cryptocurrency and this correlation of the tweets sentiment and price of crypto shows the way to predict the fall or rise in the price using the sentiment. We used the public dataset which helps us to beause we cannot able to gather the data and we create graphs that show the relationship for the price and sentiment works for bitcoin price and ethereum price. As our task is to check the correlation of price and tweets of twitter and by working on this we found that it there is an correlation after that we created model on bases of that which gives the sentiment of it.

### Tweets with positive sentement shows the raise in the plot as this method works well but gathering the data is quite hard as creating the application using this it requires the twitter streaming api that helps to feed the tweets into the model and that shows the postive, negative and netural. Future work requires more data upto 2 years data and twitter streaming api which makes it better to make an application using this. Our work just able to get sentiment and on the basis of positive sentiment shows rise and negtive shows fall and suggestion for future work is that we also want to predict price on the base of the tweets and the cost which requires combintion of regression or any price forecasting algorthims and nlp(natural language processing) classification.

# Contributions

Alay- data cleaning, Exploratory data analysis, Researching related work, model building.

## Priyank-data cleaning, Exploratory data analysis, Researching related work, model building.

From this project both team members did everything from scratch together so the contribution is almost same for the whole project.

##### References

1. Matta, M., Lunesu, I., & Marchesi, M. (2015, June). Bitcoin Spread Prediction Using Social and Web Search Media. In UMAP workshops (pp. 1-10).
2. Wołk, K. (2020). Advanced social media sentiment analysis for short‐term cryptocurrency price prediction. Expert Systems, 37(2), e12493.
3. Huang, X., Zhang, W., Tang, X., Zhang, M., Surbiryala, J., Iosifidis, V., ... & Zhang, J. (2021, April). Lstm based sentiment analysis for cryptocurrency prediction. In International Conference on Database Systems for Advanced Applications (pp. 617-621). Springer, Cham.
4. R. Abraham, J., Higdon, D., Nelson, J., & Ibarra, J. (2018). Cryptocurrency price prediction using tweet volumes and sentiment analysis. SMU Data Science Review, 1(3), 1.
5. Li, T. R., Chamrajnagar, A. S., Fong, X. R., Rizik, N. R., & Fu, F. (2019). Sentiment-based prediction of alternative cryptocurrency price fluctuations using gradient boosting tree model. Frontiers in Physics, 7, 98.
6. Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system. Decentralized Business Review, 21260.
7. Sharma, T. K. (2018, July 23). *How does blockchain use public key cryptography?* Web3.0 & Blockchain Certifications. Retrieved April 24, 2022, from https://www.blockchain-council.org/blockchain/how-does-blockchain-use-public-key-cryptography/#:~:text=Public%20Key%20Cryptography%20in%20Bitcoin&text=Bitcoin's%20protocol%20uses%20what's%20called,key%20and%20corresponding%20public%20key.
8. Best, R. de. (2022, April 20). *Bitcoin market Cap 2013-2022*. Statista. Retrieved April 24, 2022, from https://www.statista.com/statistics/377382/bitcoin-market-capitalization/#:~:text=In%20April%202021%2C%20the%20Bitcoin,U.S.%20dollars%20in%20June%202021.
9. Twitter. (n.d.). Twitter. Retrieved April 24, 2022, from https://twitter.com/
10. *Reddit*. reddit. (n.d.). Retrieved April 24, 2022, from https://www.reddit.com/
11. Facebook. (n.d.). Retrieved April 24, 2022, from https://www.facebook.com/
12. Bollen, Johan, Huina Mao, and Xiaojun Zeng. "Twitter mood predicts the stock market." Journal of Computational Science 2.1 (2011): 1-8.ence,
13. Zacharias, C., & Poldi, F. (n.d.). Twintproject/twint: An advanced twitter scraping & OSINT tool written in python that doesn't use Twitter's API, allowing you to scrape a user's followers, following, tweets and more while evading most API limitations. GitHub. Retrieved April 24, 2022, from https://github.com/twintproject/twint
14. Kash. (2022, April 18). *Bitcoin tweets*. Kaggle. Retrieved February 27, 2022, from https://www.kaggle.com/datasets/kaushiksuresh147/bitcoin-tweets
15. aroussi, R. (n.d.). *Ranaroussi/yfinance: Download market data from Yahoo! Finance's API*. GitHub. Retrieved March 18, 2022, from https://github.com/ranaroussi/yfinance
16. NLTK. (n.d.). Retrieved March 24, 2022, from https://www.nltk.org/
17. Costa-Luis, C. (n.d.). *TQDM/TQDM: A fast, Extensible Progress Bar for Python and Cli*. GitHub. Retrieved March 20, 2022, from https://github.com/tqdm/tqdm
18. loria, S. (n.d.). Sloria/Textblob: Simple, Pythonic, text processing--sentiment analysis, part-of-speech tagging, noun phrase extraction, translation, and more. GitHub. Retrieved April 22, 2022, from https://github.com/sloria/TextBlob
19. McKinney, W. (n.d.). *Pandas*. pandas. Retrieved April 25, 2022, from https://pandas.pydata.org/
20. *TensorFlow*. (n.d.). TensorFlow. Retrieved April 25, 2022, from https://www.tensorflow.org/
21. Trask, A. W. (2019). *Grokking deep learning*. Manning Publications.