

**CRYPTOCURRENCY SENTIMENT ANALYSIS USING PYTHON AND TWITTER**

|  |  |
| --- | --- |
| Nitishkumar S | 01JST18CS072 |
| Pratham A | 01JST18CS089 |
| Kiran Kumar H | 01JST18CS050 |
| Dhanush R | 01JST18CS032 |
| Mohith K S | 01JST18CS045 |

**CONTENT**

1.

1.1 INTRODUCTION

1.2 TWITTER

1.3 SENTIMENT ANALYSIS

1.4 CRYPTOCURRENCY

2. CASE STUDY

3. DATA AQUISATION

4. DATA CLEANING

5. DATA MANIPULATION

6. DATA VISUALIZATION

7. CONCLUSION

**ABTRACT**

Social media has its influence in pretty much every field, cryptocurrency is no exception. Many influencers like snoop dog, Justin Bieber, **Elon Musk** etc., had tweeted or talked about **Dogecoin** during the end of 2020. This led to surge in its price and people started investing in it.

Crypto market is volatile, which means nothing is stable. So inevitably the price fell and lot of people lost their life savings.

This project aims at giving people insights about how tweeter users feel about a particular crypto coin by analyzing the tweets. This helps them take sensible decision on whether to sell, hold or buy that coin. Thus, saving them from going bankrupt.

**1.1 INTRODUCTION**

As internet is growing bigger, its horizons are becoming wider. Social Media and Micro blogging platforms like Facebook, Twitter, Tumblr dominate in spreading encapsulated news and trending topics across the globe at a rapid pace. A topic becomes trending if more and more users are contributing their opinion and judgements, thereby making it a valuable source of online perception. These topics generally intended to spread awareness or to promote public figures, political campaigns during elections, product endorsements and entertainment like movies, award shows. Large organizations and firms take advantage of people's feedback to improve their products and services which further help in enhancing marketing strategies. Thus, there is a huge potential of discovering and analyzing interesting patterns from the infinite social media data for business-driven applications.

**1.2 TWITTER**

**Twitter** is an American microblogging and social networking service on which users post and interact with messages known as "tweets". Twitter was created by Jack Dorsey, Noah Glass, Biz Stone, and Evan Williams in March 2006 and launched in July of that year. Twitter is an online networking site driven by tweets which are 140-character limited messages. Thus, the character limit enforces the use of hashtags for text classification. Currently around 6500 tweets are published per second, which results in approximately 561.6 million tweets per day. These streams of tweets are generally noisy reflecting multi topic, changing attitudes information in unfiltered and unstructured format. Twitter sentiment analysis involves the use of natural language processing to extract, identify to characterize the sentiment content.

**1.3 SENTIMENT ANALYSIS**

Sentiment analysis is the prediction of emotions in a word, sentence or corpus of documents. It is intended to serve as an application to understand the attitudes, opinions and emotions expressed within an online mention. The intention is to gain an overview of the wider public opinion behind certain topics. Precisely, it is a paradigm of categorizing conversations into positive, negative or neutral labels. Many people use social media sites for networking with other people and to stay up-to-date with news and current events. These sites (Twitter, Facebook, Instagram, google+) offer a platform to people to voice their opinions. The sentiment analysis has wide applications and include emotion mining, polarity, classification and influence analysis.

Sentiment Analysis is often carried out at two levels

1) coarse level

2) fine level

In coarse level, the analysis of entire documents is done while in fine level, the analysis of attributes is done . The sentiments present in the text are of two types: Direct and Comparative. In comparative sentiments, the comparison of objects in the same sentence is involved while in direct sentiments, objects are independent of one another in the same sentence.

**1.4. CRYPTOCURRENCY**

A cryptocurrency is a digital or virtual currency that is secured by cryptography, which makes it nearly impossible to counterfeit or double-spend. Many cryptocurrencies are decentralized networks based on blockchain technology—a distributed ledger enforced by a disparate network of computers. A defining feature of cryptocurrencies is that they are generally not issued by any central authority, rendering them theoretically immune to government interference or manipulation.

Bitcoin is a type of cryptocurrency. There is no physical bitcoin, only balances kept on a public ledger that everyone has transparent access to. All bitcoin transactions are verified by a massive amount of computing power. Bitcoin is not issued or backed by any banks or governments, nor is an individual bitcoin valuable as a commodity. Despite it not being [legal tender](https://www.investopedia.com/terms/l/legal-tender.asp) in most parts of the world, bitcoin is very popular and has triggered the launch of hundreds of other cryptocurrencies, collectively referred to as [altcoins](https://www.investopedia.com/terms/a/altcoin.asp). Bitcoin is commonly abbreviated as "BTC."

**2.** **CASE STUDY**

During the years 2020 and 2021, due to the covid-19 pandemic and lockdown people got more exposure to cryptocurrency trading other than the stock market trading. As we can see there is much more of profit in cryptocurrency trading than in stock-market which comes with so much of the risk in losing money and cryptocurrency trading is so much volatile compared to stock market trading. As cryptocurrency didn’t make much of its buzz in 2010’s but it was growing slowly with little amount of attention to it. The “Bitcoin” is the most valuable cryptocurrency in the whole cryptocurrency world with Ether with it. Doge coin as introduced earlier, it started growing slowly. As the interest of people started to shift towards the cryptocurrency trading world some the influencers of the world started giving advices to the people on which coins, they should invest in, in order to get profit. The profit was in two ways long term profit and short-term profit.

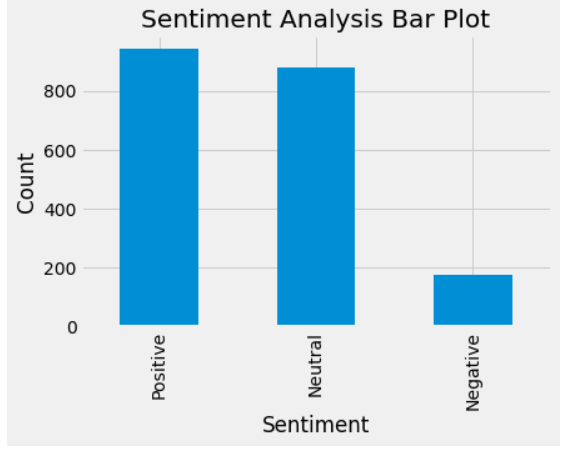
The influencers started providing their advices in form of “tweets” using specific # in order to reach the people. As people saw ample number of insights on the cryptocurrency people started investing in some of the coins as the hype was created. Elon Musk is an entrepreneur and business magnate. He is the founder, CEO, and Chief Engineer at SpaceX and investor, CEO, and Product Architect of Tesla. He has a large impact on crypto world especially on the rise of the Doge Coin. In the months of February and April of 2021, Elon musk tweeted some of the tweets related to the Doge coin. Here are some of the snippets.





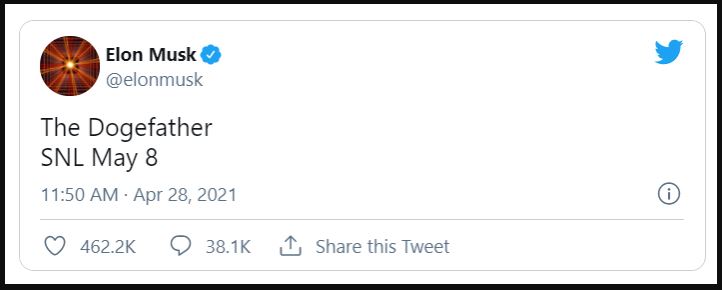
So, these tweets hyped up to the doge coin to the peak. From this tweets people who were new to this crypto trading started investing blindly on doge coin with large amount of money believing it will reach mark of Rs.150 from Rs.20. Though the price reached Rs.60 it did not move further. By using our twitter sentiment analysis, we got the insights of the tweets during that period.

As of this bar plot the positive and neutral responses were nearly same. After this hype SNL wanted to conduct an event with Elon Musk in April. People thought there is going to be some big announcement and large spike in the doge price and they kept on investing on doge.

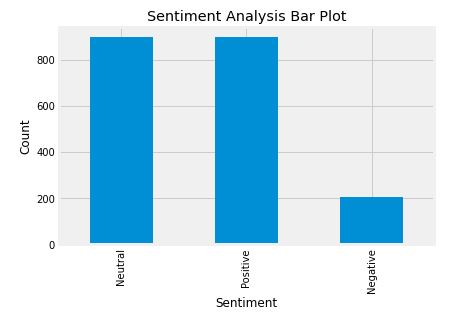


Elon musk again tweeted something related to doge coin, SpaceX and the SNL event. The SNL event was dated on May 8th 2021.Till then the hype had come down and people who knew about cryptocurrency predicted that doge won’t reach or won’t leave up to its expectations. People were confident and kept faith and held the coin. In SNL event nothing new happened and the doge price fell very badly. People lost lots of money.



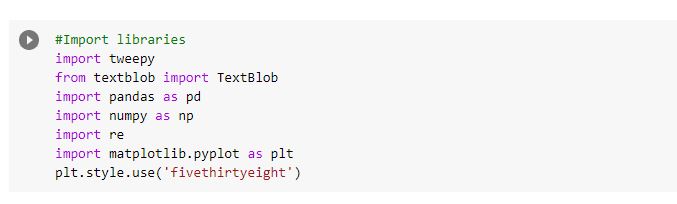


As we tried to extract the sentiment analysis after May 8th positive results came down compared to neutral results.



**3. DATA AQUISATION**

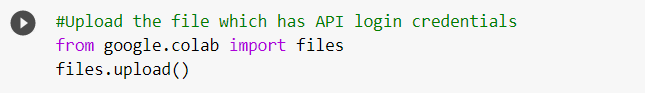
This program gets the sentiment of Bitcoin from Twitter users.



Here, we have imported some of the libraries in order to proceed with the sentiment analysis. The first of the libraries includes **Tweepy.** It is an open source Python package that gives you a very convenient way to access the Twitter API with Python. Next the **TextBlob** is a Pythonlibrary for processing textual data. **Pandas** is an open-source library that allows to you perform data manipulation and analysis in Python.

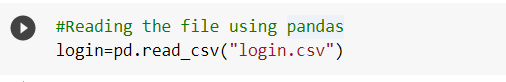
**NumPy** is an open-source numerical Python library. NumPy contains a multi-dimensional array and matrix data structures. It can be utilised to perform a number of mathematical operations on arrays. A **regular expression(re)** is a special sequence of characters that helps you match or find other strings or sets of strings, using a specialized syntax held in a pattern. **Matplotlib** is an amazing visualization library in Python for 2D plots of arrays and an extension of Numpy. At last we have set the graph plot style to **‘fivethirtyeight’.**

Uploading the file which has API login credentials



In order to upload the .cvs file we imported files from google colab.

Reading the file using pandas



Getting twitter API credentials from login file



Here, the join() method provides a flexible way to create strings from iterable objects.



Creating authentication object using tweepy. OAuthHandler function which is used to callback URL that needs to be supplied dynamically.

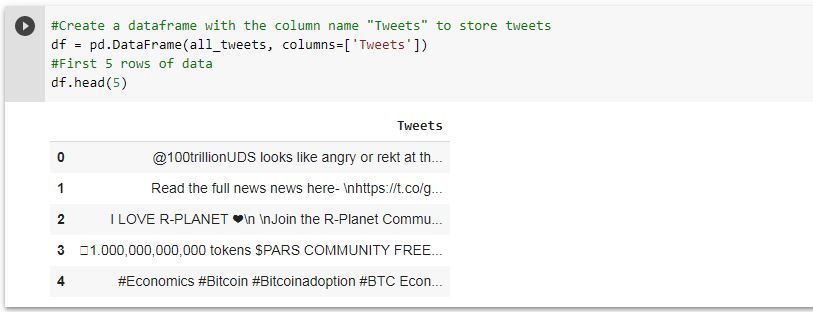
After creating authentication, we need to set the access token and access token secret for using the twitter api using “authenticate.set\_access\_token()”.

At last, we need to create an API object using tweepy. API function by using the above credentials.

Next step is to gather tweets about bitcoin and filtering out the retweets. This can be done by search\_term function which searches the term #bitcoin.

After this we have to create a cursor object in English language since the date 2021-05-01. This gathers about 2000 tweets in English language of #bitcoin.

After gathering of the tweets, those tweets are stored in a variable ‘all\_tweets’.



Here, we are creating a dataframe using pd. DataFrame() function which takes two parameters from where the content is to be drawn from and how it should be stored. Here, we store it in columns format with the name ‘Tweets’ and storing them in the variable df.

The df.head(5) gives the first 5 columns stored in the variable df. This are the 5 of the tweets which are displayed.

**4. DATA CLEANING**



Now, we are defining a function called ‘cleanTwt()’ this function helps in refining the tweets with re.sub() function. The re.sub() function is used to replace occurrences of the particular sub-string with another sub-string. In this function, we are removing the

‘#’,’\\n’(next line),’http?:\/\/\S+’(hyperlinks),’@’ and replacing with sub-strings and ‘ ‘.Storing them in a variable ‘twt’.

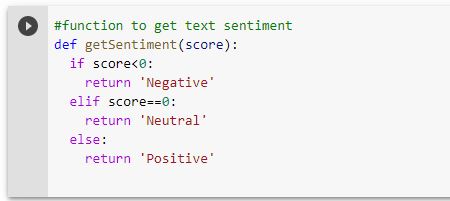


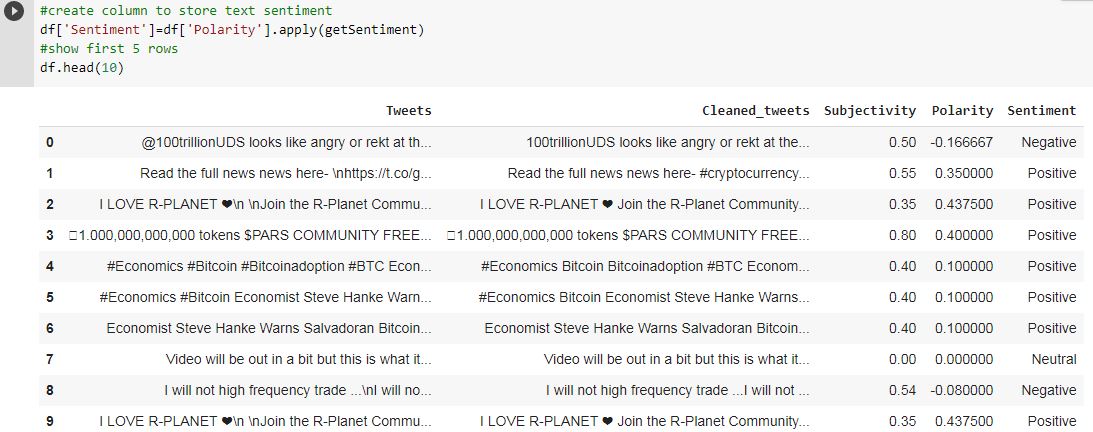
The new data frame is created with new column ‘cleaned\_tweets’ and all the tweets in ‘Tweets’ column is filtered and copied into a new column ‘Cleaned\_tweets’. The df.head(5) gives the 5 tweets stored.

**5. DATA MANIPULATION**



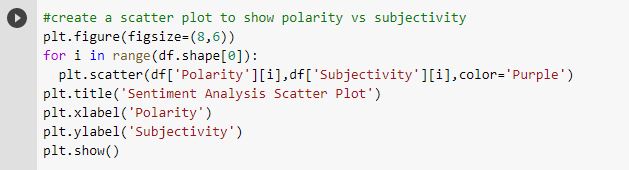
Now, we are defining a new function getSubjectivity() and getPolarity(). The sentiment function of textblob returns two properties, polarity and subjectivity. Subjective sentences generally refer to personal opinion, emotion or judgement whereas objective refers to factual information. The aim is to find the opinionative data and classify it according to its polarity, i.e., positive, negative or neutral feedback. The subjectivity reads the tweet and gives us the personal opinion, emotion or judgement based on that the polarity is decided. After the polarity and subjectivity are found they are stored with new columns named ‘Polarity’ and ‘Subjectivity’ respectively and some of the values are displayed.



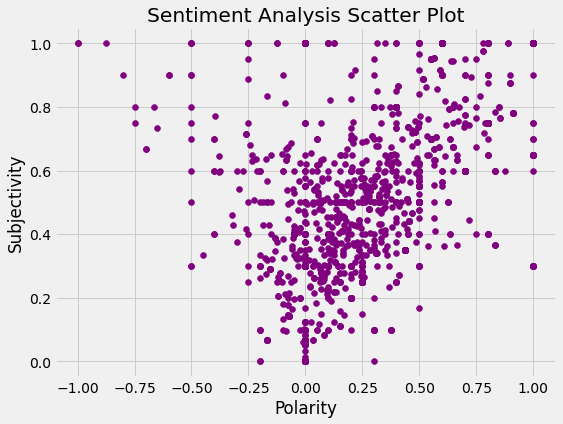
Again, a new function is defined ‘getSentiment()’ where are trying to decide the polarity given to the tweets are positive on the basis of the value of the polarity. This function returns ‘Negative’ if the value of the score in the sense polarity is less than 0,returns ‘Neutral’ if the value is equal to 0 and returns ‘Positive’ if the value is greater than 0. 

Now, the data that we got from the function ‘getSentiment()’ are stored in a new column ‘Sentiment’ which contains ‘Negative, positive or neutral’.

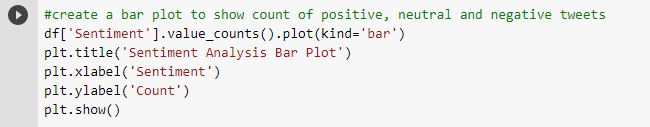
**6. DATA VISUALIZATION**



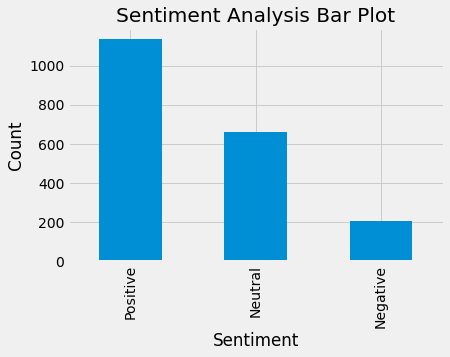
Creating a plot to show polarity vs subjectivity by using some of the function like plt.figure(figsize=()) here the figsize attribute allows us to specify the width and height of a figure in unit inches. The figsize attribute is a parameter of the function figure() which create a figure object. The ‘for’ loops has the range shape[0],where the shape[] function returns the shape of an array where the numbers denote the number of elements per axis of the plot.’plt.scatter()’ has the variables ‘polarity’and ‘subjectivity’ with dots of the colour ‘purple’.



The sentiment analysis scatter plot is plotted between Subjectivity in the y-axis and Polarity in the x-axiz.1 is the highest value of the subjectivity where as 0 is the lowest and all the values varies between 0 to 1. The polarity varies between -1 to +1. As you can see in the scatter plot we can say that most of the values lies in middle of the plot or more towards the positive side of the polarity.



In this final cell, we have tried to create a new plot of graph between the number of counts and the sentiment analyzed column in the form of bar plot.



The sentiment analysis bar plot is plotted between the positive negative and neutral values of the sentiment analyzed and the count.

As you can see it visually the result of the analysis in numbers is more in number for positive results and very less in terms of negative tweets.

**7.** **CONCLUSION**

This analysis technique can be used for predicting the feelings of crypto investors and general public regarding bitcoin, which will help us derive a lot of worthy insights. A good example of insight can be, by using this analysis one can predict the flow of market in near future. One can even pinpoint the price of bitcoin in next few hours or days. We can also decide whether to invest in bitcoin or not. Precisely speaking, 2000 tweets don’t give an accurate picture about the trend of market.

Here we used only 2000 tweets for the analysis, but if we get a lot of tweets ranging over a few days, the result can be even accurate. This can be done twitter allows people to take as much information as one wants but there is a limit on how many tweets can be pulled in a day. If we set the number of tweets to pulled very high and if we have enough days for collecting all those tweets, the analysis can be more accurate.