



# BITCOIN SENTIMENTAL ANALYSIS USING TWITTER DATA.

## Abstract

[1] Since the inception of Cryptocurrencies in the year 2009. Because of its autonomy and convenience, the new type of currency has become a popular and viable source of currency across the world. To serve as an alternative currency, several varieties of cryptocurrencies have been developed. As of today there are upwards of 2000 cryptocurrencies that exist and around 36mil owners are in the US alone. Because of its peer-to-peer and many other benefits, this type of currency is enjoyed by many. Because of its vast popularity it gets influenced by big stakeholders and famous personalities who can sway the worth of these digital currencies in matter of hours. These trend changes can usually be observed through social platforms most discretely. This presents us with an opportunity of predicting and forecasting these trends for one's benefit. This project dives deep into such aspects of the online social platforms and predicts the onslaught of trend changes.

## Project Report by,

Purushothama Rajanna (SUID: 770079902)  
Manoj kumar Gundapaneni (SUID: 239719638)  
Nirmit Pankajkumar Patel (SUID: 992323441)  
Aniket Andraskar (SUID: 417802350)  
Pushkar Nath Mishra (SUID: 770079902)  
Manjunatha sangappa (SUID: 924427618)

## **Table of Contents:**

<b>Introduction .....</b>	<b>03</b>
<b>Significance .....</b>	<b>03</b>
<b>Motivation .....</b>	<b>03</b>
<b>Flow of Project .....</b>	
<b>Tools and Libraries used .....</b>	
<b>Fundamentals:</b>	
What is Sentiment Analysis? .....	04
How Sentiment Analysis Works? .....	04
Sentiment Analysis using VADER .....	06
Data Resources .....	06
 <b>Literature Survey .....</b>	 <b>07</b>
<b>Project Execution .....</b>	
<b>Results .....</b>	
<b>Future Scopes .....</b>	
<b>Conclusion .....</b>	
<b>References .....</b>	

## **Introduction:**

Bitcoin is a decentralized digital peer-to-peer currency which is essentially a payment system created to facilitate financial transactions whilst it being not issued by any financial institution be it government or private.

Bitcoin was invented back in the year 2008. By the year 2009, people started using this very currency as its implementation became open-source. It is a decentralized digital currency without any single administrator or a central bank and can be sent from user-to-user bitcoin network without the need of intermediaries. Transactions are verified by network nodes through cryptography and recorded in a public distributed ledger called a blockchain. Bitcoins are created as a reward for a process known as mining. They can be exchanged for other currencies, products, and services. It's like an online version of cash. You can use it to buy products and services, but not many shops accept Cryptocurrency yet and some countries have banned it altogether. But still in 2022 crypto is considered as a valuable investment asset.

## **Significance:**

The main aim of this project is to analyze twitter data using the twitter API and sentiment analysis. The outcome will be compared with past/ongoing market trends of popular coins such as Bitcoin, Ethereum, Dogecoin etc. to see the importance of social media impact on the crypto world.

## **Motivation:**

Recent research studies have emerged that involve the impact of twitter on bitcoin. We intend to analyze how bitcoin and other cryptocurrencies are being affected by tweets from famous personalities on twitter. In the project we compare the sentiment scores of Twitter text and Bitcoin prices to show how twitter has a significant impact on the cryptocurrency market. The sentiment analysis includes converting tweet text into a sentiment score that is representative of its emotion. In its simplest and most widely used form, sentiment analysis concerns the polarity of the entire text: whether it is positive or negative.

## Flow of Project:

In our proposed timeline for completing the project, we divided amongst ourselves the task for the 4-week schedule.

### Week1:

- Identify popular nodes (Influencers) using twitter API.

Starting with some strong and compelling data was pace setting for our project. This gave us the proper view of how the project was going to shape in its maturity. Popular personalities and verified accounts like the ones of Elon Musk and Warren Buffet skew the trends heavily.

### Week2:

- Fetch tweets using keywords.

As we learned from the assignments, we were able to successfully fetch the tweets using their keywords such as – ‘#btc, #bitcoin, #crypto, #doge, #eth’.

- Perform sentiment analysis on the data set.

Performing Sentimental analysis was the next step. This was performed on the aforementioned popular verified twitter accounts and on the grounds of the likes it has received, how many times it was retweeted based on the #hashtag it has like the ones mentioned above.

These are the only tweets fit for being analyzed and carry any credibility for analysis.

- Plotting graphs.

Graphs were plotted for the results.

### Week3:

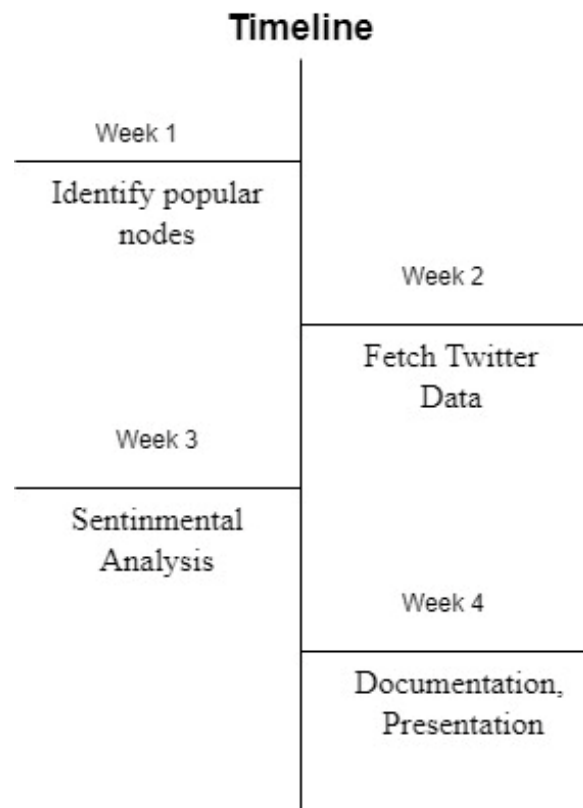
- Compare the outcome of analysis with past/ongoing market trends.

Comparing this with ongoing trends was crucial.

### Week4:

- Documentation and presentation.

After we were satisfied with the outcome and results of the project, it was time to document every last thing in the entire process.



## Tools And Libraries Used:

### 1. Tweepy:

It is an open-source Python package that gives us a very convenient way to access the Twitter API with Python. It includes a set of classes and methods that represent Twitter's models and API endpoints, and it transparently handles various implementation details, such as:

- Data encoding and decoding
- HTTP requests
- Results pagination
- OAuth authentication
- Rate limits
- Streams

### 2. TextBlob:

Python offers us various libraries for performing Natural Language Processing tasks in the most convenient way possible. One of the most prominent and easy-to-use libraries is TextBlob. It is a Python (2 and 3) library for processing textual data. It provides a simple API for diving into common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more.

### 3. Dote NV:

dotenv is a Python module that allows us to specify

environment variables in traditional UNIX-like ".env" (dot-env) file within a Python project directory. Environment variables is the set of key-value pairs for the current user environment. They are generally set by the operating system and the current user-specific configurations. It helps us work with SECRETS and KEYS without exposing them to the outside world and keep them safe during development too.

### 4. OS Module:

In Python the OS module provides functions for interacting with the operating system. It comes under Python's standard utility modules. This module provides a portable way of using operating system-dependent functionality. The `*os*` and `*os.path*` modules include many functions to interact with the file system. It provides functions for creating and removing a directory (folder), fetching its contents, changing and identifying the current directory, etc.

### 5. Matplotlib:

Matplotlib is a multi-platform data visualization library built on NumPy arrays and designed to work with the broader SciPy stack. One of the greatest benefits of visualization is that it allows us visual access

to huge amounts of data in easily digestible visuals. Matplotlib consists of several plots like line, bar, scatter, histogram etc.

6. Pandas:

pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating

numerical tables and time series.

7. NumPy:

NumPy stands for Numerical Python, which is a library consisting of multidimensional array objects and a collection of routines for processing those arrays. Using NumPy, mathematical and logical operations on arrays can be performed. NumPy is a Python package.

## What is a Sentiment Analysis?

[2]Sentiment analysis is text mining (or opinion mining) that finds and extracts subjective information from a given source material, allowing a company to better understand the social sentiment of its brand, product, or service amongst people whilst monitoring online conversations.

However, most social media stream analysis is limited to simple sentiment analysis and count-based metrics. This is equivalent to only scraping the surface and missing out on high-value discoveries. So, how should a company go about grabbing that low-hanging fruit?

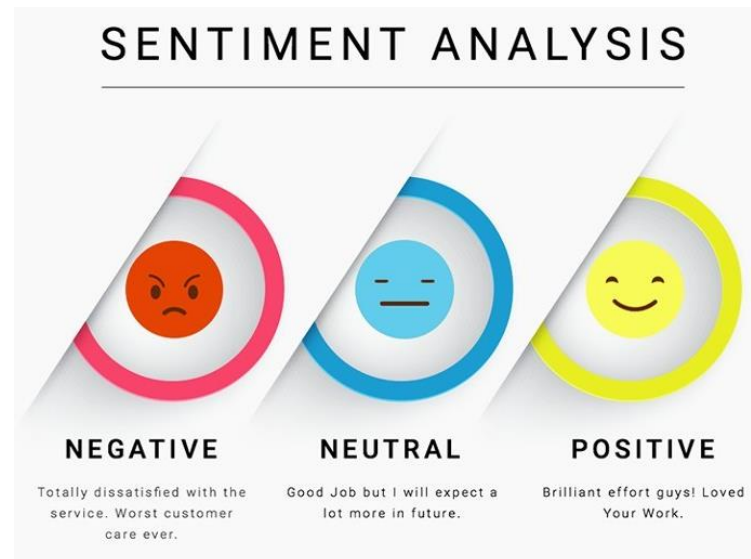
The answer is simple, the researchers have come up with sophisticated deep learning algorithms at the disposal of the companies and brands to maximize their profit margins. The capacity of algorithms to analyze text has greatly increased because to recent advancements in deep learning. Advanced artificial intelligence algorithms used creatively can be a valuable tool for conducting in-depth research.

We believe it is critical to categorize incoming customer conversations regarding a brand along the lines below:

- Customers care about certain features of a brand's goods and service. This is a very good indicator to determine a trend.
- The underlying motivations and reactions of users in relation to those features.

When these fundamental notions are combined, they form a powerful tool for

evaluating millions of brand dialogues with human-level accuracy.



## How Sentiment Analysis Works?[3]

### ○ Training and Prediction Processes:

Based on the test samples used for training, our model learns to correlate a specific input (i.e., a text) with the associated output (tag) during the training process (a). The text input is converted into a feature vector by the feature extractor. To create a model, feature vectors and tags (such as positive, negative, or neutral) are fed into a machine learning algorithm.

The feature extractor is utilized in the prediction procedure (b) to convert unseen text inputs into feature vectors. The model then uses these feature vectors to generate predicted tags (again, positive, negative, or neutral).

- Feature Extraction from Text:

In a machine learning text classifier, the first step is to modify the text extraction or text vectorization, and the traditional method has been to use a bag-of-words or bag-of-ngrams with their frequency.

New feature extraction algorithms based on word embeddings have recently been utilized (also known as word vectors). This type of representation allows words with

similar meanings to have comparable representations, which can help classifiers perform better.

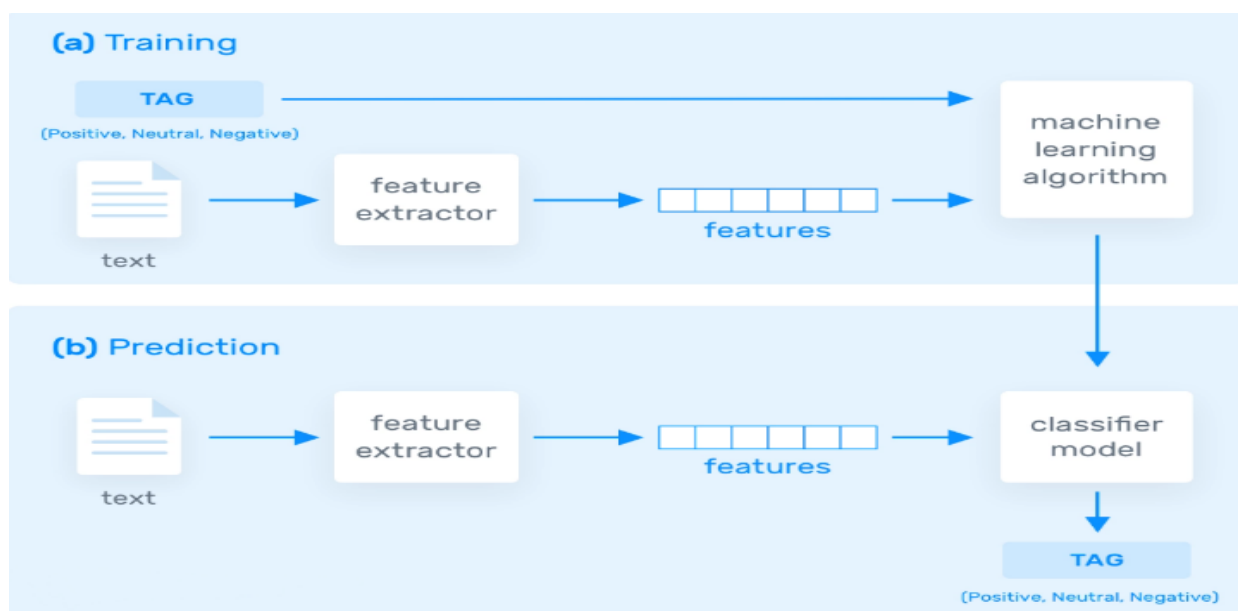
- Classification Algorithms:

A statistical model such as Nave Bayes, Logistic Regression, Support Vector Machines, or Neural Networks is commonly used in the classification step:

- Bayes' Theorem is used to forecast the category of a text in

the Nave Bayes' family of probabilistic algorithms.

- Linear Regression is a well-known statistical approach for predicting a value (Y) given a set of features (X).
- Support Vector Machines (SVMs) are a non-probabilistic model that employs a multidimensional space to represent text instances. Different categories (sentiments) are assigned to different areas within that space. Then, based on similarities with existing texts and the regions they're linked to, new texts are assigned a category.
- Deep Learning is a collection of techniques that use artificial neural networks to process data in an attempt to replicate the human brain.





## Sentiment Analysis using VADER:

Sentiment Analysis by VADER. VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool that is customized to sentiments of social media but also works on texts from other domains.

### [4]Data Resources:

Our project utilizes the data from the most highly visited sites and search engines like Google and the inputs from Google Trends, and social media platforms that actually skew the prices of these digital currencies in certain direction. Another source of data is the twitter API and the volume of tweets. We will talk more about how these data is collected ahead.

Twitter API: Twitter is a major social media platform where users can exchange messages. The Twitter API can be used to mine data. It is necessary to have keys. One can obtain by submitting an application on Twitter's developer site. Many researchers utilize Twitter as a source because they believe it is reliable. The bitcoin market is influenced by public opinion.

Google Trends: Google's data is especially very helpful in setting a paradigm as it is single handedly the most used search engine worldwide. Google amounts for a whopping 74.5% of all the web searches. This means that Google search data can reveal more about what the public are interested in, as well as how interested it is in any specific issue or topic. Data from Google Trends shows how popular certain search keywords are in comparison to other search terms at any given period. Furthermore, the popularity of these search terms can be compared over time. The popularity of cryptocurrencies over the previous five years might be found through Google trends, and this data, when contrasted to the price charts of Bitcoin, is beneficial for our analysis.

## **Literature Survey:**

### **Paper 1:**

#### Price Prediction Methodology Survey[4]:

Sentiment Analysis is a text categorization tool for determining if a sentiment is neutral or positive, negative. There are some academics who have only relied on public mood data to forecast stock market and cryptocurrency prices.

A very good paper that talks about conducting a Sentiment Analysis to predict Crypto Prices.

### **Paper 2:**

The paper[5] conducted a Twitter Sentiment analysis. This was to predict cryptocurrency prices as follows.

The data was evaluated to see if it could be useful in the final model. Because neutral sentiment does not normally reveal a trend for buying or selling, VADER sentiment analysis indicated that tweets are more neutral, which would make the outcomes less efficient as if the public sentiment is neutral. Price was substantially associated with both Google Trends and volume of tweets. The closing daily price of Bitcoin was predicted using a linear regression model.

Regardless of potential price swings, Twitter sentiment on cryptocurrency is mostly positive. Because the data from this paper were obtained when prices were only going up, a more complex model than linear regression could be utilized in future work to improve the results.

There is another very good paper on Social Media Data Prediction by the name,

### **Paper 3:**

This paper[6] proposes a model utilizing machine learning tools. Using machine learning algorithms and social media data, a model forecasted the prices of a few cryptocurrencies. Neural networks (NN), Support Vector Machines (SVM), and random forest were used in this model (RF).

The results of this model suggest that combining sentiment analysis and machine learning methods, it is possible to predict cryptocurrency. Furthermore, some cryptocurrency prices, notably those with a significant following, can be forecast just using Twitter data.

#### **Paper 4:**

While considering social platforms for forecasting these factors the news can also prove to be a very powerful ally. The paper[7] mentioned talks about the impact of News on the Crypto.

Lamon forecasted bitcoin prices using news and social media sentiment. The model employs a classifier to learn feature weights for data labeling. The researchers experimented with linear support vector classification, multinomial Naive Bayes, and Bernoulli Naive Bayes. The best results, however, came from logistic regression. This model properly predicted the biggest price rises and declines. Because this study looked at news and Twitter data separately, a more efficient result may be produced if the algorithm could deal with both types of data.

#### **Paper 5:**

To anticipate[8] the price of Bitcoin, Stenqvist and Lönn looked at public sentiment from a tweet. 2.27 million tweets on Bitcoin were gathered for sentiment analysis in the hopes of predicting a price change in the near future. This is accomplished by just assigning a decline or rise to the severity of aggregated Twitter sentiment change throughout time intervals ranging from 5 minutes to 4 hours, and then shifting these forecasts ahead in time 1,2, 3, or 4 time periods to reflect the relevant BTC interval time. The accuracy of this approach was 83%. Furthermore, only when the mean of sentiment was limited by a minimum 2.2% shift was a prediction produced. In the future, this analysis could be improved by using a domain-specific lexicon, which would result in a more representative sentiment.

#### **Paper 6:**

The paper[9] dives into what they call a LSTM based analysis model on one of China's biggest microblogging platform called Sina-Weibo since it was launched back in 2009.

Xin Huang suggested a sentiment analysis model based on the LSTM.

The data used to determine the sentiment came from Sina-Weibo, China's most popular social media network. A recurrent neural network based on long short-term memory (LSTM) was utilized in conjunction with past bitcoin prices to forecast future price trends. The results showed an accuracy rate of 87 percent. This was 15.4 times greater than the standard autoregression method.

## Paper 7:

The paper[10] deals with analyzing the Stock Market using the same sentiment analysis.

This study used social media data to predict stock market prices and looked at popular sentiment. The association between stock market values and sentiments in social media data is established using a sentiment analysis algorithm.

To predict prices, the moving average approach was utilized as an indication.

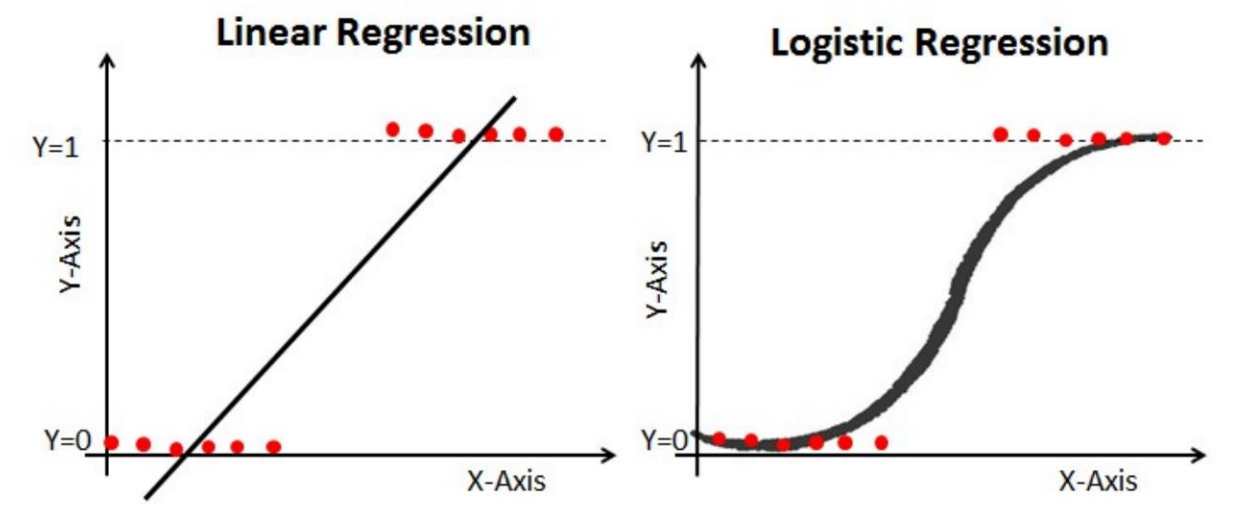
The results demonstrate that using the moving average method in conjunction with sentiment analysis leads in a 14.43 percent higher efficiency than using only the moving average method.

## Machine Learning Algorithms:

The Paper[4] also mentions about two algorithms namely logistic regression and linear regression.

Logistic Regression: By computing the likelihood of each element of the set, logistic regression is used to classify elements of a set into two categories (binary classification). For a given collection of characteristics (or inputs),  $X$ , the target variable,  $y$ , can only take discrete values in this classification issue.

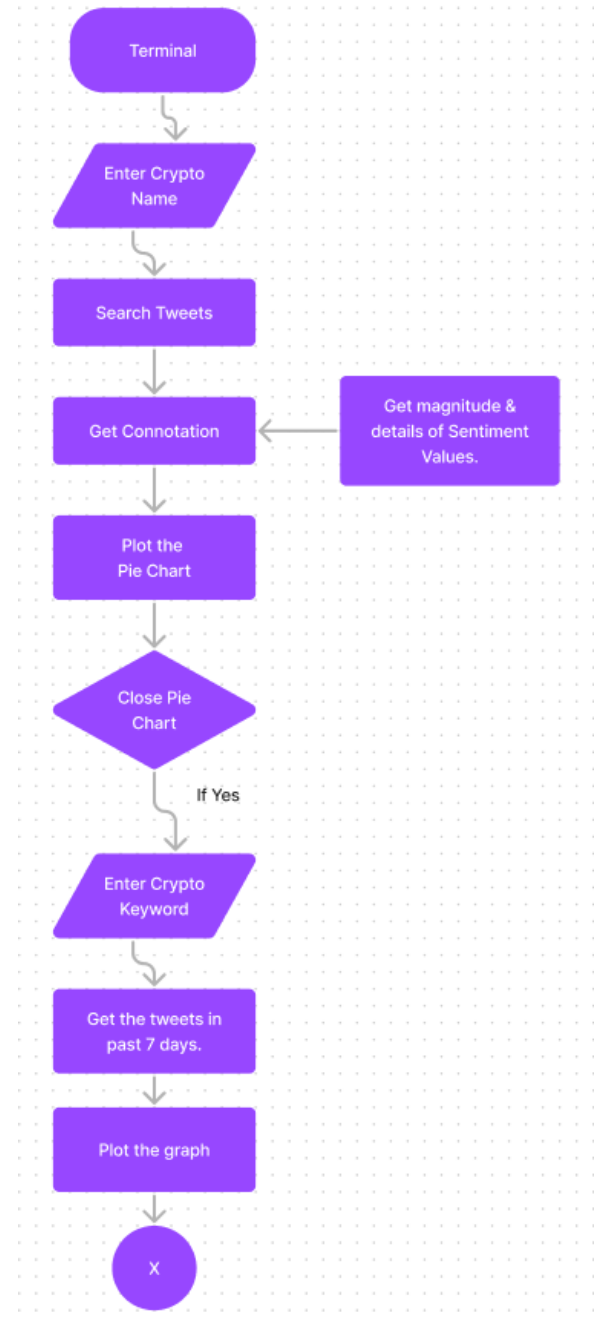
Linear Regression: Linear regression is a supervised regression method-based machine learning algorithm. Based on independent variables, regression models aim for prediction values. It is used to determine the link between variables as well as predictions. The relationship between the dependent and independent variables, as well as the number of independent variables utilized, distinguishes different regression models.



## Project Execution:


### 1.1 Project Flow Chart:

- Terminal:  
The landing screen of our program output.
- Enter Crypto Name:  
Next, we input the cryptocurrency name.
- Search Tweets:  
Using twitter API we search for tweets related to the cryptocurrency.
- Get Connotation:  
Perform Sentimental analysis on the aggregated data.
- Plot the Pie Chart:  
Plot the chart according to the Sentimental analysis data.
- Close Pie Chart:  
Pie Chart must be closed to execute remaining part of the program.
- Enter Crypto Keyword:  
Enter the Crypto keyword again to execute the remaining program.
- Get the tweets in past 7 days.  
Fetch the tweets in past 7 days.
- Plot the graph:  
Plotting the graph.
- End of the execution.



## 1.2

Tweets that include *#bitcoin* and provides you with information on which account tweeted it and also how impactfull a certain tweet has proved to be.

```
puru@puru:~$ python3 smdmMain.py
Select the Crypto Currency Type
What coin are you looking for? bitcoin
BITCOIN
Number of tweets found: 95
-----
Verified: True
saylor tweeted:
-----
Monday morning is time to get back to work. #Bitcoin https://t.co/JLufLXRT9W
-----
# of likes: 87645
-----
# of retweets: 11435
Tweet Value: 77.94153829944709
Account Value: 1.1
Tweet Magnitude: 0
Sentiment: Neutral
-----
Verified: True
nayibbukele tweeted:
-----
#Bitcoin City is coming along beautifully ❤️https://t.co/A6ay8aAREW
-----
# of likes: 34784
-----
# of retweets: 5448
Tweet Value: 36.766151305308746
Account Value: 1.1
Tweet Magnitude: 37.86615130530875
Sentiment: Positive
-----
Verified: True
nayibbukele tweeted:
-----
El Salvador just bought the dip! 
```

### 1.3 Program explanation in brief:

1.3.1 It performs an analysis on the tweet and returns if the tweet is positive, negative or neutral.

```
def getSentiment(score):  
    if score < 0:  
        return 'Negative'  
    elif score == 0:  
        return 'Neutral'  
    else:  
        return 'Positive'
```

1.3.2 This shows how popular the user is. To check user's popularity we are using the follower count of that user and whether or not the user is verified.

```
def find_account_value(tweet):  
    #check verification  
    #check followers  
    value = 1  
    ver = tweet.user.verified  
    if(ver == True):  
        value *= 1.10  
    elif(tweet.user.followers_count > 100000):  
        value *= 1.05  
    else:  
        value += 0.05  
    return value
```

1.3.3 obtaining the magnitude of the tweet by using the polarity and sentiment values. We are using inbuilt functions for extracting the data.

```
def getMagnitude(list_of_tweets, tweet):  
    pol = getPolarity(tweet.text)  
    sen = getSentiment(pol)  
  
    mag = 0  
  
    if(sen == "Negative"):  
        mag = find_account_value(tweet) + find_tweet_value(list_of_tweets, tweet) * -1  
    elif(sen == "Neutral"):  
        mag = 0  
    else:  
        mag = find_account_value(tweet) + find_tweet_value(list_of_tweets, tweet)  
  
    return mag  
  
def getConnotation(list_of_tweets):  
    con = 0  
    for tweet in list_of_tweets:  
        con += getMagnitude(list_of_tweets, tweet)  
    return con
```

1.3.4 Plots the pie chart. Uses matplotlib library for sentiment analysis graph.

```
fig = plt.figure(figsize=(5, 3))
plotArray = [positiveCount, negativeCount, neutralCount]
plt.pie(plotArray, autopct='%1.1f%%', explode=None, shadow=False, startangle=90)
plt.title(user_search)
plt.axis('equal')
sen = ["Positive", "Negative", "Neutral"]
plt.legend(labels=sen, loc='upper right')
plt.show()
```

1.3.5 Picks the date range of data for market price graph. Here we can see that the price trend is picked for last 7 days since the twitter API is fetching tweets of last 7 days itself. So comparing these two data makes sense to derive the impact of twitter data on ongoing market price.

```
def getData(cryptocurrency):
    now = datetime.now()
    current_date = now.strftime("%Y-%m-%d")
    last_year_date = (now - timedelta(days=7)).strftime("%Y-%m-%d")

    start = pd.to_datetime(last_year_date)
    end = pd.to_datetime(current_date)

    data = pdr.get_data_yahoo(f'{cryptocurrency}-{CURRENCY}', start, end)

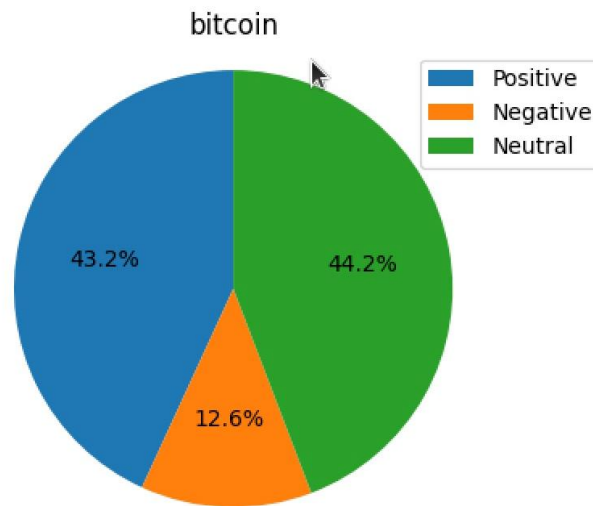
    return data
```



## Results:

We analyzed the trends for the most popular cryptocurrencies.

1. We started with Bitcoin:



**Fig1.1: Pie Chart for Bitcoin.**



**Fig1.2: Candle Stick graph for Bitcoin.**

2. Another Crypto we analyzed the Ethereum.

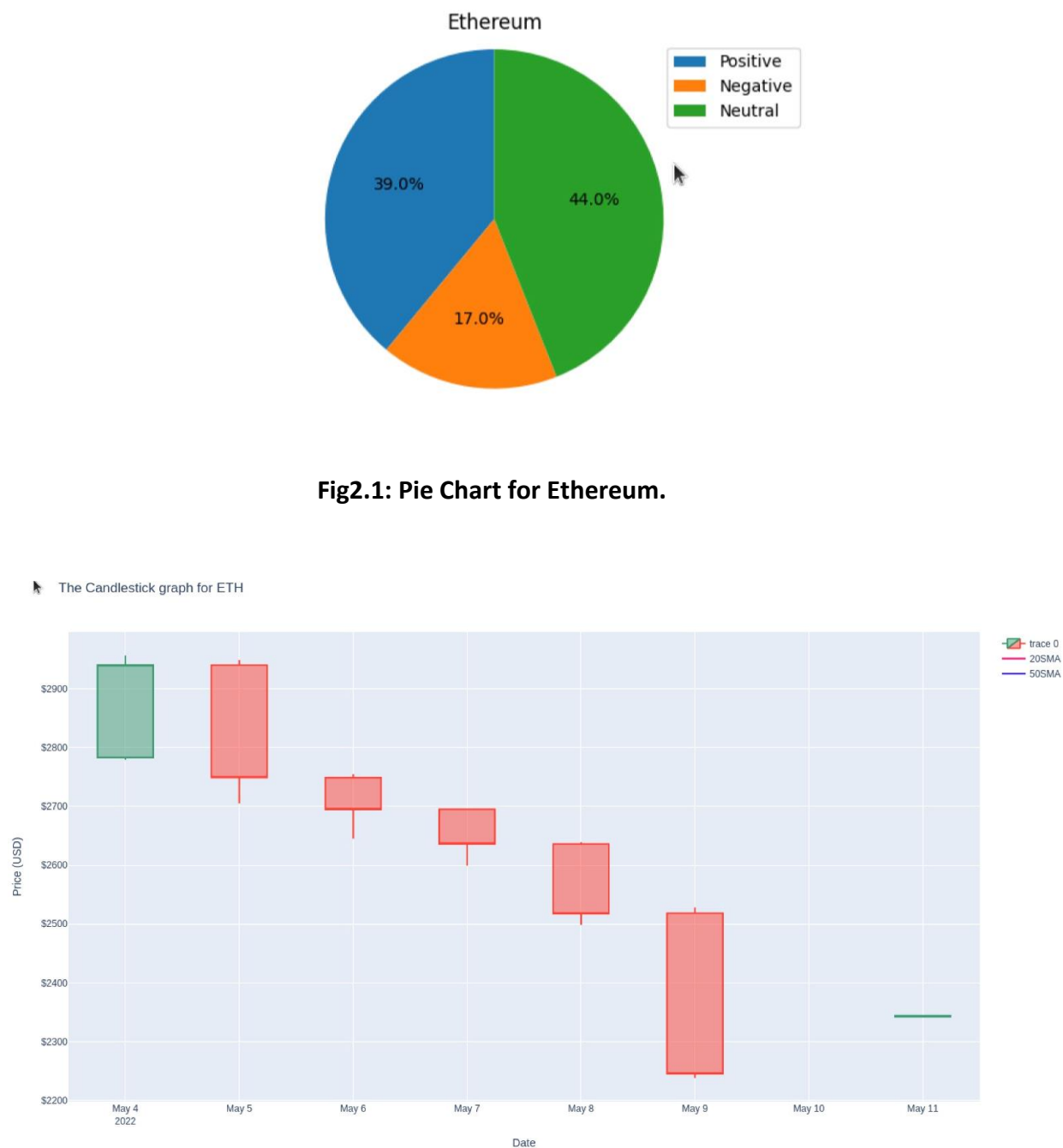


Fig2.2: Candle Stick graph for Ethereum.

## **Future Scopes:**

Forecasting cryptocurrency trends using bitcoin Sentimental Analysis Using twitter data.

## **Conclusion:**

We hypothesize earlier that we can take sentimental analysis of twitter user data and use that data to forecast trends in the prices of different crypto currencies is in fact possible. in our project we compared the data gathered from the twitter's most influential users in crypto space with the recent price history of that particular crypto currency and we found out that the recent fluctuations in the price of a crypto currency is directly correlated with the amount and the type of user engagement it is getting. So, therefore by using the data gathered from the public domains like twitter we can predict future trends to a certain extent. This is just one of many examples that shows us the importance of sentimental analysis. We can use this technology not just to find out trends but also to get an idea about how a certain entity is performing in the given time. Let's take an example of a movie, if we can perform the same sentiment analysis on tweets from influential figures in movie industries and general population to get an idea how that movie is performing and how it will perform in the coming times. This project is a great example of how this technology has potential to change the field of data science and provide crucial insights as well.

We can conclude that use of online social media platforms the likes of twitter by businessmen's, industrialists, big stakeholders, how heavily it skews the ongoing market price by the

## References:

[1] The Economic Impact of Cryptocurrency.

globalEDGE

Author: Kaleb Davis

Published: 11/3/2021.

[globalEDGE Blog: The Economic Impact of Cryptocurrency >> globalEDGE: Your source for Global Business Knowledge \(msu.edu\)](#)

[2] Sentiment Analysis: Concept, Analysis and Applications.

Published in:

Towards Data Science

By, Shashank Gupta

Jan 7, 2018

[Sentiment Analysis: Concept, Analysis and Applications | by Shashank Gupta | Towards Data Science](#)

[3] How Does Sentiment Analysis Work?

MonkeyLearn

[Sentiment Analysis Guide \(monkeylearn.com\)](#)

[4] Cryptocurrency Price Prediction using Sentiment Analysis.

Abdul Rehman Khurshid [akhurs18@earlham.edu](mailto:akhurs18@earlham.edu)

Computer Science Department at Earlham College

Richmond, Indiana, USA.

[Cryptocurrency Price Prediction using Sentiment Analysis \(earlham.edu\)](#)

[5] Jethin Abraham, Daniel Higdon, John Nelson, and Juan Ibarra. 2018. Cryptocurrency price prediction using tweet volumes and sentiment analysis. SMU Data Science Review 1, 3 (2018), 1.

[6] Price movement prediction of cryptocurrencies using sentiment analysis and machine learning.

Franco Valencia, Alfonso Gómez-Espinosa, and Benjamín Valdés-Aguirre. 2019. Entropy 21, 6 (2019), 589.

[7] Cryptocurrency price prediction using news and social media sentiment.

Connor Lamon, Eric Nielsen, and Eric Redondo. 2017.

SMU Data Sci. Rev 1, 3 (2017), 1–22.

[8] Predicting Bitcoin price fluctuation with Twitter sentiment analysis.

Evita Stenqvist and Jacob Lönnö. 2017.

[9] LSTM Based Sentiment Analysis for Cryptocurrency Prediction.

Xin Huang, Wenbin Zhang, Yiyi Huang, Xuejiao Tang, Mingli Zhang, Jayachander Surbiryala, Vasileios Iosifidis, Zhen Liu, and Ji Zhang. 2021.

arXiv preprint arXiv:2103.14804 (2021).

[10] Sentiment analysis for effective stock market prediction.

Shri Bharathi and Angelina Geetha. 2017.

International Journal of Intelligent Engineering and Systems 10, 3 (2017), 146–154.

[11] pandas (software)

[pandas \(software\) - Wikipedia](#)