Analyzing Impact on Bitcoin Prices Through Twitter Social Media Sentiments

by

Jay Patel

A thesis submitted in partial fulfillment of the
Requirements for the degree of
M.Sc. in Computational Sciences

The Office of Graduate Studies
Laurentian University
Sudbury, Ontario, Canada

THESIS DEFENCE COMMITTEE/COMITÉ DE SOUTENANCE DE THÈSE
Laurentian Université/Université Laurentienne
Office of Graduate Studies/Bureau des études supérieures

Title of Thesis Titre de la thèse	Analyzing Impact on Bitcoin Prices Through Twitter Social Media Sentiments
Name of Candidate Nom du candidat	Patel, Jay
Degree Diplôme	Master of Science
Department/Program Département/Programme	Computational Sciences
Date of Defence Date de la soutenance	April 28, 2022

APPROVED/APPROUVÉ

Thesis Examiners/Examinateurs de thèse:

Dr. Ratvinder Grewal
(Supervisor/Directeur(trice) de thèse)

Dr. Ramesh Subramanian
(Committee member/Membre du comité)

Dr. Kalpdrum Passi
(Committee member/Membre du comité)

Dr. Aniket Mahanti
(External Examiner/Examinateur externe)

Approved for the Office of Graduate Studies
Approuvé pour le Bureau des études supérieures
Tammy Eger, PhD
Vice-President Research (Office of Graduate Studies)
Vice-rectrice à la recherche (Bureau des études supérieures)
Laurentian University / Université Laurentienne

ACCESSIBILITY CLAUSE AND PERMISSION TO USE

ABSTRACT

Many cryptocurrencies exist in today's date, and many more are on the verge of being brought into circulation. It is a form of a digital currency but instead of being run by a centralized authority and government, it is a decentralized structure that is created using blockchain technology. These currencies are highly influential and unpredictable with their factors of influence ranging high and low all over the world. This research revolves around the most well-renowned cryptocurrency which is Bitcoin. The focus here is on the discussion around the relationship of bitcoin with the prominent online media platform called Twitter. Twitter has been taking part in the discussion of almost all major as well as related incidents and events all around the world. It is a social media platform that is informative as well as useful for the public so much, that even major personalities, as well as politicians, take to the platform in order to express their views quickly on an important matter. The research included firstly gathering the tweets and was divided into two parts - Verified and Non-Verified users and then a cleaning process was done on the data to make sure that only the desired and necessary data if left for further research. The tweets regarding bitcoin were analyzed and utilized for a deeper observation so that the sentiment can be extracted and can be visualized against the bitcoin prices to derive a conclusion regarding the relationship between Twitter and Bitcoin prices. The analysis returned a lot of insights as well as inference relating to the influence that the Bitcoin prices and related tweets have on each other. The results of the report mention the outcome of the analysis that was found stating the original hypothesis to be true or not

ACKNOWLEDGEMENT

I would like to thank everyone that helped me and supported me throughout this research project to be successful. I would like to start by thanking my mentor and my teacher for the immense help and guidance provided by them for me to achieve my goals as well as the outcome. Firstly, I would like to express my sincere gratitude to my thesis supervisor Dr. Ratvinder Grewal for the constant support of my master's study and research, for his patience, motivation, and extensive knowledge. His advice and guidance provided gave me enough motivation, support, as well as knowledge, and wisdom that helped me to finish this thesis on time. I could not have imagined having a more desirable advisor and mentor for my master's study. To those who will be judging this thesis, I would also like to thank you for providing feedback and your valuable opinions on this research.

I would also like to thank my parents, grandparents, sister, and all my friends for providing me with enough love and support due to which this research was possible. Their love and presence were enough to provide me with the courage to finish this till the end successfully. They provided me with enough support throughout my years of study that enabled me to be knowledgeable and modest enough to finish this project successfully. I would also like to thank my pet for helping accommodate my anxieties surrounding this research that I had. I would also like to thank my relatives for providing me with enough support to complete this research.

This thesis would not have been possible without all of them.

Thank you

Table of Contents

List of Figures

List of Tables

ABBREVIATIONS

API	Application Programming Interface
NLP	Natural Language Processing
NLTK	Natural Language Toolkit
BTC	Bitcoin
ETH	Ethereum
DOGE	Dogecoin
JSON	JavaScript Object Notation
CSV	Comma-Separated Values
VADER	Valence Aware Dictionary Sentiment Reasoning

CHAPTER 1

INTRODUCTION

1.1 Background

Online platforms like social media are the most active communities that engage in discussions about the famous digital decentralized currencies than any other platform. These digital currencies are usually highly secure and utilized for encrypted transactions, hence being called cryptocurrencies. Many Cryptocurrencies exist in today's date, and many more are on the verge of being brought into circulation. It is a form of a digital currency, but instead of being run by a centralized authority and government, it is a decentralized structure that is created using blockchain technology. These currencies are highly influential and unpredictable in nature with their factors of influence, ranging high and low, worldwide. One such currency is known as Bitcoin and is the most popular and most talked-about currency yet. Bitcoin was created in 2009 by Satoshi Nakamoto after he released an open-source paper that discussed a brand-new decentralized structure that can support a form of digital currency. In the beginning, bitcoin was worth almost nothing, and today, its worth is almost $58,000 for just a single coin. This jump-started in the year 2013 when bitcoin, which was flat for most of the past years, suddenly jumped to a value of $250 from a mere ~$0.8. This is when the hype around bitcoin really took to the communities, and people started trading it like a stock market entity on a large scale. People who already had purchased bitcoin previously became millionaires overnight and, not too late, became billionaires in just a year.

This hype also got around the internet in various ways, such as articles, blogs, and the most prominent of them all, social media. This research will provide a similar analysis of social media compared against the famous cryptocurrency itself. This project will focus the discussion on the relationship of bitcoin with the prominent online media platform called Twitter. Twitter has been taking part in the discussion of almost all major, as well as related incidents and events all around the world. It is a news/social media platform that is informative as well as useful for the general public, so much so that even major personalities, as well as politicians, take to the platform in order to express their views quickly on an important matter. Twitter quickly takes the matter that is trending and holds an immediate discussion with or without any scope. These discussions often involve many different sorts of people and are a very strong point for the platform. People can voice their opinion no matter what community they belong to. Trending news and influential topics often involve major people such as politicians as well as celebrities from whom people want to hear the stance on the matter. This information about the influential online platforms can result in quite a few observations that are crucial to a product or an entity and can enforce cause and affect relationships. In this report, the relation of bitcoin with these tweets will be analyzed and observed to find whether a cause-and-effect relationship exists within these two entities. These textual updates can be analyzed and processed using appropriate modules and scripts so that the sentiments and feelings of the posts and updates can be derived.

1.1.1Relationship of Twitter and bitcoin

There is a very ambiguous relationship that exists between the prices of various Cryptocurrencies and the tweets that contain their hashtags and keywords. This relationship is visible through many different sources all over the internet, whether it be experts and researchers or articles and blogs

that track both entities carefully and respectfully. Uploading information on Twitter and any other online platform is a consensual form of providing information to the world, and the data can be used for any research as well as analysis, as long as the ethical concepts of the research are clear and morally straightforward. This information is majorly used for sentiment analysis. These sentiments and feelings can decide the stance of an entire online community, particularly certain products, events, person, and any entity that exists virtually or physically. This sentiment analysis can set a base for the research and comparing against the prices of the bitcoin around the same time period. (Ajmi, Youssef and Mokni, 2022.In conclusion of experts, the crypto market cannot be considered as a safety measure situations since the price fluctuates at such high rates that the economic uncertainty index shoots up on the scale.

In other studies, done through the use of Twitter, the results of the relationship were derived from the number of tweets made a day before, rather than the usual analysis of the sentiment polarity of tweets. The relationship between the investors and the return on the investment on the bitcoin currency, the volatility as well as the volume have been studied by experts in order to study the relationship of the two main entities. (Shen, Urquhart, and Wang, 2019) This format is very different from previously held research since they do not incorporate the volume of the tweets as a significant feature in the analysis. It is found that the numbers of tweets that are uploaded on the platform beforehand are very optimum as well as resourceful in order for ascertaining the number of trades bitcoin will have the next day. The various studies conducted for this relationship conclude that both the volume and the returns are affected by the number of tweets uploaded before the day.

Another (Shevlin, 2021) prominent example of Twitter and crypto relations can be given by the famous and influential people present on Twitter. Certain celebrities, as well as many successful

financial analysts, take to Twitter and social media platforms to announce relative updates online about upcoming factors that can influence the prices. These updates are mostly vague and indirect towards the influence and still can cause a massive effect on the price of the currencies. For example, Elon Musk, a very renowned idol of people and CEO of Tesla and an avid entrepreneur, uploaded a tweet regarding bitcoin in January 2021 that nearly caused the bitcoin prices to rise by a staggering 20%. This kind of sudden rise was only explained by the musk effect that is effective because of his immense number of followers on Twitter. Many blockchain researchers are on the case, constantly measuring the polarity and volume of his tweets in order to ascertain the effect it will have on the price. Musk is not only after a single crypto coin, but instead, he actively tweets about various other Cryptocurrencies, such as the popular meme coin called Dogecoin. His collection also includes Shiba Inu, a currency that saw an increase of 100% in the last few months. Many people believe that Elon not only knows about this effect but also utilizes this for his own profit across the market. Just a slight whim can cause the prices to rise and drop a significant amount, causing a favorable outcome for the entrepreneur.

These relationships are very delicate and uncertain in nature and cannot be derived and used to accurately suggest the direction as well as the intensity of the relationship. A lot of more info is needed, along with appropriate research to capture the true correlation and its coefficient.

1.1.2 Tweet counts of different cryptocurrencies

The number of tweets regarding each cryptocurrency is given in the graph above, and it evidently shows that the most discussed currency is bitcoin itself. This is the reason that the chosen entity for this research is bitcoin. It surpasses most of the other famous and renowned currencies like Ethereum and Cardona.

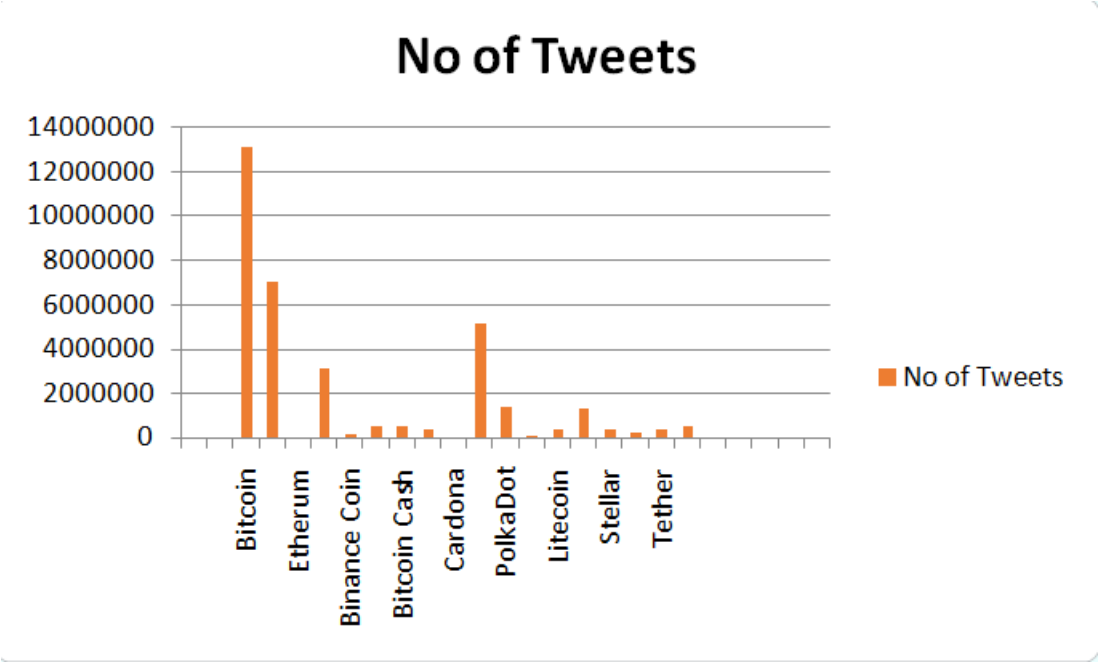

Figure 1: Tweet counts

1.1.3 Bitcoin tweet counts

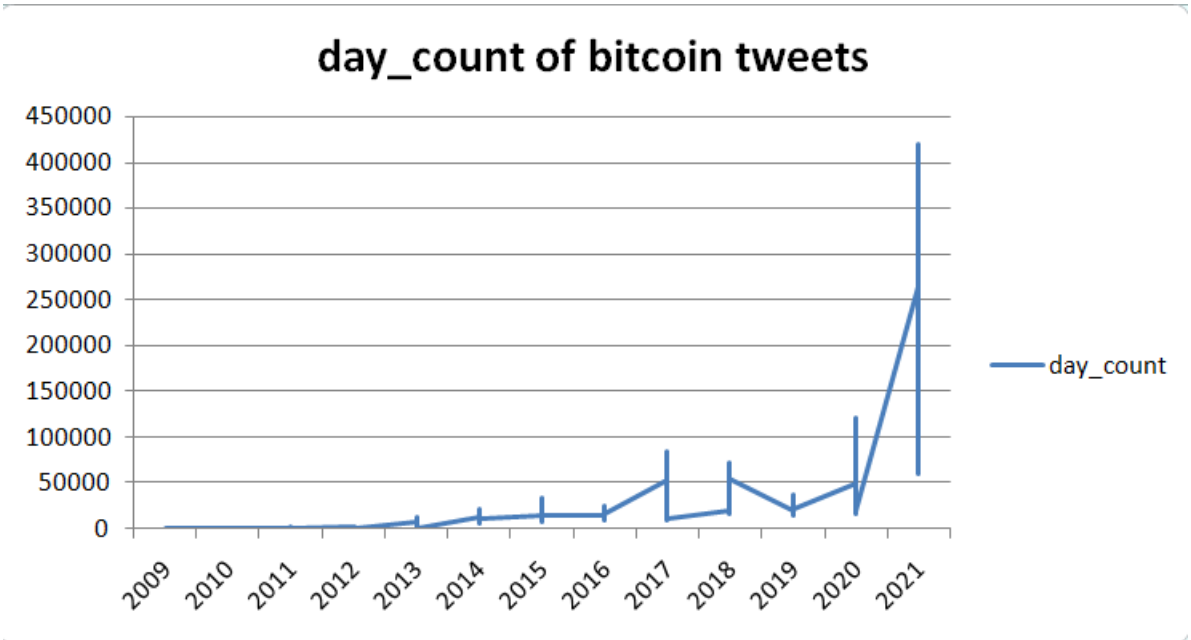

Figure 2: Bitcoin Tweets count

The line chart depicts that the famous currency bitcoin shows a promising increase in the number of tweets per day and also growing per year. The result is due to increased influence in social media, as well as the growing popularity in the currency itself.

1.1.4 Verified users and their impact

On any social media platform, there are generally three types of users. They include the normal public or the everyday user, influencers and celebrities, and business accounts. The need to distinguish between three of them arises since a lot of confusion as well as ambiguity can occur because if a celebrity's original social media handle is not distinguished, a lot of fake impersonators can emerge and claim to be the particular person creating a case of identity theft. Similarly, business accounts also need to be distinguished so that the common public can examine who to reach out to and where to go in need of service. The latter two types of accounts, celebrities and business accounts, are generally created as verified account that contains a distinguishable element to their profile. The standard norm to follow in order to create a verified-looking account is to add a verified symbol in front of the handle. This symbol is generally a blue tick that is global to all of the social media handles and can be found on all of the social media websites on a celebrity's page. Since they are influencing people, the following usually falls between hundreds and thousands of people, sometimes millions hence influencing a lot of people with just one post. Any research similar to ours, these types of influences become very crucial since a single tweet by a verified handle can reach up to hundreds and thousands of different people hence influencing their everyday decisions such as relating to bitcoin. Hence, verified handles need to be examined as well as analyzed so that it can be ascertained how much percentage of the influence on the price of Bitcoin is impacted by a verified handle, for example, Elon Musk.

It is hypothesized that a single tweet from a respected person such as Elon Musk can create a lot of influence and fluctuations in the prices of Bitcoin and other Cryptocurrencies such as dogecoin. People accuse Elon Musk over and over again; start he tweets only to manipulate the price of the cryptocurrency in his own favor. A lot of evidence also points towards the same accusations, such as after Tesla announced that it would be looking for more renewable resource energy and mining and hence dropping Bitcoin out of the transactions for its merchandise since it is very harmful and degrading to the environment, it became clear that Tesla was selling its Bitcoin holdings which sent a wave through the cryptocurrency community and brought down the Bitcoin price by 20%.

Figure 3: Tweet that started the fall of Bitcoin

This wasn't all; after the market continued to decline, Elon Musk again announced that it had not sold any of its Bitcoin holdings, and everything was just an ambiguous mistake.

Figure 4: Tweet for Trying to save the fall of Bitcoin

What continued was a back-and-forth brawl between Elon Musk and the crypto community on Twitter that is believed to have fluctuated the price of the bitcoin even more. The below figure depicts a timeline from May 11 to June 11 on how a series of tweets are believed to have brought down the price of bitcoin by more than $ 20,000.

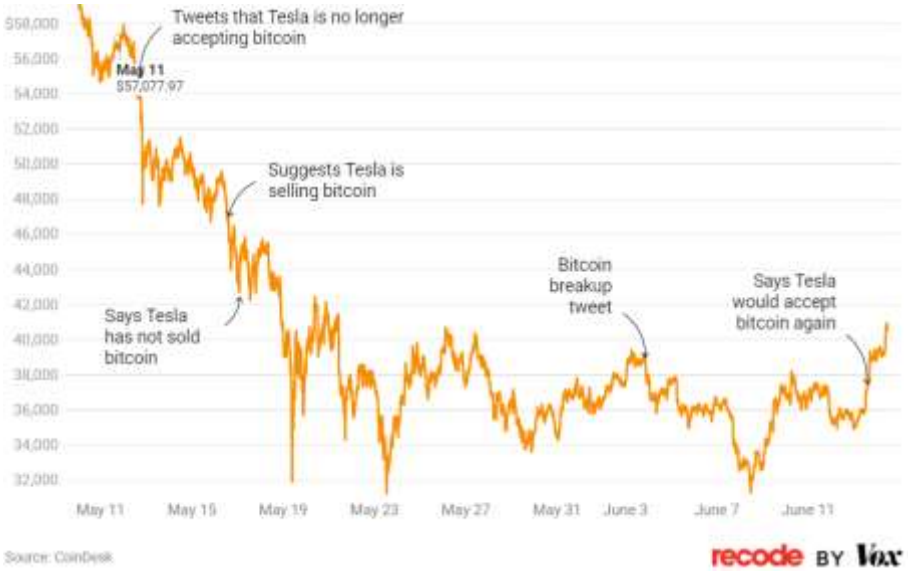

Figure 5: Timeline by CoinDesk showing the effect of a tweet on Bitcoin Price.

1.2 Motivation

Cryptocurrency is one of the most talked-about subjects around the world due to its popularity as a digital currency for the purpose of investment, exchange, or trading like in the stock market. This has also made cryptocurrency price prediction one of the prominent areas for research. A key characteristic of the crypto market is that the fluctuation of its price does not depend on institutional regulations as it is not governed by single entities; but it depends on variables like the dynamics between demand and supply, market capital, attractiveness which relates to people's perception and opinion about the crypto. This leads to further investigation on the variable that fluctuates its prices. With the mass amount of data, information, and opinions laid out on social platforms like Twitter, it is only natural and beneficial to sieve the information from social media for a better prediction model, therefore, leading us to the correlation between social media and cryptocurrency price. The purpose of this research is to find the effect of social media sentiments on price fluctuations between different cryptocurrencies by utilizing Twitter data. The research will be conducted by selecting a cryptocurrency, then gathering tweets for that cryptocurrency, doing sentiment analysis on them, and then comparing them to the price of bitcoin coins during the same intervals to see if there is an effect on the price fluctuation during the peak high and low times. This will help prove whether there is a correlation between the public sentiments and the price or not.

1.3 Objectives

The objectives to be achieved by presenting this respective report are as follows:

- An effective description as well as detailed insights regarding the various aspects of sentiment analysis.

- Presenting the readers with detailed comprehension regarding the various working procedures as well as the implementation of sentiment analysis upon various aspects.
- Apprehension towards the detailed procedures, with the extraction of data- from preprocessing it and manipulating it, to prepare it for the effective sentiment analysis to take place.
- Present the readers with detailed discussions regarding the various aspects to be utilized, like that of Bitcoins as well as Twitter and so on, to provide a briefing regarding what exactly is being utilized within the entire report.
- To present an effective series of outcomes that could clearly distinguish the effects on the prices of Bitcoins based on the various tweets presented by multiple users.
- Draw out definitive conclusions from the conducting of the entire procedure of the sentiment analysis along with the various other tools to be utilized to provide the necessary outcomes regarding the relationship between the tweets posted up on the Twitter platform with respect to the prices of Bitcoin.

From the thesis objective the following research question are derived to which the research will find answers to:

1. Is it just the effect of one single tweet that can cause the price of Bitcoin to fall or rise?
2. If fluctuations are observed in the price of Bitcoin because of the tweets, is it just a couple of times that it is happening or is it a continuously following event?
3. What is the actual number of tweets to be posted by various verified authors so as to experience a drastic or evident rise or fall regarding the popularity of Bitcoin or any particular cryptocurrency?

1.4 Methodology

The methodological approach, which was taken towards the efficient carrying out of sentiment analysis, was starting with the initial phase of scrapping the data to extract the relevant information required to be utilized for the sentiment analysis. The entities which would be extracted while the scrapping of the Twitter feeds comprise the respective unique IDs of the individual along with the usernames of the individual to mitigate the occurrence of duplicity. In addition to this, the most important aspect, which is the date, is also extracted to differentiate between the tweets from those posted within the period apart from those which were not. Furthermore, the most obvious aspect, which is the text written in the tweets, is also extracted to get cleared in the further procedures to be implemented within the detection of sentiments by the sentiment analysis; and the last aspect is the verified and non-verified status. After the completion of the scraping procedure, pandas are utilized to implement the benefits of data manipulation, to implement a low-level Data Structure capable of implementing support to the various multidimensional arrays along with a significant broad approach towards the multiple types of mathematical operations, along with various other procedures necessary towards the efficient manipulation of data to be utilized for the sentiment analysis. After the completion of this procedure, the preprocessing of data is initiated by the utilization of the natural language toolkit. This procedure would allow the cleaning of the entire data to extract only the keyword, which could be utilized for the efficient detection of the emotions while being operated under the sentiment analysis. This procedure would entail the efficient carrying out of the lower casing as well as the special characters, along with the removal of stop words as well as stems. In addition to these, various visualizations would also be provided by the implementation of matplotlib and seaborn, along with the rotting of various types of lots and visual

representations. Lastly, the implementation of sentiment analysis would take place with the assistance of Vader sentiment analysis

1.5 Contributions

Previous research has demonstrated that studying social media sentiments and trends can provide useful findings that can be used to predict public opinion on different topics. The aim of the research is to look into the sentiment trends of the social media platform Twitter for verified and nonverified users and then compare it with the price of bitcoin to see if there is a correlation between them and see if it does play a factor in price functions. This will give an insight into if it is just a single user, a single tweet, or a number of users and tweets playing a factor in the price fluctuation of Bitcoin through their influence. The finding of the research will also help understand that what is the number of totals verified and unverified tweets made along with how many numbers of tweets do actually play a role in price fluctuation. The main issue with the past research was that the researchers used the Twitter data as a whole and did not split them into verified and unverified user tweets. They also did not have access to the whole historical data as that feature was not available until the research was conducted and this is the first time it has been used for the purpose of this research. This also gives access to a wide range of data that will be available for the research. The methodology used for the search will give the findings that will also be useful in future research as well as in future price prediction models along with a combination of other factors to do a precise and close calculation of price prediction of Bitcoin and other cryptocurrencies.

1.6 Relevance and Scope

The relevance towards presenting such a thesis regarding the implementation as well as the display of the practical implementation of the various technologies towards performing an effective sentiment analysis procedure upon the various Twitter feeds related to the Bitcoins in an attempt towards the presenting of their impacts upon the deviation within the prices of the Bitcoins, couldn't be referred to as the presenting of a detailed comprehension to the readers so as to spread awareness regarding the various modern technological aspects. Due to the fact that the technology of machine learning is indeed significantly booming throughout the technological market due to aspect of artificial intelligence has experienced a significant amount of growth throughout the prior years because of its wide range of implementation for getting resolutions throughout the multiple technical issues and problems. In addition to this, various other aspects of machine learning have indeed also seen a significant amount of increment with regards to getting utilized throughout various industries and markets due to the fact that they are indeed capable of providing a significant amount of sophistication to the various procedures throughout various businesses along with providing ease and efficiency by the implementation of the various operations as well as procedures capable towards cutting down the manual labor by a significant amount.

With regards to such aspects of machine learning, the respective report also aims towards presenting the various insights as well as detailed discussions regarding the various operations

along with technical implementations like the natural language processing as well as the natural language toolkit along with how it is utilized for the pre-processing along with the various libraries of python which are utilized for the cleaning and extraction along with the analysis of the necessary data. In addition to all these technical aspects, the report further draws the attention of the readers

to the major technological implementation of the sentiment analysis, which has been popular as well as widely utilized in various industries.

With regards to the scope of the thesis presented within this respective report, it could be said that the scope towards the development as well as making this specific project grow is indeed quite high. This is due to the fact that there are still multiple types of experiments being conducted regarding the various implementations of machine learning, especially the sentiment analysis due to the aspect that it is indeed capable of comprehending the emotions behind the words of the text as well as the speech provided by the human beings, thus making it as a basis towards the development of a much more sophisticated form of artificial intelligence which is capable of converging with humans are much more accurately and efficiently along with behaving itself like a human being. Also, operating upon the data along with its effective comparison would indeed provide various detailed insights along with assisting towards the effective conducting of the respective procedures due to the fact that the wide variety of data would provide a wide range of scope towards drawing out conclusions.

1.7 Thesis Outline

This thesis consists of several chapters, the contents are as follows:

Chapter 1 is the introduction. It mainly introduces the research background of this research, and briefly introduces the main content and organization of this thesis.

Chapter 2 is a literature review. It is a discussion of background on Bitcoin and Twitter along with related work that discussed peer-to-peer papers related to the research.

Chapter 3 is the methodology, which describes in detail the methods and techniques used for the research purpose.

Chapter 4 is data extraction and preprocessing, which is a description of how data was gathered and the steps involved in the preprocessing for getting the data ready for analysis.

Chapter 5 is the results and discussion, which is an in-depth discussion of the results optioned from the research and a discussion on the overall outcome.

Chapter 6 conclusion and future work, which summarizes the results in total and comes to a conclusion along with the future work in the area.

CHAPTER 2

LITERATURE REVIEW

2.1 Background on Bitcoin and Twitter

Upon conducting reviews of multiple types of literature materials, described in detail regarding the various aspects which would be utilized within the conducting of the respective experiments which are to be discussed throughout the report, the various research findings presented and published on the Internet regarding what Bitcoin is as well as insights regarding the workings and operations of sentiment analysis along with it have been widely utilized for various purposes.

2.1.1 Bitcoin

Bitcoin or simply BTC could be explained like that of earth's newest currency which was developed or introduced in the year 2009 by an unknown individual who utilized an area by the name of Satoshi Nakamoto. The transactions which occur utilizing or regarding Bitcoins are indeed carried out without the necessity of any middleman, which means that there is no requirement of the banks. Bitcoin could indeed be utilized in two words the carrying out of various things which would be done in exchange for money or currency like the booking of hotels along with shopping furniture as well as upon placing the money overstocks and much more. Upon conducting further research throughout various pieces of literature, it was identified that some places had been established by the name of Bitcoin exchanges which influence the people towards

buying and selling Bitcoin stations of various currencies. And with regards to the transfer switch, people are unable to send and receive Bitcoin to each other to the utilization of mobile applications and their personal computers. The transfers which occur have been repeatedly reported as easy as sending digital cash. Furthermore, it was also commonly identified to work wonderful kinds of literature that the aspect of mining is indeed a necessary requirement towards the acquiring or getting of Bitcoins initially. Individuals in computers towards the mining of Bitcoins to the utilization of computers and solving various complex mathematics puzzles. This is the procedure for how Bitcoins are developed or introduced.

The Bitcoin mining has been observed as a procedure that could be explained like that of the digital addition of various transaction records within the entire blockchain; otherwise, it could also be explained as the publicly distributed ledger, which comprises the entire history of each, and every transaction related to the Bitcoin. Mining is a procedure that is executed by individuals to the utilization of immense computing powers. The process of mining Bitcoins and that the agent every individual aspiring to be a miner of Bitcoin to the world must contribute towards the decentralized network which has been established as peer to peer, to implement assurance regarding the network of the payment being trustworthy and highly secure.

More than 30,000 tweets per day can we collect it relating to a particular cryptocurrency such as Bitcoin that extends the quality of the analysis confirmed worldwide by experts. Since the hypothesis that the verified handle is still able to predict the price fluctuation of Bitcoin is still under surveillance, the number of tweets per day is also increasing significantly hence increasing the size of the data making the task even easier.

In addition to this, the commonly observed that the impact which Bitcoin introduced throughout the entire market of the world was quite significant throughout various aspects. Multiple reports

are provided which confirmed that more than 2300 businesses across the entire US have indeed accepted the utilization of Bitcoin, running along with observing in a significant increment within the number of companies throughout the entire globe with regards to indulging within the practice of utilizing Bitcoin along with the various other digital assets like that of the host of investment as well as the transaction and operational purposes. The utilization of crypto, which is being carried out throughout various businesses, indeed provides the host with a significant number of opportunities as well as various challenges. It is similar to that of any other frontier where there are both various challenges along with various strong incentives which are unknown. Comparing Bitcoin with other Cryptocurrencies, it is certainly one of the most influential and quietly used objects of transactions that did the rest of the available resources. Other most-active currencies on social media include Ethereum as well as dogecoin, a particular meme coin that gained popularity and has been used in circulation since 2013.

	Bitcoin	Ethereum	Dogecoin
Symbol	BTC	ETH	DOGE
Year developed	2009	2015	2013
Initial purpose	Created to be used as a currency or store of value	Created to sell processing power of the decentralized network	Created as a joke spoof of Bitcoin and the doge meme
Approximate market capitalization*	$893 billion	$451 billion	$26 billion
Number of coins*	18.90 million	118.72 million	132.49 billion
Maximum number of coins	21 million	Unlimited, but issuance is fixed	Unlimited, but yearly issuance limited to 5 billion coins

Table 1: Comparison of 3 different Cryptocurrencies

The table shows the information available as of December 2021. I choose the gap between Bitcoin and the two other prominent Cryptocurrencies that are discussed actively. The market capitalization of the coins is calculated by the total number of coins that are in circulation multiplied by the current trading price. Here it can be seen her capitalization has a huge gap between Bitcoin and Ethereum, with dodge coin not even in the top 5

2.1.2 Twitter

Twitter has been quite popular like that of a well-established platform which is popularly and widely utilized by multiple individuals for microblogging and enabling them towards posting updates along with the utilization attaching videos as well as images. The social media network operator launched in the year of 2006 and has since then acquired more than any amount of 300 million active users on a monthly basis throughout the entire globe. Various literature has presented the reports that multiple individual papers for the sharing of those thoughts as well as news along with the information within real-time along with weather posting of jokes allowed by hotel written in 280 characters of text or in less. This particular microblogging website offers to incorporate various features of the social networking websites like that of my space as known as Facebook, along with the aspect of instant messaging technologies which were developed within networks of the various users who are capable of communicating throughout the entire day utilizing the very brief messages written on Twitter with a regarded as tweets.

The platform of Twitter has since been playing significantly active role-playing the emotions as well as like that of a platform enabling the users towards posting or saying what they feel open. Therefore, the platform of Twitter has been popularly utilized by multiple users towards posting what they are thinking and feeling, without certain situations also provide an impact to the various

aspects and markets throughout the globe. In the essence of this, the plural form of theatre as well as that Tweets posted on it regarding a particular subject or topic has indeed been popularly utilized towards a conducting of sentiment analysis so as to obtain the various types of sentiments hidden behind the words posted by individuals' station of the respective approach of technology.

2.2 Related Work

In this section, we are discussing related works about predicting and analyzing the result of bitcoin sentiment analysis using Twitter along with other discussions as well. We marked that researchers use a different way concerning this problem. A few researchers try to find the relation between tweets made on Twitter and their relation to cryptocurrency prices along with the Bitcoin using sentiment analysis.

2.2.1 Recurrent neural network

According to (Pant et al., 2018), social media consists of several aspects and is being used widely for the different types of community communications and individual communications and use of this social media analysis. Several kinds of these facts can be and rise to the difference of the user behaviors related to the different aspects of the living can be analyzed. The social media analysis and using up the social media channels and data in the adhesive for the beginning of the research and finding the influences that are being returned by them can make the finding of the prediction of the price easier. And here, the other analysis is to be done in an effective and efficient manner. They were utilizing consisting of the capability to infer the society and the various kinds of aspects to the demand-supply. In the making process, the analysis product to the Bitcoin prices based

analysis on the impact on the cryptocurrency all can be considered and analyzed in vitamins of this in an effective manner.

Through using up the various aspects associated with sentiment analysis on the Bitcoin prices by the minutes of the social media platforms of Twitter the different generations and the aspects related to the making of the effect on the prices are all can be considered and can be obtained in such a manner that the different availability of the data can be done in an effective means of this investigation of the fluctuation in the prices can be obtained efficiently. The Twitter sentiment analysis consists of the requirement of various kinds of data from different users. The means of using attitude to date of the different ways to improve the other age groups and their impact on the prices on can be gathered and can we analyzed in such a way that the aspects of their requirement developed and in the reasons behind the possible with the perfect addition to the price are all can be obtained. By the use of the Twitter sentiment analysis, the price fluctuations can be obtained in such a way that targeting the different kinds of the age group audiences and the related persons who show interest in the sentiment and into the keywords that they were searching by such as the Bitcoin and cryptocurrency means of their messages and finding up of the assist with the use of certain types of the words and their positivity rates the aspects related to the increment and decrement and other kinds of fractions into the Bitcoin price here in so that the prediction and the impact of the sentiment on the price fluctuations are all can be find out.

2.2.2 Price prediction using sentiment analysis

According to the (Abraham et al., 2018), In the making of the sentiment analysis by using Twitter data and messages so that the effective prices fluctuations of the Bitcoin can be gathered and is the need of using up certain types of the algorithm’s technologies and by the analysis of the data can

be done and required aspects can find out. As the data consists of two different and a very large chunk of the data, there is a need of using up the multiple types of tools and technologies, and for that, the more precise information is required to be gathered and more accurate predictions to the fluctuating into the prices and fulfilling the process. Here in the use of Twitter sentiment, there is the need for different kinds of sentiment analysis to predict the long- and short-term effects on Bitcoin prices. And hence social media use can be done in an effective manner, and different kinds of prices can be predicted by means of this and its impact on the various kinds of Cryptocurrencies. Yeah, this consists of the data set Feb 2020 to April 2020. There is a need of describing the Twitter data so that the prices are limited and impact and fluctuations onto the Bitcoin can be obtained by means of this time study.

Since social media is a very distinguished platform, a lot of influencing entities also exist in the mix. These influencing entities include celebrities as well as experts and professionals whose opinions as well as the advice can influence a lot of people and hence an entire community such as the cryptocurrency community. This will be analyzed and examined as to if verified people impact the price of the currency more than the general public.

The different types of Cryptocurrencies theatres available in the market have their values increased, and they also don't like the good option for the making of the investment and making profit a profit booking share. And hence as the Bitcoin mining and buy into the other types of this increase in there is a need of making the answer is here so that the impact of the sentiments of the people on the prices of the Bitcoin work can get the summit of understanding the rocks and technology and the cryptocurrency such a way that various aspects related to this and its regulations and its use and impact on this by means of Twitter sentiment analysis the of making up the process in such a manner and on the relationships among the other variables can be cathode

so that the Twitter sentiment analysis can be done in and for the making of the research there is a need of picking up of the use of the data and by means of the collection of the are there is a need of making of the various kinds of the filters data and data processing so that it can be done to the different control algorithms and other means of the using the python. The several other algorithms to the deep learning and making up of the use and finding its patent with the price predictions can be third and based on such kind of the data analysis the means of using up of the volume and the different kind of the generations of sentiment can be analyzed in such a way that it was can be predicted here (Lamon et al., 2017).

2.2.3 Techniques for sentiment analysis of Twitter data

In the paper (Desai et al., 2016), there are several kinds of examinations associated with the making of the cryptocurrency and the Bitcoin and its several concepts tutorial provider discussed, and there are measures that can be taken in for making up of the analysis into the making of the prize predictions by the use of the sentiment analysis of the data also considered into this review. But the making of the story of the cell control is aspects there is the need of making of the Bitcoin and cryptocurrency analysis into the free and the post time duration there is the need of making of this certain country as per so that the returns variable to that is being emphasized into the Bitcoins prices fall and rise all can be done and by means of the positivity rates the price consideration here can be done and, so are the prices can go of the higher and by means of making up of the negative reviews on the majority of the people the thing can be considered in which can be stated that the prices can befall in this case and in the making of the position of the Bitcoin and cryptocurrency is as they are also used in days in and as an investment option there is a need of making of the analysis in such a manner that the effective use of the literature review and such kind of the aspects can be done in making of the investments plans better. Here in this is the need of considering the

different kinds of technical aspects related to the study and there is a need of making up of the descriptive statistics based on the data analysis, and there is a need of taking up the data of the two months. By using the different kinds of keywords for sentiment analysis can be done in and health care in this is obtained in a much larger efficiency.

The analysis will be done in such a way that the demand and supply analysis of the Cryptocurrencies and especially the Bitcoin all can be done and by the determination of the demand and supply of the Bitcoin and cryptocurrency prices rise and fall can be method in under the predictions on this computer in such a way that the positive and the negative impact on the cryptocurrency

2.2.4 In the VADER-Based Sentiment Analysis of Bitcoin Tweets

The making of the sentiment analysis (Pano et al., 2020) by the means of the Twitter data there is the need of maximum of the analysis of the several kinds of the variables making and finding up of the trends based on the changes to the variable and there are several concepts of the aspects are required to be considered and here in the hope of such kind of the address here and there is a need of making up of the finding and this there are certain aspects that are required to be considered in such as the keyword that are stated and the different kinds of the models and statistics can be maintained in such a manner so that the difference between the different variables can be made here and in this the current scenario details can be find out and the based on the impact of the onto the situation and the prices of earth and we chatted and can be used such a way that the different kinds of the keyword and the Bitcoins impact of the cases and that's all can be measured in this and hence the critical value center risk associated with them and all required to be maintained in this and by the means of these the several variables such as the tweets Google trends cases

regarding Bit coins for patients how can be obtained and performance of the different components that are being associated with this.

In the making of the variable and associate in need of finding the estimate sheds and use of the asymptotic value, there is also the need of funding of the different kinds of the aspects related to the making of the approach and making of this certain kind of the graphs so that the impact on the other Cryptocurrencies can be cathedral and this is also the need of finding the effect on to the Bitcoin fluctuation prices. There is the net of seeing the different parts of the long-term and short-term aspects and making up the returning to the Bitcoin based on the different kinds of the trend.

2.2.5 Netizen's opinion on cryptocurrency

According to (Hassan et al., 2021), There are different keywords that are being used consisting of the finding up of the Bitcoin and its availability percentage so that the representative of the different sweets can be done and based on the making of the python charged the sentiment analysis can be main can hear, and fluctuations can be detected. In consideration of making up of the various kinds of the prices here to make of the objective and the neutral percentage here is the need of the analysis, so that the variability can be that and to make up of the comparative study. It is also the rate of using up of the aspects that are associated herein and requirement of their needs to be done in such a manner that the prices and the impact of the different keywords and trends on to the Bitcoin prices can be obtained.

In the making of the sentiment analysis, there is the need for the collection of the data, and hence the data connection requests several kinds of these factors to be considered in it so that the variability in the data can be brought in. The use of the data collection data that is being collected is required to be processed so that it can be prepared for the required format into the making of

defective analysis. In the making of use of the data preprocessing, there is the need of filtering of the data and caring and the gathering only the required aspects from the data and the removal of the several types of complexities that are involved in the data and the unwanted data that is being presented to those. For the making of the data preprocessing more efficient, here is the need of removal of the incomplete data and the entries that are not completed nature and finding of the keyword specific data so that the unwanted efforts that are required to put in can be prevented. To improve the data efficiency and to make the sentiment analysis in an efficient manner, there is the need of making of the data preprocessing and hence is crucial processor and method that is being involved here. After the mixing up of the data preprocessing here, there is the need to extract the different kinds of features from the data, and after that, the procedure is doing followed to the preparation of the test set so that the testing of the various algorithms on those data set can be done. In order to obtain the extraction of the future, there is also the need for the classifiers so that the data can be classified in the different kinds of subsets. Hence more efficient and effective analysis can be done. This is also the need for fore classification of tour based on the quality of the data, and by means of this, there are more chances of getting up analysis done much effective manner. Such means of using of this there is the need of this sentiment analysis, and hence there is a need of using different kinds of the sentiment classification techniques involved and the different kinds of the skilled models to be trained and being made.

2.2.6 Price fluctuations using gradient boosting tree model

Here, the sentiment analysis by means of the Twitter data there is the supervised machine learning algorithms are being used here and consisting of the various kinds of the techniques also uses several concepts such as the negative bias approach maximum entropy and the support vector machine along with the other types of the algorithm such as a random forest and here in this

evolution of the supervised machine learning algorithms and the diameters evaluation and also required. For the use of the need of finding the kinds of the chances to top in fall which consisted of the algorithm such as their understanding complexity theoretical accuracy performance with the number of the observation and the training speed along with the classifier for all the different factors that are required to be considered and before the using up of the different algorithm in making such kinds of the analysis. For making the use of the equation to support vector machine and the algorithm such as the decision tree and the random forest here and there is a need of making of the classifiers and thus requires the labeling of the data and the maximum what criteria are used in here for cleaning up the protections and different route nodes and the finding of the number of the observation along with the classification required to be cut out so that the performance and the tracking of the training speed and sound into the observation are they classified we can we get that with the maximum accuracy obtaining interest (Li et al., 2019).

The maximum entropy that is being used here consists of the total reliability of the probability distribution, and by means of this, the estimated classification of the performance of the technique can be used in the making of the extraction of the different control features this requirement of providing the computer weights and the detective in such a way that the different information related to the data can be gathered and the uniform and non-uniform distribution can be obtained. Here in this the support vector machine as consisting of providing the concept of the winding up of the margin, and by the winds of the support vector machine here, the categorization of the problems can be done here in an efficient manner, and the use of the winding up of the representation and find enough of the maximum margins can be gathered here, and the equivalent analysis of the sentiments can be done.

2.2.7 Bitcoin price prediction through opinion mining

Here in (Cheuque Cerda et al., 2019), The naive Bayes here consisting of the use of these features that can be used for the classification performance can be created by the use of probability classifiers. The evolution of the supervised machine learning and the different parameters that are being informed here is all can you provide to the use of reducing and understanding the complexity that dubbing consists here and understanding the classified problems that are associated with this, and gathering accurate algorithm based on their performance analysis and the probability classified as on the other classified as the linear classifier and the decision-based classifiers. By catering to the information, you are the reason for the use of the natural language processing, and in the middle of the development by means of social media, there is a requirement of handling and emotion detection, and when this the domain-specific documents are required to be considered and pink used to ensure that the analysis and the required outcomes are order to get the research and the various use of the algorithms our help in providing the different criteria. So that the document the research outcome obtained can we have active efficiency and increased performance.

2.2.8 A short-and long-term analysis of the nexus between Bitcoins

In the tweet volume and making up of the cryptocurrency prices prediction, the final voter list is required to be prepared based on the review of the volume of the tweets. Here into the gathering of the price changes and protection of the sentiment with the use of such as pets here and the gathering of the interested to the cryptocurrency, there is the need of using of the technology so that the high and low of the prices can be gathered.

Here, making about the consideration the fluctuation there is also the requirement of volume and based on them the metric sentiment changes and amount of the people the talking about and its

impact on to the Bitcoin price can be gathered, and such broad can be made into the use of the python matplot library. There is also a requirement of considering and making up of the performance and adjustments with the use of python and other types of the algorithms as such uses, and the different price changes are also required to be considered so that the comparative analysis and their research can be done and the objective and the neutral price which can be used here for the slope of the analysis. While making the Bitcoin objectivity and making up of the use of several kinds of the different currency consideration the classification is required of the unlabeled and unclassified data and the feature extraction is also an important part here and the gathering of the negative phrases the parts of speech and the terms and the term presence frequency into the sentence also required to be gathered. By means of getting a confirmation letter to the frequency and the presence of the individual and specific terms into the messages and their analysis, the positive and negative phrases gathering can be obtained based on which the sentiment analysis can be done (Béjaoui et al., 2021).

2.2.9 Review on Bitcoin Price Prediction Using Machine Learning

Here, use of the data and the sentiment analysis of such kinds of the messages but a means of and their impact on to the Bitcoin and the other cryptocurrency prices the laws of the improvements and break down into the demand-supply chains and the markets and other investment methods are also working used in for making up of the analysis so that the various kinds of the aspects that can be obtained. Into the making of the different control these such aspects should have been mentioned here, there is the need of using several aspects of their the stocks and all can be used in here, and in this, the Cryptocurrencies and there jump and fall of the prices can be predicted by the use of this as and such times the unemployment and the demand-supply and the crucial factors that make the people optimized on investing into the cryptocurrency is and which will lead to the

considerable changes into the rise and fall of the prices of Bitcoin. Hearing the different aspects also being involved and by using of officer's kinds of the data and descriptive statistics the different kinds of the values can be obtained which consisting of finding of the mean median and the standard deviation along with the maximum and the minimum sweet trends that are used into the messages and does a number of detections of the cases and the impact of those on the prices also required to be mentioned (Sibel KERVANCI et al., 2020).

2.3 Summary

The initial part of the chapter consists of the background description of Bitcoin and Twitter with a detailed understanding of what bitcoin is along with a comparison with other cryptocurrencies. It also describes the role of users on Twitter and how their emotions conveyed through tweets help in analysis through sentiment analysis. The other part is the replated work which describes how other research related to this we done and list their plus points along with pointing out the flaws as well. There is a discussion on Recurrent neural network, price prediction using sentiment analysis, techniques for sentiment analysis of Twitter data, VADER-based sentiment analysis if Bitcoin tweets, Netizen's opinion on cryptocurrency, price fluctuation using gradient boosting tree model, Bitcoin price prediction through opinion mining, A short- and long-term analysis of the nexus between Bitcoin and review on Bitcoin price prediction using machine learning. The major take from this was learning different methods that were used in previous research and seeing how they lack. The overall understanding from all the research was found that they all used the same time of data which was a general data gathered from Twitter that consists of bots and misleading tweets which was a generalized tweets data that did not found to be effective in finding the actual relation between Twitter and Bitcoin. This led to the objective of the thesis in splitting the data and then finding how the verified users share a relationship with the price of Bitcoin.

CHAPTER 3

METHODOLOGY

In this chapter, there is a detailed discussion on the methodologies used and data filtering done after the data was extracted. The steps involved were as follows:

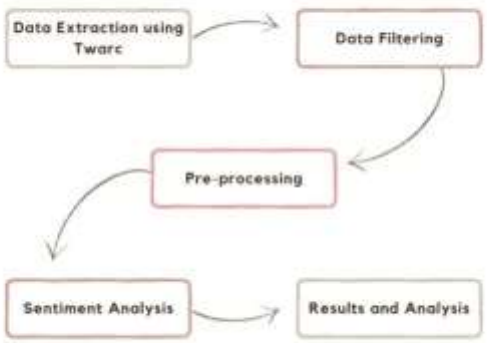

Figure 6: Flow diagram of Methodology

3.1 Data Extraction

Web scraping or simply scaping could be defined as the term which has been popularly utilized for the referencing of a program or an algorithm with regards to the extraction and the processing of a significant quantity of data from the web. It is a popular practice utilized by multiple data scientists as well as engineers and even common individuals who aspire towards the analyzing of

large sets of data. The ability or the aspect of scraping the data from the data cluster upon the internet so as to only acquire or gather the necessary information or data required for conducting the needed procedure could be achieved with utmost efficiency.

3.2 TWARC

Throughout the respective report created for the presenting of conducting effective sentiment analysis of the tweets, implementation of the concepts and operations of scrapping are indeed utilized within the practical implementation by utilizing the module of TWARC (Twitter academic research account) (Mehta et al., 2020).

The TWARC could be explained like that of a scraper utilized specifically for the extraction of the relevant information from that of the Services of the Social Networking. It was utilized towards implementing the scraping function in attempts towards the extraction of the various specific details to be utilized for conducting the sentiment analysis.

For TWARC, it is needed to install the module and configure it with your bearer token and/or API key using the following structure

```
#!pip install twarc
#!twarc2 configure
```

Figure 7: Installing Twarc

TWARC is a package in Python that can be utilized through a command-line interface to extract as well as fetch tweets from Twitter. It uses the Twitter API device for the tweets in a JSON format filtering to all the needs as well as the index. TWARC is a great way to fetch the data since it provides a high-quality representation of a tweet and can be used filtered through older as well as

complex filter tweets will stop. It already does most of the work inside this function, such as switching on the rate limits for the API, waiting for the API penalty time to be over, conversion into different data types such as a JSON → CSV, etc.

And this is such specifically the TWARC module was utilized, which uses version two of the Twitter API that was released in 2021 that was Built for researchers so that they can Fitch and gather tweets from any known archive from Twitter.

For the purpose of collection of tweets based on a specific query, a specific command can be used followed by the query and the desired filters such as:

!twarc2 search "\"Elon Musk\" is:verified" twitter-data/dfw_exact.jsonl

This command specifically asks for tweets related to Elon Musk and sets the filter of verified author equals to true since we only want tweets from a verified account. Moreover, the total data is then extracted and stored into a JSON file called DFW_EXACT.jsonl.

The entities which were extracted comprised a lot of data of which the required data is filtered to limit the data necessary for the research.

3.2.1 The ID of the respective individual tweets.

This was extracted and gathered in attempts towards the mitigation or prevention of duplicity within the gathered information. This particular aspect was implemented to train the machine with a larger variety of data and sentiment behind the tweets posted by various users to make the algorithm significantly capable of distinguishing among the large variety of text regarding the efficient recognition of sentiments (Bharathi et al., 2017).

3.2.2 Username of the Individuals who posted the tweets.

This entity was extracted for the sole purpose of distinguishing the tweets posted by specific individuals and to ensure that the tweets which are being gathered upon a daily basis since the initiation of the designated time period of collecting the Twitter data The usernames were eventually discarded from being utilized in the practical research conducted to let the user who posted the various types of tweets on the platform remain anonymous, thus maintaining their respect integrity (Kamyab et al., 2018).

3.2.3 Date of the tweet being posted on

The date on which the tweets were posted by the various respective users was also extracted. This was done due to the aspect that the tweets which were posted on the Twitter platform were to be extracted according to the time period. The dates of the postoperative platform are one of the most crucial aspects to be utilized within the experiment due to the aspect that the comparison would indeed be implemented prior to the time period along with the impact of both the scenarios upon the prices as well as the market of Bitcoins within the entire financial sector. Automated algorithm subah applied to extract five feeds on a daily basis regarding the keywords Bitcoins and other related terms to both these aspects, due to this particular aspect of collecting the tweet data according to the various dates falling within the time period required for the comparison between the impacts experienced on the Bitcoin prices. This factor enabled the various capabilities towards episode conducting of the various comparative procedure as well as the efficient extraction of the keywords for the conducting for sentiment analysis in the most effective manner (Gayathri et al., 2020).

3.2.4 Text from the Tweets

The text or the entire context of the tweets were also extracted as they would serve the most crucial purpose within the effective carrying out of the Sentiment Analysis. It is due to the aspect that the main procedure and operation of the sentiment analysis could only be carried out by operating upon the cleared and clean data or text. This is due to the factor that the detection of emotion or the sentiments could only be found out by operating upon the text, which is the key value responsible for holding the emotion or the sentiment of the users behind it, who posted the respective tweets he Twitter platform. It could also be explained or comprehended by the notion that only a slight inflection within some individual's voice or words, like that of any poignant comment passed by any individual user upon the internet, could be referred to as passing a sarcastic comment, and similar such situations could be referred to like the various types of complexities encountered on the processing of the Human Emotions. The text to be scrapped or extracted would indeed be utilized towards serving like that of an aspect which could prove to be of significant assistance with respect to establishing a means of communication along with the base towards the effective deciphering of the various feelings or the sentiments which the user might have felt writing or posting the text on Twitter (Das et al., 2018).

3.2.5 Verified Identity

Along with all the necessary information relating to a tweet, another feature that will also assist in making decisions is the verified identity that specifies whether the tweet was made from a verified handle or not. As mentioned earlier in the introduction, verified handles have a huge impact, generally much bigger than what a normal account would have on the price of a cryptocurrency. Following the example of Elon Musk, many influential personalities have also engaged themselves

in updating the community about various types of Cryptocurrencies that then, in turn, affect the price and fluctuate prices in the entire market. It is kept in mind that certain individuals can and will manipulate the market at their will and sometimes for their own benefit or with a bigger plan in mind.

3.3 Pandas for Data Manipulation

Pandas could be explained like that of an open-source library to be utilized within Python. It provides the user with significantly high performance throughout the operation to be performed on the data structure and the analysis of data, as well as the tools required to do so. The Module of Pandas runs upon the top of NumPy along with being popularly utilized for the various operations entailed within the Data Science as well as the Analytics to be carried out regarding the respective data. (Adwan et al., 2020) The role fulfilled by both the Pandas as well as NumPy is that of providing the Low-Level Data structure by NumPy, which is capable of supporting the multi-dimensional arrays along with a Significantly broad reach of the various types of mathematical operations. Pandas can also be defined as that of a higher interface that is utilized towards providing or implementing a streamlining as well as the alignment of the temporal data along with the implementation of significantly efficient functionality with regards to the time series.

After the completion of the successful installation of pandas within the python, there are various types of data structures provided by the founder model which were utilized as per the need occurred. These three times are namely referred to like the series as soon as the date of name and the panel. The panda’s data frame was utilized like that of the most important as well as the popularly utilized data structure along with the aspect that it serves as the standard approach towards the efficient storage of multiple data. A data frame comprises data that has been a line

within a pattern of rows and columns like that of any traditional relational database as well as in a table or the spreadsheet. Furthermore, it also offers a provision towards implementing the hard coding of each and every data into the provided data frame or just simply importing through the utilization of a CSV file. There are several approaches to the creation of an effective data frame, but the main objective behind the development of a data frame object is that of creating it from the provided dictionaries so that it might comprise of a list of various doubles as well as the CSV and excel file and so on.

The first and foremost step which was taken towards the efficient implementation of the data manipulation was that of the developing of an efficient dictionary, followed by a step regarding the passing of the entire dictionary like that of an argument within the data frame method. The third and final step to be carried out is print the respective data frame. Throughout the respective experiment, the date frame to be utilized for indeed created by importing a CSV file. Han CSV file can be explained like that of a text file comprising of 1 record data with respect to each line present within it. The values comprised within the respective record are provided in separations to the utilization of the effective commas in the right places. The next phase which was carried out was that of inspecting the data comprise within the data frame, which encompassed the procedure of running the entire data frame to the utilization of its respective name by displaying the whole table.

For carrying out the efficient analysis of all the data, they need to word the establishment of inspecting the data in the most Ottoman sophisticated manner from the large volumes of the provided sets of data is indeed significantly crucial. (Pawar et al., 2020) With regards to providing assistance with this procedure, the panda's library or the module provides its users with various assist functionalities, which proved to be of significant help with regards to implementing inspection throughout the data that the user is in need of. After two successful completions of this

step, the creation of the statistical summary of each and every record is to be started. The statistical summary code indeed is developed or acquired on the basis of the data to theatre laceration of the df. describe() function. The respective function would then be utilized towards displaying the entire statistical summary regarding the percentage of effectiveness on the Bitcoin prices corresponding to the tweets posted of the data stored in columns. Then computer efficient sorting of each and every record, which can be achieved throughout each and every column through the utilization of df, sort_values() function. (Biwas et al., 2020) After the completion of the above-discussed aspects or steps, the necessity towards mentation of slicing the records as well as filtering the data is very crucial.

In order to efficiently slice the records and extract the data from within a particular column, the slicing of records could indeed be utilized by leveraging the column name. In addition to this, multiple columns can indeed be sliced at a single moment by preventing the inclusion of the entire number of various column names which have been enclosed within the accurate square brackets along with the various names of the collars which have been separated by the utilization of commas. For the filtering of the entire data, it is indeed quite possible to achieve the efficient filtering of each and every value stored within the columns to the implementation of the comparison operator, which is used towards fetching and filtering of the data, on the basis of the provided conditions. Followed by these procedures, conducting the efficient data wrangling is indeed a crucial aspect throughout the field of data science which comprises of the various types of processing to be implemented on the data so as to make the respective data-capable towards working efficiently with being integrated with various types of data algorithms. The data engine can be explained as a procedure implemented towards the processing of data along with introduced in various aspects like merging or concatenating as well as the grouping of the various data.

3.4 Data Pre-Processing by NLTK

Natural language processing or simply referred to as NLP could be explained as that of a unique subset comprised within the technical limitation of machine learning which is concerned with the aspect of the real-life data, which is unstructured in nature. Even though the computer machine is not capable of working efficiently identifying along with processing the various inputs provided in the format of strings, however the libraries within the natural language processing like that of the NLTK, which is the natural language processing toolkit, along with similar to school interview utilized implementing the efficient processing of all the inputs provided in the format of strings in a mathematical approach. (KEMALOĞLU et al., 2021) Twitter social media platform serves like that of the interface throughout multiple individuals aspire towards expressing their individual feeding by posting their individual parts upon some specific context. It is due to this aspect that the most capable platform from which the various types of text comprising of multiple data sources from various users could be extracted for conducting the sentiment analysis on it. There are various procedures to be followed prior to the starting of the sentiment analysis, which is necessary for the efficient pre-processing of the entire data.

3.4.1 Lowercasing

Prior to the initiation of the processing as well as operations and details within the carrying out of efficient sentiment analysis regarding each and every review, it is significantly important to implement the cleaning of the entire textual data. Text data regarding the application of the lower casing throughout the entirety provides assistance throughout the procedure of normalization and should be regarded as the most important procedure in attempts towards keeping up with the words within a distinct containing a largely uniform manner. The programmed machine does not have

any concept of language; therefore, it treats capital words and lower-case words differently. Since social media text is very disorganized and non-uniform in nature, all the different words will be treated differently, thereby increasing the size of the process text, which will then, in turn, decrease the efficiency of the dataset. Lower kissing brings back the entire data into a normalized form where each and every character and word is lowercase so that two different words that existed in different alphabetical cases can correspond to the same original word.

3.4.2 Special characters

Special characters are those characters that could not be categorized as being alphabetically as well as a numeric value like that of the hashtags as well as special symbols. Operating upon the numerical data is an easy task for the machines to carry out as well as the machines could indeed be capable of operating upon the textual data, but the difficulty arises when the special characters are introduced, which could prove to be sometimes quite tricky. Due to this aspect, throughout the procedure of tokenization, the special characters are indeed developed within their own tokens and are disregarded due to the aspect that they are not at all helpful towards describing any sentiments throughout any algorithm, likewise, the numbers. They only serve the purpose of increasing the size of the data set, which makes it less efficient and takes a lot more time to process. The special character also includes punctuation marks such as full stops (.) and commas (,). These special symbols, such as exclamatory marks (!), are also deemed useless for this particular research since they have no purpose in extracting the sentiment of the text and further.

3.4.3 Stop word removal

The stop words are the most widely utilized characters of the worlds within the natural English language. But still, 30 respective words provide no value regarding the predictive power when

talking in real-time. Due to their syntactical nature of them, the stop words do not serve any purpose or do not provide any assistance to us in carrying out efficient sentiment analysis. Stop words are filled words that are present in a sentence only for chromatic as well as syntactical purposes but contribute very little to know the meaning in the actual underlying meaning of the text. The English language is pull-up stop words, and some of the examples are, my, to, the, while can because, by, of, etc., and many more. These words only increase the size during the process of vectorization later in the sentiment analyzer. Each vertex up space in the coppers, and since all in the sentence constitute multiple of these words, if a frequency chart is to be drawn, these words will shine on top certainly. In order to determine the sentiment as well as the intensity, only the keywords within a sentence are required, and the process of step word removal ensures that only the keyword in a sentence remains. A fine example of this scenario is the voice command of setting the alarm when conditions are needed.

‘Set an alarm for 5:15 in the morning.’

The sentence above is a simple command to set the alarm at 5:15 in the morning, but the way it is processed by the algorithm is different. The stop words are removed, and the remaining sentence becomes,

‘set alarm 5:15 morning.’

Notice how the underlying meaning of the sentence is still intact, but the sentence is much more efficient for a model or algorithm to process since it contains fewer amounts of words than before.

3.4.4 Word Stemming and Lemmatization

This is the most crucial aspect to be utilized in the pre-processing of the text due to its significant use with respect to the field of text mining as well as its assistants towards the updating of a

significant amount of relevant information due to the aspect that introduces a sophisticated amount of reduction throughout each and every word which comprises of the similar roots to that of a common fine through implementing the efficient removal of the various of suffixes (Mehta et al., 2020).

Stemming inherently focuses more on the production part rather than retaining a word in its full grammatical spelling. This means that in a stemming algorithm, the word is reduced down to the basic form that each and every word in the entire corpus follows. Stepping is usually much quicker and more efficient than its counterpart since it has easy access it has many variations to choose from depending upon the type of task at hand; however, the only task that these transformations are forming is removing the suffixes at the end of the word so that to give more abstract meaning that is still acceptable. The following table shows the stem operation performed on each word and the result acquired.

Word	Stem
Closed	Close
Sleeping	Sleep
Easily	Easy

Table 2: Stemming

Lemmatization, however, utilizes an existing knowledge base or a Dictionary of words in order to transform and stem the required words back to their root form. This overcomes the disadvantage of a stemmer in cases where a generic English word is not retained at times. Lemmatization reduces the word back to its base form and converts the words from any degree to their normal one. The

only disadvantage of lemmatizer is that since it uses a dictionary, the execution is really slow, and often times it is not used in pipelines of data mining for this exact reason. It is also hard to create a lemmatizer for other languages because of the same reason, but the results are much better under often the deciding point between the 2. The following table shows the output of a lemmatizer on certain words.

Word	Lemmatize
Higher	High
Best	Good
Fastest	Fast

Table 3: Lemmatization

3.5 Data Visualizations

3.5.1 Matplotlib

The visualization of data could be explained like that of a crucial aspect throughout the various activities conducted in a business environment due to the factor that the organization in today's modern era is moving forward with the approach of collecting a significant amount of data to be processed according to their needs. In addition to this sophisticated representation, visually indeed assess the readers towards easily comprehending the entire place of any complex issue or problem, with respect to the efficient identification of various patterns as well as the outliers in data along with the relationships that are shared among the various entities within it. The inside, which is

gathered from reviewing any visual representation of data, provides assistance with respect to the development of various strategies, which could be of significant assistance to the businesses regarding experience in increment within the overall growth.

Plotting of the visually represented data through the utilization of matplotlib has been observed to be of significant ease. What usually happens is that when the plot in all the data is being carried out, the user has to follow certain steps throughout each and every plot. The matplotlib consists of a module by the name of pyplot, which assists the users towards the plotting of various figures in an efficient manner. The Jupyter notebook was utilized towards the running of the various types of plots by importing the library of pyplot.

3.5.2 Plotting of the categorical Data

The statistical approach could be utilized towards the denotation of a procedure regarding the comprehension of the various types of relationships among the multiple types of variables present throughout a provided dataset along with the insights towards how that relationship is responsible towards affecting as well as depending upon the other variables.

The Seaborn is Widely and popularly utilized towards the presenting of various visualizations, like:

- Scatter plot
- SNS.relplot
- Hue plot

3.5.3 Scatter plot by Seaborn

The scatterplot could be defined as that of the significantly common example utilized for the presenting of effective visualizations regarding the relationships among the two variables (Kolasani et al., 2020). Each and every point displayed within the respective observations throughout the datasets, as well as the observations, are indeed presented by the specific format of visuals comprising of the representing the data as dot-like structures.

```
sns.relplot(x="Views", y="Upvotes", data = df)
```

3.5.4 The SNS.relplot by seaborn

SNS.relplot is the function from within the SNS class within the Seaborn, which is utilized when imported above that of the various other dependencies. The various parameters like -x,y, and the data which is operated upon, represent the relationship among the variables upon the X-axis as well as the Y-axis along with the aspect that the data which is being utilized for the developing of the visuals.

```
sns.relplot(x="Views", y="Upvotes", hue = "Tag", data = df)
```

3.5.5 Hue Plot

Seaborn also provides the users with the aspect towards the inclusion of adding another dimension within the already developed plot with the assistance of the Hue. This aspect is carried out so as to implement a color throughout the various points along with providing meaning to the utilization of each of the colors, which indeed have some hidden meaning behind it.

```
sns.replot(x="Views", y="Upvotes", hue = "Answers", data = df);
```

3.6 Limitations

There are several kinds of associated factors in the world, and here in this, there is a need to study the different kinds of roaches so that the various kinds of stations can be observed. The limitation of the work done into this project consisted of requiring the different forms of analysis and also the major limitation con assisting of the statement that the sentiments and the other social media platforms sentiments and volumes of the searchers and trends are not a sufficient thing to study the fluctuations into the data. In the prediction of the Bitcoin prices and other cryptocurrency prices is the need to study different kinds of the aspects and things that are being hard and helpful in acquiring the ups and downs of the Bitcoin currency. In the making up of the use cases and finding of the proper productions only the sentiment analysis of the social media is not they can be the criteria and hence their needs to the different kinds of the specs to be stated in such a way that the utilization of the effective opinion and winding up of the predictions can be done. Into the making up of the resources here, there is the need of using a physical quantity inspection by means of these the research has the limitation that there cannot be the relation among the sentiment analysis of the Twitter data and into the situation of the Bitcoin prices as there is another control the factors in fault. Also, the relation does not have authenticity, such as the only criteria of analyzing the sentiments of the tutor, and by the use of the Twitter data, the proper estimations cannot be taken in.

3.7 Summary

The chapter had a detailed discussion of the methodologies for data extraction and filtering the data once it has been collected. The methods involved in filtering the data were explained in detail along with the ways used to plot the data in a graphical format. The key intake was how the data

was collected using Twarc and then further cleaned for further preprocessing so that to make sure that proper data was forwarded ahead to further analyze. The data that was collected from the twitter stating the total number of verified and unverified users for the given time period can be summarized in the below table.

Time Period	Verified	Unverified
2020/02/01-2020/02/28	12558	1015215
2020/03/01-2020/03/31	16621	953071

Table 4: Count of Verified and Unverified tweets made

CHAPTER 4

Data extraction and Preprocessing

4.1 Collecting the data

Cryptocurrency has been the most trending subject in terms of investments as well as the various subjects regarding the financial status of the social communities. It is a digital form of decentralized currency that is controlled entirely by the people who own it and not some other third independent party that expects to control the monetary value of the currency, expecting a favor or decrease according to will. The cryptocurrency market is still at large and is very popular among the various communities that are present online, and together they all get together to engage in the buying as well as selling of a currency of their favor. There are a lot of different types of currency involved, whereas many new ones are being created every now and then. Entire blogs, as well as forums are devoted to providing information about these distinct types of currencies and what kinds of factors affect them and their price. One such factor that affects the price of cryptocurrency is social media and its users. Bitcoin is a very popular and widely known cryptocurrency that made it so big that a single coin is worth around 55,000$. This amount of worth for a single coin has made a lot of people rich, and a lot of people analyze its worth and dive deep into the research as to what affects the price of the crypto coin the most. This leads to a lot of research as well as studies around the field of crypto coins and their influences as well as the factors that influence them.

Twitter is a prominent social media website as well as a giant hub for very detailed discussions of various kinds of communities of people; it is a great choice for research based on a specific entity such as Bitcoin. People tweet a lot of their responses and views as well as opinions regarding the stage or the price of bitcoin and as well as anything that is upcoming that might affect or influence the price of the bitcoin by a considerable amount. This is the reason that tweets are selected for the purpose of this research, and conclusions have been derived based on the same data. It is uncertain as to what drives the flow and which direction is the cause-and-effect relation pointing. Twitter data contains a lot of hashtags as well as keywords that can easily be utilized to determine the identity of the sentence. In order to arrive at the semantic details of the sentence that can be used to determine the impact of the sentence, a bunch of modules will be used in python so that the sentiment can be derived.

4.2 Twitter Academic research API and Twarc

In this research, the tweets will be extracted using Twitter, which is useful for fetching and extracting tweets from the website at random. It can also be filtered so that specific tweets can be extracted that certain match criteria for targeted and specific research. This API can be used to fetch tweets based on a lot of filtering as well as for a long time period. There are a lot of modules in python that can manage to extract tweets using this API, but many suffer from certain drawbacks as well as limitations that do not let them perform to the fullest extent. Many of these disadvantages include not being able to derive tweets past a single week from the date of extraction. Another drawback includes not being able to clearly filter out re-tweets from normal tweets, etc.

These problems are solved in a single module called Twarc and research API. This module supports a lot of functionality as well as beings forward solutions to all the drawbacks by providing

a better token and key for the scrapping so that the data can be filtered exactly according to the needs. Usually, for Twitter data scrapping, a bunch of tokens and encrypted keys are needed to authenticate the API and make a request to the server for the tweets. This is the usual process that includes access tokens and well as customer keys in order to fetch the tweets. The Customer keys are of two types, and both help identifies the user and helps the Twitter server know who is making the request and which account it is attached with. For fetching the tweets, the keys are derived from a Twitter developer account that enables us to get the access tokens as well. The access tokens are equally important since they assist in informing the server of the reason for this request to fetch tweets. The reason can be a lot of things for the server, but the most used one is for research and study purposes.

4.3 Data Preparation

The first part comprises of obtaining the data set along with loading it so as to read the various contents comprised within the data sets regarding the prices of Bitcoins. Followed by this, a detailed discussion, as well as insights, are provided regarding how to implement data processing, along with a detailed explanation of why this particular procedure needs to be carried out.

Using these modules, the tweets were extracted, and the filters were provided such that a specific timeline was involved. Moreover, the tweets were filtered so that only those tweets that contain the tags or word ‘bitcoin’ are extracted. So, for the research, different datasets were extracted from the internet.

- **Bitcoin Prices dataset**: The bitcoin prices on the dates that match the dates of the tweets in both cases.
- **Tweets time period**: Feb 2020 → Apr 2020.

```
from google.colab import drive # loads a library to mount your google drive
drive.mount('/content/drive', force_remount=True)
```

```
Mounted at /content/drive
```

```
!ls "/content/drive/My Drive/" # shows all files in your google drive root, including the project data file Tr0
path = "/content/drive/My Drive/" # sets the path to the root with the file Tr0
```

```
# all_tweets = pd.read_csv('/content/drive/My Drive/Feb_mar_verified.csv')

nltk.download("vader_lexicon")
nltk.download("stopwords")
stemmer = SnowballStemmer('english')
vectorizer = TfidfVectorizer(use_idf = True, tokenizer = nltk.word_tokenize,stop_words='english', smooth_idf = True)
remove_punctuation_map = dict((ord(char), None) for char in string.punctuation)
sid = SentimentIntensityAnalyzer()
stopw = set(stopwords.words('english'))
```

```
[nltk_data] Downloading package vader_lexicon to /root/nltk_data...
[nltk_data] Downloading package stopwords to /root/nltk_data...
```

Figure 8: Bitcoin price dataset

4.4 Processing and cleaning of data

Tweets were extracted using a manual loop that extracted all the tweets for each day between the time periods for both the tweets dataset specified.

Out[ ]:

	id	created_at	text	author_id	author.verified
0	1244776383108043904	2020-03-30T23:59:51.000Z	$BTC SELLING PRESSURE ALERT Price tradi...	915882541864775680	False
1	1244776267112538112	2020-03-30T23:59:23.000Z	中国ビットコインのロゴ入りTシャツ販売中！ #Bitcoin #ビットコイン #仮想通貨 #...	3167286050	False
2	1244776243679068160	2020-03-30T23:59:17.000Z	#Bitcoin: BTCUSD ($BTCUSD) DOWN 5.7521%! Last ...	2332763818	False
3	1244776224767053827	2020-03-30T23:59:13.000Z	#Bitcoin: BTCUSD ($BTCUSD) DOWN 1.9116%! Last ...	2332763818	False
4	1244776217376485376	2020-03-30T23:59:11.000Z	relong or add to longs 6250--6300 $BTC #BTC h...	836500153737474050	False

Figure 9: Script for scrapping the data

After extracting the tweets, the task at hand was to observe and analyze them so that the sentiment analysis process could be executed. The Twitter accounts which are focused randomly for the data collection includes verified user as well as the non-verified user. Most of the data are collected from verified users. The tweets and other related data need to be structured accordingly such that they can be processed inside a data frame. Other than tweets, a few other details were also extracted, such as the date, for time-based analysis, as well as the tweet ID and the user's name.

Due to ethical concerns, the usernames were dropped from the dataset after extraction so that the data could remain anonymous. Moreover, the research and study conducted in this project are totally random and not carried out on a specific group or a community in order to generate bias. The tweet ID was extracted so that duplicate tweets can be dropped easily, even if they are tweeted on different days.

```
In [18]: preprocessed = []
         for i in all_tweets["text"]:
             preprocessed.append(clean(i.strip().lower()))

         all_tweets.insert(3, "processed_tweet", preprocessed, True)
         # all_tweets

         all_tweets['processed_tweet'] = all_tweets['processed_tweet'].apply(lambda row: str(row))


         sentiment = {"Positive":[], "Negative" :[], "Neutral":[], "Compound":[]}
         for i in all_tweets["processed_tweet"]:
             sentiment["Positive"].append(sid.polarity_scores(i)['pos'])
             sentiment["Negative"].append(sid.polarity_scores(i)['neg'])
             sentiment["Neutral"].append(sid.polarity_scores(i)['neu'])
             sentiment["Compound"].append(sid.polarity_scores(i)['compound'])

In [19]: all_tweets.shape

Out[19]: (2008110, 6)
```

Figure 10: Twitter data after extraction presented in a data frame

Only the date and the tweets were kept after the details were derived and necessary functions were performed. The next step was to clean the textual tweets so that they could be easily processed effectively as well as efficiently while also taking up less space and ending up cleaner for the analysis. Preprocessing of data is based on verified and non-verified users, as most of the data is based on verified accounts of Twitter. Processing and cleaning the tweets get rid of the unnecessary textual features that are present only to serve a syntactical structure to the sentences and phrases. After cleaning the tweets only, the keywords that are important are left, and they can be easily utilized to ascertain the sentiments and polarity of the tweets. The date column was also converted to a date-time data type such that it can be ordered and not result in ambiguous results.

Usually, if the date is not in the correct format, it is treated as a string that can be a very big hindrance in the process since the date will result in a very ambiguous result.

```
In [21]: all_tweets['created_at'] = pd.to_datetime(all_tweets['created_at'], format='%Y-%m-%d %H:%M:%S')

         all_tweets['month'] = all_tweets.created_at.dt.month
         all_tweets['day'] = all_tweets.created_at.dt.day
         all_tweets['hour'] = all_tweets.created_at.dt.hour
         all_tweets['year'] = all_tweets.created_at.dt.year
         all_tweets['weekday'] = all_tweets.created_at.dt.weekday


         for col in all_tweets.select_dtypes(['datetimetz']).columns:
             all_tweets[col] = all_tweets[col].dt.tz_convert(None)


         all_tweets.insert(11, "Neutral",  sentiment['Neutral'],  True)
         all_tweets.insert(12, "Positive", sentiment['Positive'], True)
         all_tweets.insert(13, "Negative", sentiment['Negative'], True)
         all_tweets.insert(14, "Compound", sentiment['Compound'], True)

In [23]:

In [24]: # all_tweets.to_csv('/content/drive/My Drive/all_tweets.csv', index = False)

         # saving the preprocessed dataset so that it can be accessed later

In [25]: all_tweets = pd.read_csv('/content/drive/My Drive/all_tweets.csv')

In [31]: all_tweets.head()
```

Figure 11: Performing necessary functions before processing the tweets

The cleaning of these tweets required a lot of specific procedures as well as performing them in a strict and necessary order. This is done to ensure that all the text, no matter how differently presented, in the end, boils down or reduces to the same pattern as well as syntax. The order is necessary so that the reduction that is performed can be applied to the full dataset equally and in a definitive manner.

It includes converting the text into the lower case at all places so that all the words and phrases are the same. In the scripticase-sensitive text is a feature that prevents different cases of the same word from representing the same thing. Capital words and smaller case words are treated differently, such that ‘Hello’, HELLO, and hello are three entirely different words. Reducing them to the lower case can result in a huge improvement in efficiency so that the dataset can be cleaner.

The next step was to detect patterns in the text so that they could be removed from the tweets. A tweet consists of a lot of the same phrasings and symbols such as a '@' in order to tag people and usernames. These must be removed since they serve little to no purpose in this research. The next pattern and symbol were '#'. These are used to indicate a topic or a keyword that explains what topic or entity the tweet is concerning or is relating to. Hashtags often contain redundant information in tweets, and it is advised as well as efficient to drop them and clear the space so that the text is short and to the point. Sometimes the hashtag symbol (#) is removed, and the word preceding it is kept so that information is not lost at a level.

Similarly, website links that are present in text are removed, and it is done by utilizing the same pattern of detecting them first and deleting them. All of the detection and replacement is done by the regular expression module in python. It detects the links by capturing words that start with the keywords like 'https:/'. Usually, the links start with this notation for the websites and are removed using the substitute method of the module.

```
In [ ]: all_tweets['processed_tweet'] = all_tweets['processed_tweet'].apply(lambda row: str(row))

In [28]: dates =list()

         all_tweets['created_at'] = pd.to_datetime(all_tweets['created_at'])

         for i in all_tweets['created_at']:
             dates.append(i.date())
         all_tweets['date'] = dates

         all_tweets.drop('created_at',1, inplace = True)

In [32]: verified, not_verified = all_tweets[all_tweets['author.verified']==True], all_tweets[all_tweets['author.verified']==False]

         x, y = verified.groupby('date')['processed_tweet'].count(), not_verified.groupby('date')['processed_tweet'].count()

In [33]: x, y = x[x.index>date(2020,2,1)], y[y.index>date(2020,2,1)]
         x,y = x[x.index>date(2020,2,1)], y[y.index>date(2020,2,1)]

In [34]: plt.suptitle('Daily Influential vs Non Influential Tweets', fontsize=25)
         plt.plot(x.index, x.values, label ="verified", marker="x")#.grid(axis='y')
         plt.plot(y.index, y.values, label = "not_verified", marker = 'x')#.grid(axis='y')
         axes = plt.gca()
         axes.yaxis.grid()
         plt.legend()
         plt.xticks(rotation=45)
         plt.figure(figsize=(15,10))
         plt.show()
```

Figure 12: Function to clean the text and remove unnecessary words

The price data of bitcoin was also adjusted so that it is much closer to a similar distribution and scale of the polarities that can be easily visualized. The bitcoin prices were normalized through max value normalization. The entire array of prices was divided by the maximum value in the list.

4.5 Sentiment Analysis with VADER

The incorporation of the VADER module is utilized towards implementing the uses of the sentiment analysis, and the Natural Language toolkit, along with the input to be provided to be operated on, which comprises the entire single sentence (a single tweet). The outputs received from the completion of the procedures for the VADER are observed to be related to the polarity and the subjectivity of the input. The score for the polarity lies among the range of -1, 1, where -1 is used for the identification of the negative words along with one is utilized towards the identification of words that are in the positive aspect (Hasan et al., 2018). The reason VADER is specifically used for this research is that it has the edge over other modules since it is built specially to analyze text related to social media. It is familiar with emoticons as well as social media slang that make it an optimal choice for such research.

The procedure in which the VADER was utilized throughout the respective report towards the efficient conducting of Sentiment Analysis is:

At the very initial stage, the module has to be installed in Python along with the configured pip package. The code to be written for this is

- pip install nltk (if using the NLTK variant),

-pip install VADER (if using the standalone variant).

Followed by this, VADER has to be imported by the initializing of the code as- from nltk. sentiment.vader import SentimentIntensityAnalyzer

The Syntaxes utilized towards obtaining the results as the polarity score are:

res = SentimentIntensityAnalyzer ()

print(res.polarity_scores(*sentence*)

Due to the factor that the VADER is a Lexicon based analyzer utilized for the analysis of the sentiments from within the text, it comprises of some pre-defined rules that are to be followed regarding the said words as well as the weight dictionary, throughout which multiple random scores are present which are capable of providing assistance towards the calculation of a sentence's aspect of polarity.

Applying a normalization to find a sentiment score between -1 to 1 the normalization used is

Where x is the sum of the sentiment scores of the constituent words of the sentence and alpha is a normalization parameter that we set to 15.

For example:

VADER sentiment analysis takes these into account by considering five simple heuristics.

1. Punctuation is the first heuristic. "I like it." and "I like it!!!" are two different ways of saying "I like it." It's not difficult to dispute that the second statement elicits more intense feeling than the first, implying that the second sentence must have a higher VADER sentiment score. If the score is positive, VADER multiplies each exclamation point and question mark by a specified experimentally determined value. VADER subtracts if the score is negative.
2. Capitalization is the second heuristic. The phrase "AMAZING performance." is more powerful than "amazing performance." As a result, VADER accounts for this by increasing

or decreasing the word's sentiment score, depending on whether the word is positive or negative.

3. The employment of degree modifiers is the third heuristic. "Effing cute" and "kind of cute," for example. In the first line, the modifier increases the intensity of cute, whereas in the second sentence, it decreases the intensity. VADER keeps track of a booster vocabulary that includes a variety of boosters and dampeners. Depending on whether the main word is positive or negative, one modifier beside it adds or subtracts from the sentence's mood score. A second modifier adds/subtracts 95 percent, and a third modifier adds/subtracts 90 percent.
4. The shift in polarity caused by "but" is the fourth heuristic. "But" frequently connects two phrases with opposing sentiments. The latter, on the other hand, is the most prevalent sentiment. "I love you, but I don't want to be with you anymore," for example. The first phrase, "I love you," is a positive statement, but the second, "I don't want to be with you anymore," is a negative statement with a stronger emotional impact. A "but" checker is implemented by VADER. Before the "but," all sentiment-bearing words have their valence lowered to 50% of their original value, while those after the "but" have their valence increased to 150 percent of their original value.
5. To detect polarity negation, the fifth heuristic checks the trigram before a sentiment-laden lexical characteristic. A trigram is a set of three lexical features in this context. VADER keeps track of negator terms. Negation is represented by increasing the emotion score of the sentiment-laden lexical feature by a value that has been empirically calculated.

The below illustration shows the rules for labelling each user tweet into a sentiment category based on the threshold.

1. Positive Sentiment: compound score>=0.1
2. Negative Sentiment: compound score <= -0.1
3. Neutral Sentiment: -0.1 < compound score < 0.1

4.6 Summary

The chapter is focused on data extracting done and preprocessing steps done in order for the data for further processing. It also briefly describes VADER and how it works including the parameters included for classifying the type of sentiment it was give as an output. The succeeding chapter of this thesis will include the is focused on data analysis and results with a brief discussion on what the outcomes are.

CHAPTER 5

Results and Discussion

This chapter is focused on discussing the results that were obtained through the analysis performed. The first part is a discussion and comparison of the verified and unverified users' tweets and their compound sentiment obtained in the chosen timeframe of the research. The second part shows further evidence on what the authors sentiments and tweets count have effect with the price of Bitcoin. The last part is an overall discussion of the results obtained.

5.1 Verified vs. Unverified users' tweets

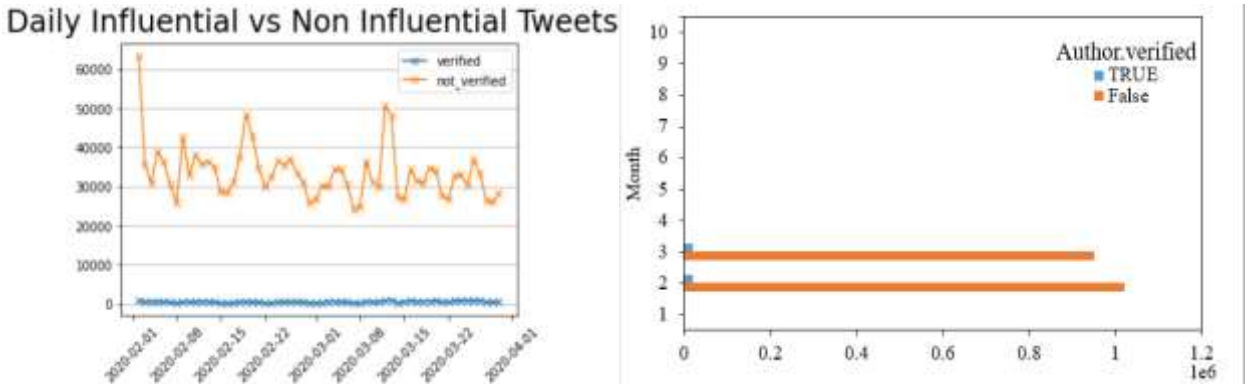

Figure 13: Number of verified tweets and unverified tweets

This graph shows start the data contains a feature that describes whether to it was made from a verified account or a non-verified account. This is important since Twitter and all of the social media is prevent bots as well as spammers that constantly tweet meaningless things that are not important and just add to the noise in the text data. Of course, some of the verified accounts are also real people who contribute to the cause and discussion by their tweets. This graph shows that

the number of tweets on each day fluctuates a lot for the unverified accounts, whereas to verify, people do not have as much frequency daily on tweets. The amount of data which are collected from the verified users is required less verification. However, here it is seen a spike for both categories of accounts. Additionally, it shows that March saw more verified tweets than February.

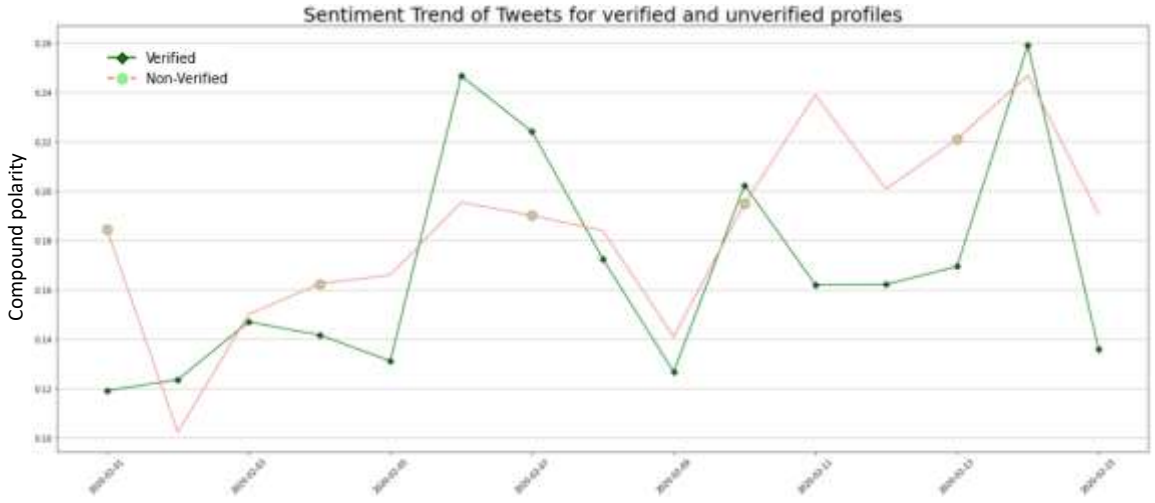

Figure 14: Polarity of verified and unverified accounts between 1st Feb and 15th Feb 2020

This line chart shows the tweet average polarity value grouped according to the verified and unverified accounts. The green line shows the compound polarity on the dates on X-axis for all the verified accounts, whereas the red line indicates the compound polarity for all the unverified accounts and the corresponding X-axis dates.

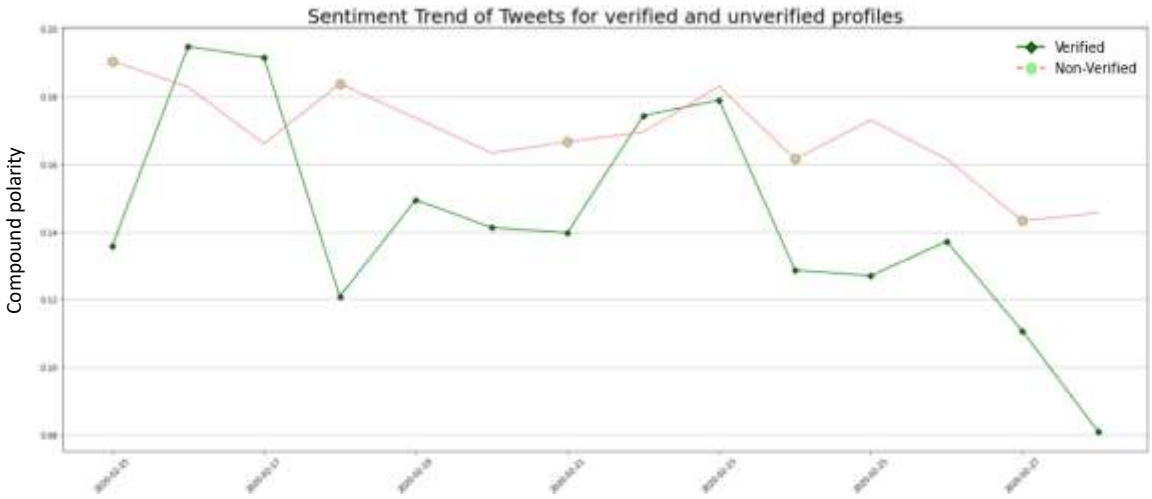

Figure 15: Polarity of verified and unverified accounts between 15th Feb and 28th Feb 2020

This graph above depicts the same relative information of the compound polarity of verified and unverified tweets between the time period of 15th February and 28th February 2020. These ground effects one important key factor that is despite the number of tweets for an unverified account being higher than the rest, the polarity of the verified tweet is still more and fluctuating than the rest (Hasan et.al., 2018).

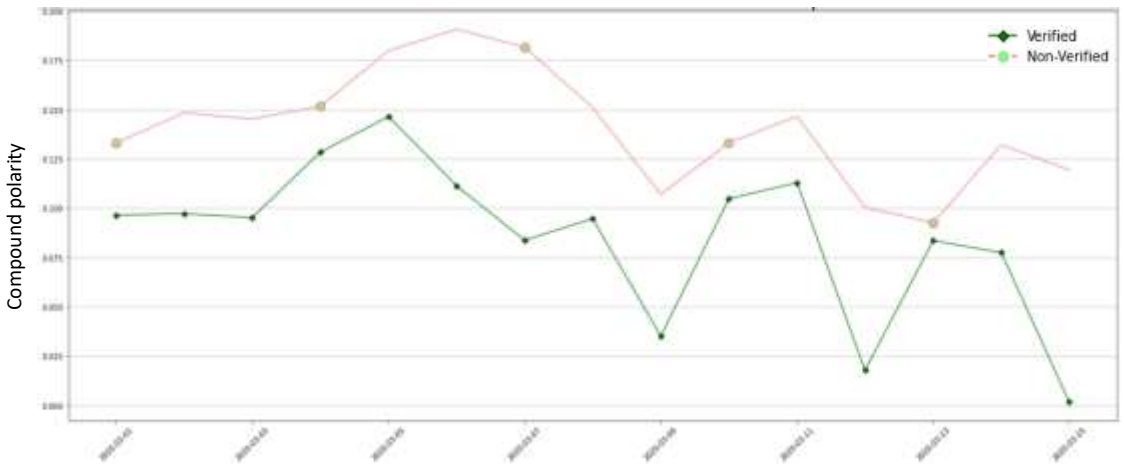

Figure 16: Polarity of verified and unverified accounts between 1st March and 15th March 2020

This graph again expands the time period in order to take more data into account as the average polarity. This shows that when a bigger time window is taken into account, the average polarity of

the unverified tweets turns out to be more than the polarity of the verified accounts. Another key factor to notice here is that they both follow a relatively similar pattern of increase and decrease in the polarities, which could be due to time-based events carried out on those dates. These events are hypothesis to be news related to the the good news about the currency Bitcoin could indicate positive tweets, and resultant unstable price-related news could point towards the fluctuation of popularity.

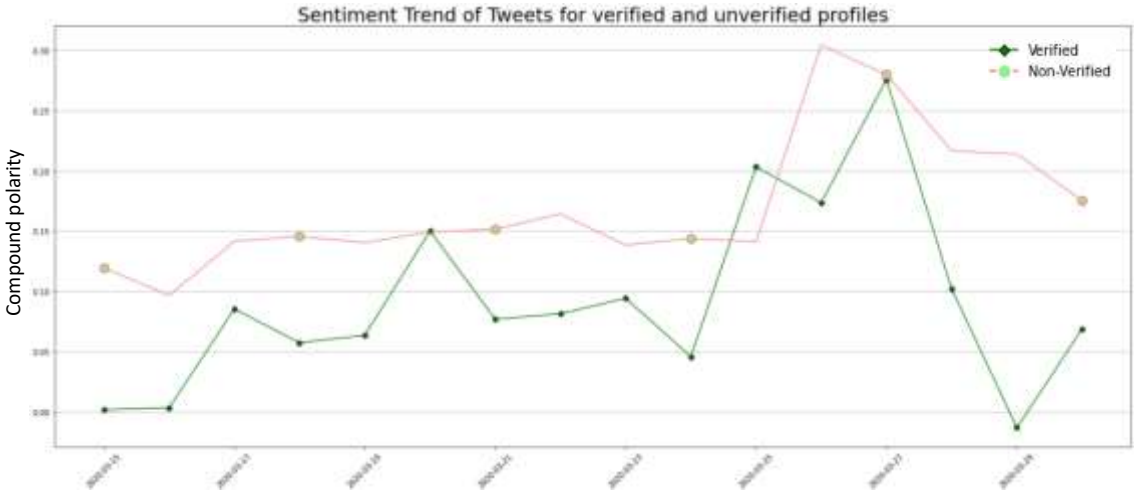

Figure 17: Polarity of verified and unverified accounts between 15th March and 30th March 2020

This graph shows the time period between 15th March to 30th March 2020. The end of March, around 25th, saw an increase in positive tweets for both verified and unverified accounts

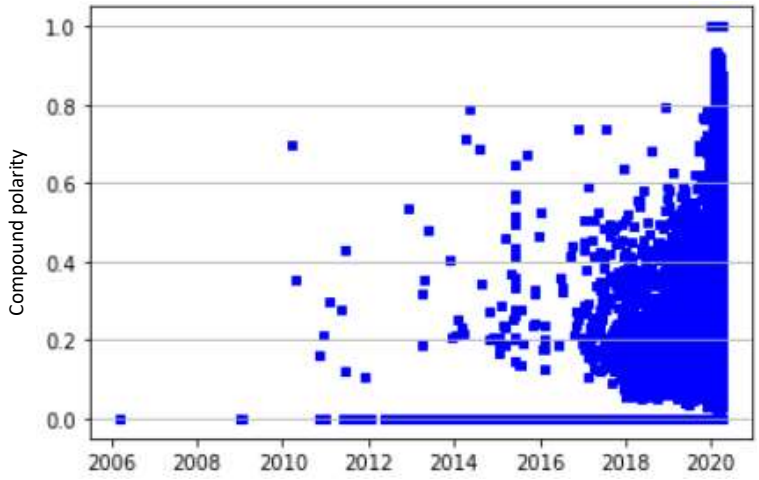

Figure 18: Polarity of the tweets mapped on a timeline

In this scatter plot, all the unique positive tweets in the data set are mapped on a timeline of 14 years. This scatter plot depicts information of the pattern on how the positive polarities of the tweets have been increasing as time passes. It shows a massive increase from the year 2017 to the year 2020, where the year 2020 has the highest polarity of positive tweets.

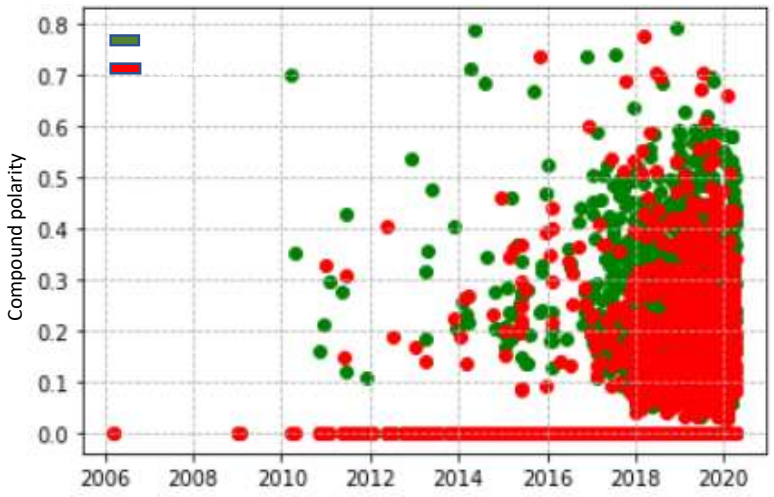

Figure 19: Polarity of the positive and negative tweets mapped on a timeline

For contrast purposes, a comparative scatter plot was also created in order to assess the difference between the polarities over the years between positive and negative tweets. For this purpose, all the unique positive and the unique negative tweet polarities were utilized to create the scatter plot

and assist the difference between them. It shows that most of the tweets are positive in nature about Bitcoin, whereas negative tweets were not far behind. Both the positive and negative tweets about cryptocurrency follow a similar trend of gaining a disruptive increase around the year 2020 (Shen, Urquhart, and Wang, 2019).

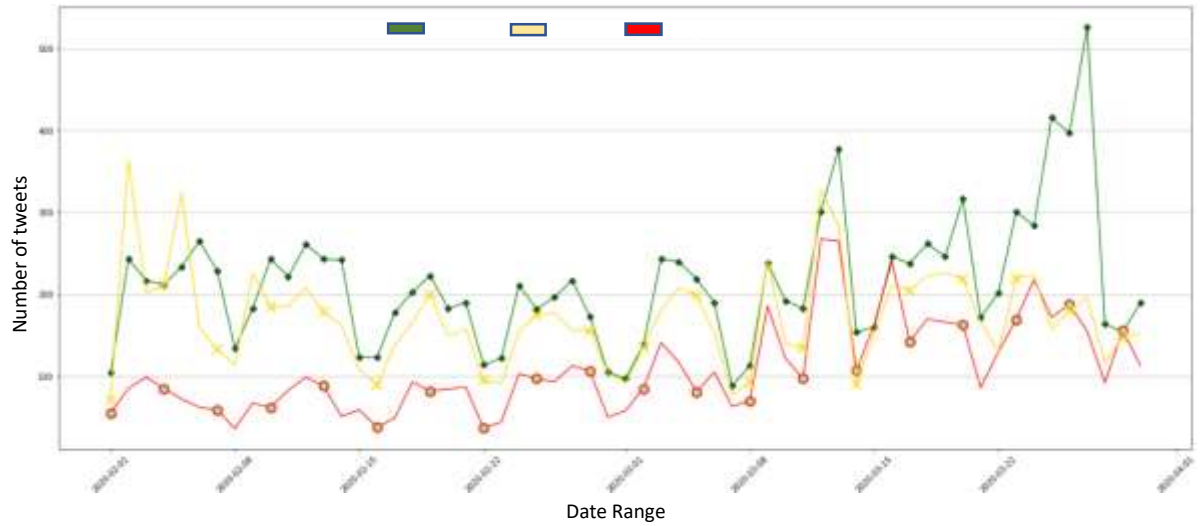

Figure 20: Number of positive, negative, neutral tweets by verified accounts on a specific timeline

This line chart represents the number of positive, negative, and neutral tweets that are made by verified accounts between the times 1st Feb 2020 to 1st April 2020. This is a very informative representation of the information since it shows that despite being totally different in nature, the number of tweets follows a distinct pattern altogether. All three of the categories of two each suffered a decline on 15th of February and can be seen have shown a similar response on 22nd Feb as well as 1st March. They also show a similar rise in Indian numbers at a constant rate.

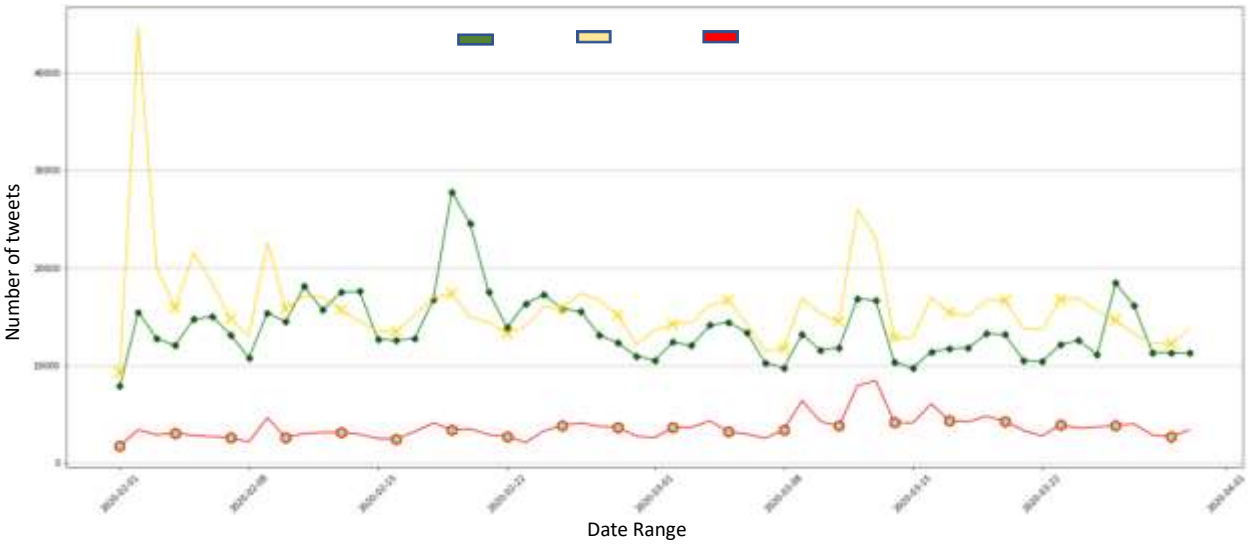

Figure 21: Number of positive, negative, neutral tweets by unverified accounts on a specific timeline

As mentioned earlier, unverified accounts are not as reliable as a source of information as to their counterpart of verified accounts. The above figure displays similar information of the number of positive, negative, and neutral tweets but only for the unverified accounts. As seen here, the neutral polarity seems to surpass both the positive and negative categories and stand out on top, indicating that most of these tweets are generic and not meaningful in nature. As stated above that, mostly these tweets consist of Twitter bots and advertisements and promotional offers about Cryptocurrencies such as Bitcoin in this case. Moreover, the unverified graph does not seem to follow a distinct pattern like the verified accounts indicating fluctuations all over (Ajmi, Youssef, and Mokni, 2022).

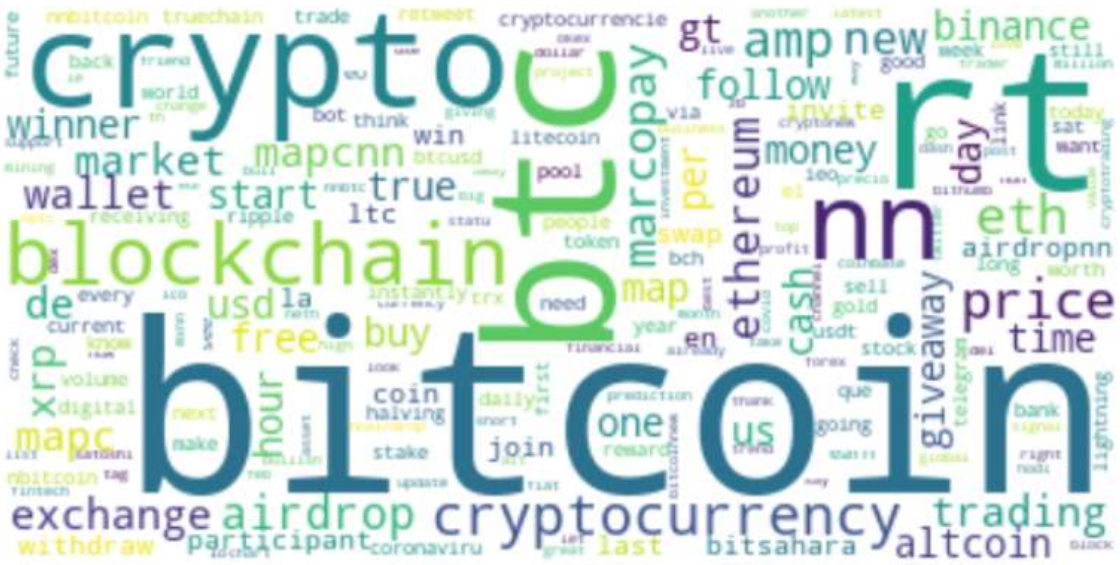

Figure 22: Word cloud formed by the tweets of unverified accounts

The above word cloud is formed by combining the tweets and words off unverified accounts in the data set. At first glance, the words are relative to the cryptocurrency but are very generic in nature, with few words like AirDrop, free, buy, macro play, giveaway, etc., indicating association with certain ads that hover around social media.

Figure 23: Word cloud formed by the tweets of verified accounts

The same word cloud can be seen by the verified accounts, and a clear difference is notable here that the words used are more relative to the pricing of Bitcoin and reliable Ness of the text. Certain words like investor, coronavirus, bank, future, money, assets, market, etc., do have higher precedence here than the word cloud created previously by unverified accounts. This shows the difference between how a verified account can have all the differences in providing information that is reliable and relatable to the cryptocurrency (Shevlin, 2021).

5.2 Authors and their effects on bitcoin price

In the second part of the entire procedures to be carried out, this could be regarded as a crucial aspect towards verifying as well as building the statistics as the various verified authors are capable of influencing the judgment of common people according to what they say or tweet regarding the Bitcoins as well as the cryptocurrency. Therefore, this particular factor is also included towards verifying as well as identifying how the tweets, as well as wordings of various authors, have impacted the pricing of Bitcoin and the cryptocurrency market. The particular section efficiently describes the detailed correlation to be identified among the number of various tweets presented by verified auth up on a daily basis regarding their positive as well as negative effects. To understand these particular positive, as well as negative in detail, a comprehensive description as well as a graphical representation has also been provided towards Pro presenting the readers with the negative linear connections as well as the positive linear connections among the authors treats as well as their respective fall and rise identified in that particular time period upon the Bitcoin.

As stated throughout the research, the type of author matters significantly if it is to influence the price of Bitcoin. Verified author unknown to influence a lot of people into buying or selling the coin it is evident through the visualizations as well as through the research. The question remains

of the quantity that how many authors will need to express a dear particular opinion that can cause the price of Bitcoin to change drastically and evidently. A few more questions arise, like how many tweets for a particular author does it take for the same change to appear. Regarding this hypothesis, analysis was focused on the number of authors and their tweets compared against the prices as well as polarities. The correlation that was discovered suggested that when the number of verified authors tweet relating to Bitcoin and its factors increase, the price of Bitcoin decreases. Since this correlation is negative in nature, when the number of tweets from the verified author about Bitcoin decreases, the price of Bitcoin increases.

A few more visualizations were generated in order to support this hypothesis, including twin graphs as well as correlation heat maps.

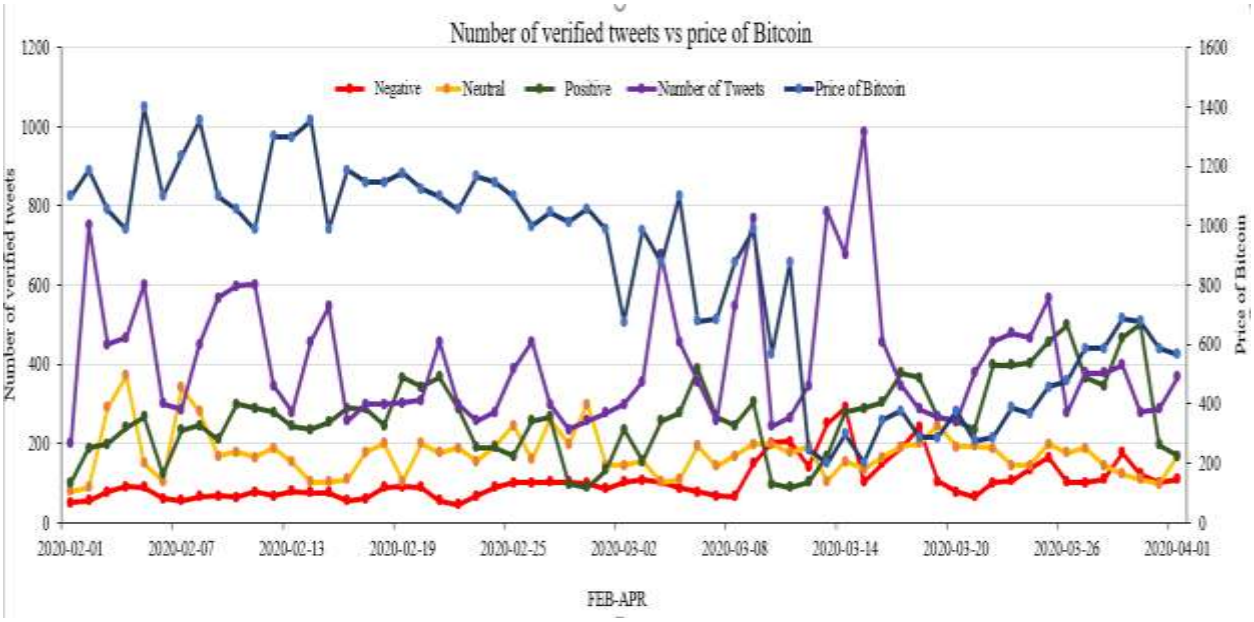

Figure 24: count of the sentiment of verified tweets per day against the price of bitcoin

The graph shows the price of Bitcoin compared with the number of tweets posted by verified authors on a particular date. It is a little difficult to derive any inference from the graph since of the fluctuating nature, and to make any decision, one has to look very closely. This is the reason

why he was selected, but this particular analysis captures this entire trend into a single positive and negative value. Even though it is very fluctuating, it can be analyzed that when the number of authors increases, the price decreases.

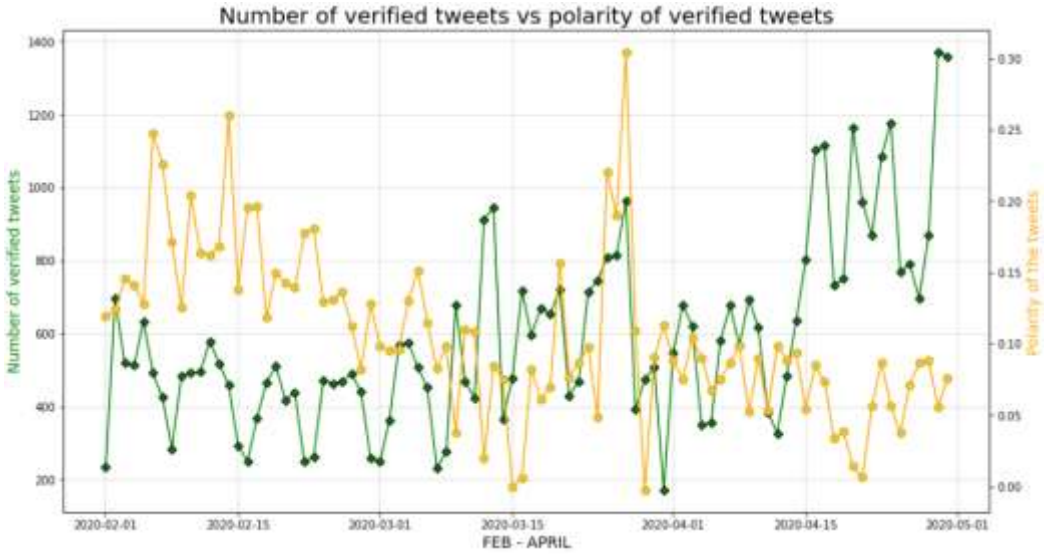

Figure 25: count of verified tweets per day against the polarity of tweets

The Graph shows same analysis result of performed with a number of tweets posted by verified authors compared against the polarity of the same tweets. From this graph, it is seen that positive relationships are formed between the two entities. It can be concluded that more people tend to appraise the factors relating to Bitcoin, and a smaller number of people tend to criticize if there is any negative comment.

An index of relationship, known as a co-efficient of correlation, is used to quantify the degree of association or relationship between two variables. We used the bivariant distribution to calculate the correlation, which might be Positive, Negative, Zero, Linear, or Curvilinear in nature. The correlation is stated to be positive when an increase in one variable (X) is followed by a matching increase in the other variable (Y). The range of positive correlations is 0 to +1, with +1 being the perfect positive coefficient of correlation.

The perfect positive correlation states that there is a proportional rise in one variable for every unit increase in the other. "Heat" and "Temperature," for example, have a perfect positive association.

Negative correlation occurs when a rise in one variable (X) causes a commensurate drop in the other variable (Y). The negative correlation varies from 0 to − 1, with the perfect negative correlation occurring at the lower end. The perfect negative correlation states that for every unit increase in one variable, the other decreases proportionally.

A zero correlation indicates that there is no relationship between the two variables X and Y; that is, changes in one variable (X) are unrelated to changes in the other variable (Y) (Y). Body weight and intelligence, for example, shoe size and monthly wage; and so on. The midpoint of the range – 1 to + 1 – is the zero correlation.

Positive correlation is seen if the line moves upward from left to right. Similarly, if the lines go downhill in a left-to-right direction, there will be a negative correlation.

The correlation was calculated using the Scatter Diagram approach in this case. A scatter diagram, often known as a dot diagram, is a visual tool for determining the relationship between two variables. The observed pairs of observations are represented by dots on a graph paper in a two-dimensional space by taking measurements on variable X along the horizontal axis and variable Y along the vertical axis in the preparation of a scatter diagram.

The location of these dots on the graph indicates whether the variable is changing in the same direction or in the opposite direction. It's a quick, easy, but sloppy way of calculating correlation. The frequencies or points are plotted on a graph using scales that are convenient for both series. According to the degree, the plotted points will tend to concentrate in a band of bigger or smaller

width. The direction of 'the line of best fit,' which is drawn freehand, illustrates the nature of correlation.

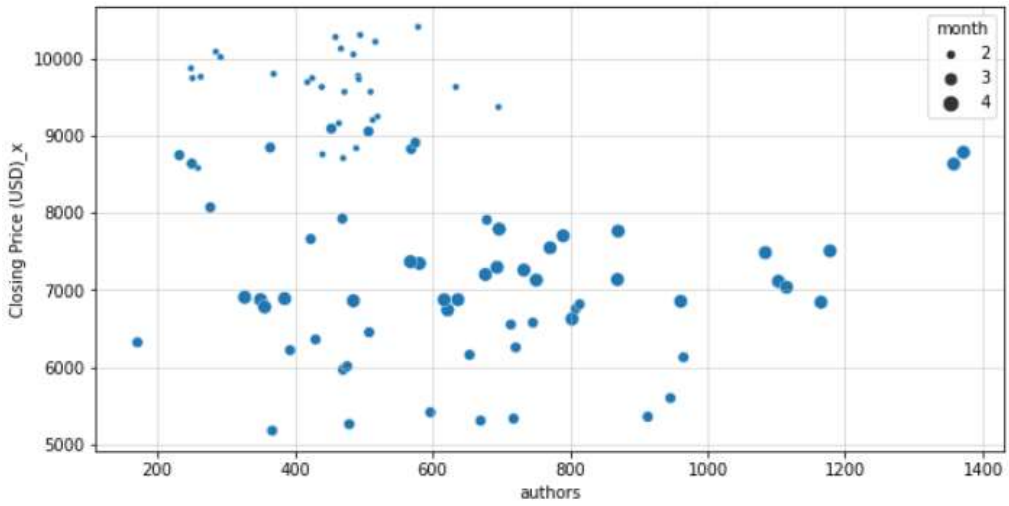

Figure 26: NEGATIVE LINEAR CORRELATION

This linear correlation was confirmed through a scatter plot of the closing price as well as the number of verified authors tweeting. It showed a negative correlation over the months as it goes from the second month of the year to the 4th month. If observed closely we can see that the taking into account the month 2 and 3 and excluding the month 4 we can see that the month 2 observed a lower number of tweets with price remaining higher, while the month 3 observed a negative scenario with increase in number of tweets tending the price to go lower. Overall, we can see that the line can be drawn from the top to the bottom for the month of 2 and 3 creating a negative correlation. However, the month 4 seemed to have a stable price for a longer period of with an increase in price at the end of the month and the reason to be not considered in the final results of analysis.

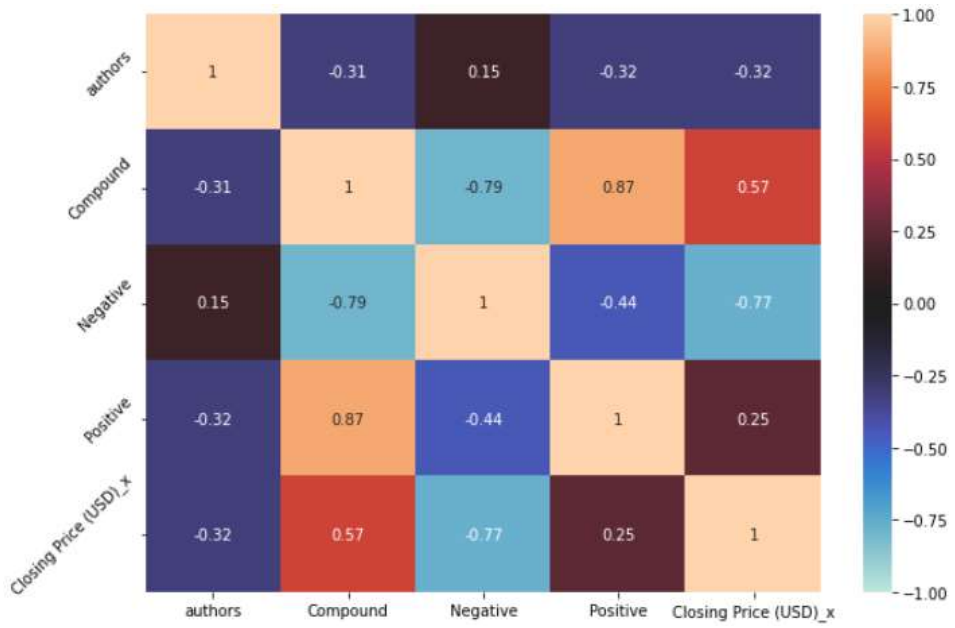

Figure 27: correlation between the number of verified tweets a day and the positive and negative effect on it

Moreover, a heat map was also created comparing various entities that depicts the correlation of the polarities compared to the number of verified authors and the closing price of Bitcoin along with the Positive negative sentiments. As seen here, a correlation can be seen here between the price of bitcoin and the compound and negative sentiments of the tweets. The relationship is between the closing price and the negative sentiments is seen to be negative in nature with a value of -0.77. which not a 100% negative but depicts a certain amount of negative nature of the tweets. Also, the relationship between the authors and the closing price is observed to be negative in nature, meaning that when there is a greater number of verified tweets, the price tends to decrease, and when there is a smaller number of verified tweets, the price tends to increase. This along with the negative correlation seen in the scatter plot lead that when there is increase in the number of tweets, they tend to bring the price down for bitcoin for the observed time frame.

Timeline	Positives	Negatives	Neutral
2020/02/01- 2020/02/28	100-200	0-100	0-350
2020/03/01- 2020/03/31	100-250	0-130	120-130

Table 5: Number of positive, negative, neutral tweets by verified accounts

The above table is a detailed description of what is seen in the graphs above consisting of counts of Positive, Negative, and Neutral tweets during the months of Feb 2020 and March 2020 for the verified user's account.

Timeline	Positives	Negatives	Neutral
2020/02/01- 2020/02/28	10000-28000	0-5000	10000-47000
2020/03/01- 2020/03/31	10000-29000	5000-8500	15000-13000

Table 6: Number of positive, negative, neutral tweets by unverified accounts

The above table is again a description of the graphical data in detail consisting of counts of Positive, Negative, and Neutral tweets during the months of Feb 2020 and March 2020 for the unverified user's account.

Timeline	No. of verified tweets	Price of Bitcoin
2020/02/01- 2020/02/28	230-650	$8562.45-$10326.05
2020/03/01- 2020/03/31	230-980	$4970.79-$9122.55

Table 7: Count of verified tweets per day vs the price of Bitcoin

In the above table comparison of the number of verified tweets vs the Price range of Bitcoin can be seen where we can see that a change in an increased number of tweets results in dropping of the price range of Bitcoin as well. To understand it clearly, we can refer to the graphical data in figure 24 and figure 25 to when on a particular date a negative correlation between these two is observed.

5.3 Discussions

The results present the reader with detailed discussion regarding the various aspects to be utilized to be utilized in this research between Bitcoin and Twitter. There are various variables to consider while analysing, studying, and analysing cryptocurrency fluctuations, particularly the need to comprehend the concepts of both, and there are several factors that play a role in the top of the fluctuation in digital currencies. This consists of the user of the various aspects that are associated with the understanding of the blockchain technology, as well as the group consisting of the several aspects that are helpful in treating the various types of changes in the fluctuations supplied to the cryptocurrency, as well as the several factors of the decentralised the structure of this technology and currency.

Here the report consists of the literature review and the methodology, which are helpful in finding out the several kinds of the past to extend into the past into the same field and the different kinds of the technical aspects and the approach that is being followed in waking up of the use case of this project are respectively

As the report contains the sentiment analysis while also requiring an assortment of data and, as a result, in a neat manner, an effective explanation, as well as deep insights into the many parts of sentiment analysis, has been carried out effectively. All of the information acquired has to be digested and treated with care in order for the final conclusion to be more concrete and relevant. All of the complexities as well as pertinent information had to be handled in order for the data to be as dependable and accurate as feasible. The processing processes were completed quickly and efficiently, employing all available resources.

The first part of the results describes the number of tweets that were either made by verified or unverified users. It has been seen as an obvious as the number of tweets made by verified users are far less than the unverified users considering the verified users on twitter. The polarity of the verified and unverified account is calculated and then they are represented on a day-to-day basis comparison. The data is further divided into a two-week format to further to further understand the difference into detailed manner. The main pattern that was noticed from this was that when the time frame was expanded for visualizing the results seemed to follow a similar kind of pattern for both verified and unverified as well.

Further in the chapter a comparative analysis was also done to see the positive, negative and neutral tweets made by the verified and unverified account was analyzed. The results of this were seen as in the verified tweets taken into the account it was observed that most of the tweets were positive in nature and with negative tweets not farther away and following the same trend as well. While

in the section of unverified tweets it was observed that the neutral tweets however overpowered the positive and negative tweets sentiments and stood above them. It was also seen that the overall results seem to not have a proper pattern to be followed altogether.

The overall tweets which were split into the verified and unverified were also observed using a word cloud. The word cloud of the unverified seemed to be more general in nature with lots of noise data that was observed along with the important stuffs mentioned. Some of the examples included were “free, buy, airdrop, play, giveaway, etc. which overall point to more association of the adds and bots. The word cloud of verified account was observed to be more focused in terms of the keywords seen to be mentioned over here. One surprising keyword that was seen highlighted into the word cloud was the “coronavirus” which is seen to have been associated with a number of tweets as it was the start of the pandemic and lockdown phase. Other than a the mentioning the word does not seem to play a particular role in overall research.

The tweets taking into account were not specified as all the tweets having a mentioning of bitcoin or btc were taken into account to see a more general picture of the effect on price of bitcoin rather than just taking a specified tweet into account like of the avid entrepreneur and CEO of Tesla – Elon Musk. Also, the time frame observed was not particularly popular to point just one particular person and focus on to see how their tweets affect just alone compared to others.

In the later section of the authors effect on the price of bitcoin a comparison was done with the number of verified tweets and the polarity of the tweets. Also a comparison was done with the price of bitcoin and the sentiments of the verified tweets made and overall it was observed that the pattern of negative tweets seems to be following the similar kind of trend along with the price of bitcoin although positive tweets also seem to have following the line but not as often as negative and does not seem to be same.

To further understand the finding on a numerical basis a scatter plot correlation method was used for comparison of the results and from that it was evident that when the number of tweets increased it was seen to have brought the price of bitcoin farther down which was observed for the month of February and March 2020. A heatmap of the results was also created which showed a negative 0.77 relation with the price of Bitcoin. Upon closer looking into the data, it was seen that the despite there were more number of positive tweets of verified users on a daily basis than the total number of negative or neutral tweets the negative tweets seemed to have more influence over the price of bitcoin than the positive tweets and it is therefore believed that smaller number of verified users were seemed to have far more influence for a vast majority of influencers resulting in the change in the price of bitcoin in a vast majority. The overall finding were also described in a table with the average number of tweets made for the following months where the same results are also observed.

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 Conclusion

With regards to the discussions as well as insights along with the practical work presented throughout this particular study, various evaluations could be made which present that the influence provided by the verified authors is indeed a truth upon the various social media platforms as they are capable of influencing the general public regarding the buying as well as selling of any particular cryptocurrency, which has also been proving one regarding the practical implementation which has been carried out towards presenting the graphical visualizations in the above section.

Still, various questions or queries have arisen, such as

What is the actual number of tweets to be posted by various verified authors so as to experience a drastic or evident rise or fall regarding the popularity of Bitcoin or any particular cryptocurrency?

The prior question also provides a query for this that how many tweets, in particular, would be required by any verified author towards implementing the same change. The answers to these queries were also obtained throughout this particular research as it was identified that when a detailed analysis is focused upon the significant amount of others as well as two of it's presented by them along with comparing these two factors against that of the price aspect along with the polarities, the correlation among these particular factors could indeed be discovered with suggests that the number of verified authors along with their tweets relating Bitcoin indeed impacts in the

increase as well as decrease of the cryptography prices as well as their popularity among the general public.

Since this correlation is negative in nature, when the number of tweets from the various authors about Bitcoin decreases, the price of Bitcoin increases.

The approach of sentiment analysis implemented towards extracting and operating upon the Twitter information was carried out by the implementation of strategically developed and administration of various calculations, which comprised of a multitude of procedures. The finding were seen that smaller number of people of twitter having verified account do seem to have a far more influential approach than the other users.

6.2 Future Work

There are several kinds of future works associated with this kind of research, and this can be done in by means of protecting the various kinds of analysis. Into the making of the future work will involve considering not just the tweets from a particular time frame but expanding the time frame throughout the life span of the Bitcoin and see if the same trend follows. The same will be followed for some of the other Cryptocurrencies that have an influence on other Cryptocurrencies. This will help see if the same trends are followed for tweets made for those Cryptocurrencies for verified and non-verified users. Furthermore, investigating other factors will be taken into account into the twitter itself by focusing it on specific regions of North America and Asia, which is currently believed to be having more influence through tweets over the cryptocurrencies. The future work will also involve not just looking over the twitter data but will be expanded to other social media platforms as well, with Reddit being the first in line, followed by Google Trends and Wikipedia views data. The findings of these data of social media will then be taking into account the factors

other than social media that are believed to be involved in the change of price of cryptocurrencies along with the current research to help create a Machine Learning or Artificial Intelligence model that will help predict the future price of bitcoin and other cryptocurrencies which currently has inaccurate prediction rate. Weather means of analyzing and working on this point the more impactful analysis can be done in and fluctuations based on these factors be obtained and can be obtained and by means of such research the more efficient and effective analysis can be done and by finding the most relevant information and of the data analysis.

References

Pant, D. R., Neupane, P., Poudel, A., Pokhrel, A. K., & Lama, B. K. (2018). Recurrent neural network-based bitcoin price prediction by Twitter sentiment analysis. In *2018 IEEE 3rd International Conference on Computing, Communication, and Security (ICCCS)* (pp. 128-132). IEEE.

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.

Desai, M., & Mehta, M. A. (2016). Techniques for sentiment analysis of Twitter data: A comprehensive survey. In 2016 International Conference on Computing, Communication and Automation (ICCCA) (pp. 149-154). IEEE.

Pano, T., & Kashef, R. (2020). A Complete VADER-Based Sentiment Analysis of Bitcoin (BTC) Tweets during the Era of COVID-19. *Big Data and Cognitive Computing*, *4*(4), 33.

Lamon, C., Nielsen, E., & Redondo, E. (2017). Cryptocurrency price prediction using news and social media sentiment. *SMU Data Sci. Rev*, *1*(3), 1-22.

Hassan, M. K., Hudaefi, F. A., & Caraka, R. E. (2021). Mining netizen's opinion on cryptocurrency: sentiment analysis of Twitter data. *Studies in Economics and Finance*.

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.

Cheuque Cerda, G., & L. Reutter, J. (2019,). Bitcoin price prediction through opinion mining. In *Companion Proceedings of The 2019 World Wide Web Conference* (pp. 755-762).

Béjaoui, A., Mgadmi, N., Moussa, W., & Sadraoui, T. (2021). A short-and long-term analysis of the nexus between Bitcoin, social media and Covid-19 outbreak. *Heliyon*, *7*(7), e07539.

sibel KERVANCI, I., & Fatih, A. K. A. Y. (2020). Review on Bitcoin Price Prediction Using Machine Learning and Statistical Methods. *Sakarya University Journal of Computer and Information Sciences*, *3*(3), 272-282.

Corbet, S., Hou, Y. G., Hu, Y., Larkin, C., & Oxley, L. (2020). Any port in a storm: Cryptocurrency safe-havens during the COVID-19 pandemic. *Economics Letters*, *194*, 109377.

Karalevicius, V., Degrande, N., & De Weerdt, J. (2018). Using sentiment analysis to predict interday Bitcoin price movements. *The Journal of Risk Finance*.

Rouhani, S., & Abedin, E. (2019). Crypto-currencies narrated on tweets: a sentiment analysis approach. *International Journal of Ethics and Systems*.

Shevlin, R. 2021, How Elon Musk Moves The Price Of Bitcoin With His Twitter Activity. Forbes. https://www.forbes.com/sites/ronshevlin/2021/02/21/how-elon-musk-moves-the-price-of-bitcoin-with-his-twitter-activity/?sh=4164164a5d27

Ajmi, A., Youssef, M., and Mokni, K. 2022. COVID-19 pandemic and economic policy uncertainty: The first test on the hedging and safe haven properties of Cryptocurrencies. Research in International Business and Finance, 60(4), 1-13.

Shen, D., Urquhart, A. and Wang, P. 2019. Does Twitter predict Bitcoin? Economics Letters, 174, 118-122. ISSN 0165-1765.

Bharathi, S., Geetha, A., and Sathiynarayanan, R. 2017. Sentiment analysis of twitter and RSS news feeds and its impact on stock market prediction. International Journal of Intelligent Engineering and Systems, 10(6), 68-77.

Hasan, A., Moin, S., Karim, A., and Shamshirband, S. 2018. Machine learning-based sentiment analysis for twitter accounts. Mathematical and Computational Applications, 23(1), 11.

Kolasani, S. V., & Assaf, R. (2020). Predicting Stock Movement Using Sentiment Analysis of Twitter Feed with Neural Networks. Journal of Data Analysis and Information Processing, 8(4), 309-319.

Chethan, N., and Sangeetha, R. 2020. Sentiment Analysis of Twitter Data to Examine the Movement of Exchange Rate and Sensex. Journal of Computational and Theoretical Nanoscience, 17(8), 3323-3327.

Huang, X., Zhang, W., Huang, Y., Tang, X., Zhang, M., Surbiryala, J., ... and Zhang, J. 2021. LSTM Based Sentiment Analysis for Cryptocurrency Prediction. arXiv preprint arXiv:2103.14804.

Khokale, M. N. K., &Tundalwar, R. (2021). STUDY OF SENTIMENT ANALYSIS OF TWITTER USERS FROM THEIR TWEET'S TEXT AND DIFFUSION PATTERNS. INTERNATIONAL JOURNAL, 6(5).

Xia, E., Yue, H., & Liu, H. (2021). Tweet Sentiment Analysis of the 2020 US Presidential Election. In Companion Proceedings of the Web Conference 2021 (pp. 367-371). (Xia et al., 2021)

Meduru, M., Mahimkar, A., Subramanian, K., Padiya, P. Y., &Gunjgur, P. N. (2017). Opinion mining using twitter feeds for political analysis. Int. J. Comput.(IJC), 25(1), 116-123.

Das, S., Behera, R. K., & Rath, S. K. (2018). Real-time sentiment analysis of twitter streaming data for stock prediction. Procedia computer science, 132, 956-964.

Kamyab, M., Tao, R., Mohammadi, M. H., and Rasool, A. 2018. Sentiment analysis on Twitter: A text mining approach to the Afghanistan status reviews. In Proceedings of the 2018 International Conference on Artificial Intelligence and Virtual Reality (pp. 14-19).

Mehta, P., and Pandya, S. 2020). A review on sentiment analysis methodologies, practices and applications. International Journal of Scientific and Technology Research, 9(2), 601-609.

Adwan, O., Al-Tawil, M., Huneiti, A., Shahin, R., Zayed, A. A., and Al-Dibsi, R. 2020. Twitter sentiment analysis approaches: A survey. International Journal of Emerging Technologies in Learning (iJET), 15(15), 79-93.

Gondaliya, C., Patel, A., and Shah, T. 2021. Sentiment analysis and prediction of Indian stock market amid Covid-19 pandemic. In IOP Conference Series: Materials Science and Engineering (Vol. 1020, No. 1, p. 012023). IOP Publishing.

Biswas, S., Ghosh, A., Chakraborty, S., Roy, S., and Bose, R. 2020. Scope of Sentiment Analysis on News Articles Regarding Stock Market and GDP in Struggling Economic Condition. International Journal, 8(7).

Gayathri, M., Nisha, S. S., andSathik, M. M. 2020. Twitter Sentiment Analysis: Survey. International Journal of Engineering Research & Technology (IJERT), 8(3).

Pawar, M. P., and Agarkar, P. 2020. TWITTER SENTIMENT ANALYSIS USING TEXTUAL INFORMATION AND DIFFUSION PATTERNS. INTERNATIONAL JOURNAL, 5(7).

KEMALOĞLU, N., KÜÇÜKSİLLE, E., and ÖZGÜNSÜR, M. E. 2021. Turkish Sentiment Analysis on Social Media. Sakarya University Journal of Science, 25(3), 629-638.

Exploring the Negativity of Black ... - Data Meets Media. https://piocalderon.github.io/exploring-the-negativity-of-black-mirror-with-vader/

A Hybrid Approach to Explore Public Sentiments on COVID-19. https://link.springer.com/article/10.1007/s42979-022-01112-1

Correlation: Meaning, Types and Its Computation | Statistics. https://www.yourarticlelibrary.com/statistics-2/correlation-meaning-types-and-its-computation-statistics/92001

International Journal of Trend in Scientific Research and https://www.ijtsrd.com/papers/ijtsrd17043.pdf

Explain the various kinds of correlation . - Sarthaks https://www.sarthaks.com/1984995/explain-the-various-kinds-of-correlation