

PREDICTING BITCOIN PRICE FLUCTUATIONS BY
ANALYZING GLOBAL CURRENCY PATTERNS & SENTIMENTS

By

KSHITIJ CHHATWANI

A thesis submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Master of Science

Graduate Program in Electrical & Computer Engineering

Written under the direction of

Maria Striki

And approved by

New Brunswick, New Jersey

JANUARY, 2019

ABSTRACT OF THE THESIS

PREDICTING BITCOIN PRICE FLUCTUATIONS BY
ANALYZING GLOBAL CURRENCY PATTERNS & SENTIMENTS

By KSHITIJ CHHATWANI

Thesis Director:

Maria Striki

From the past two years with increasing geopolitical and economic issues, global currency values have been falling and stock markets have been having a poor run & investors losing wealth. This has led to a renewal of interest in digital currencies. Bitcoin, one of the most prominent digital currency has found itself in spotlight with investors wanting a piece of it and business establishments accepting it as a source of payment due to its stable performance in the last few years. A lot of research has been done on predicting Bitcoin prices using Machine Learning and Twitter Sentiment Analysis. On the same lines, we are analyzing bitcoin prices using Machine Learning and Sentiment Analysis. We also study stock market trends in order to better predict bitcoin prices quantitatively. In this work we analyze the impact of global currencies like US Dollar, foreign exchanges on Bitcoin prices and whether Bitcoin has the stability to dethrone global currencies and become the single medium of transaction.

Acknowledgement

I would first like to thank my advisor Prof. Maria Striki, faculty in the Electrical & Computer Engineering (ECE) Department at Rutgers, The State University of New Jersey for her guidance, direction and immense support throughout this journey. Prof. Striki has always encouraged and motivated me, be it in the courses I took under her and for this research.

I would also like to thank Dr. Zoran Gajic, graduate director of the ECE department at Rutgers, The State University of New Jersey for his guidance which helped me in course selection and graduation timeline. The courses at the ECE Department have been a huge support for me in gaining new insights towards my research and the faculty that have devised and taught these courses for their important contributions. I thank the rest of my committee members Dr. Athina Petropulu and Dr. Zoran Gajic for taking out time from their busy schedules to review, critique and share their feedback. It feels great to have helpful ECE administrative staff like Christy Lafferty, Arletta Hoscilowicz and John Scaffidi for assisting me and answering my queries.

A big thank you to my friends, roommates and on-campus job supervisors for guiding me and making my time at Rutgers, The State University of New Jersey memorable and a learning experience.

Lastly, I would like to express my gratitude towards my parents and sister for motivating me during all those times when I felt low, facing my tantrums, letting me give greater priority to my work over them. I would also like to thank Bhagyashri Soman for being my

pillar of support during times when I wanted to give up and found myself aimless and Raghav Bharadwaj for guiding me with his expertise during this research.

Table of Contents

Abstract of the Thesis.....	ii
Acknowledgement.....	iii
List of Figures	viii
1. Introduction	1
1.1 Introduction & Research Background.....	1
1.2 Research Background.....	1
1.2.1 Bitcoin	1
1.2.2 Machine Learning	3
1.2.3 Sentiment Analysis.....	4
1.3 Research Objectives	6
2. Literature Review	7
2.1 Introduction	7
2.2 How Market prices are affected by Sentiment.....	7
2.3 US Dollar leading world currency	9
2.4 Bitcoin as an investment	11
3. Method	14
3.1 Data Collection.....	14
3.1.1 Bitcoin Price Data	14
3.1.2 Gathering Sentiment Data	15
3.2 Bitcoin Prediction Process	15

3.2.1	Training and Testing Dataset Creation.....	15
3.2.2	Building Neural Networks Model.....	16
3.3	Sentiment Analysis Process	17
3.3.1	Reducing noise from obtained data.....	17
3.3.2	Understanding Sentiment data	18
3.3.3	Developing Sentiment Analysis model.....	19
3.4	Text Mining Data	19
3.4.1	News Articles	20
3.4.2	Reactions on Interest Rates, Foreign Exchange and US Dollar.....	20
3.5	Time Series Analysis.....	21
3.6	Conclusion.....	21
4.	Working.....	22
4.1	Introduction	22
4.2	Dataset Creation	22
4.3	Prediction of Bitcoin Price	23
4.3.1	Keras.....	23
4.3.2	TensorFlow.....	25
4.4	Time Series Analysis on Data	26
4.5	Web Scrapping Data.....	27
4.6	Sentiment Analysis.....	30
5.	Findings and Analysis	34

5.1	Results from Bitcoin Prediction Process.....	34
5.2	Sentiment Data Results	36
5.3	Time Series Results.....	37
5.4	Impact of US Dollar & Federal Interest rates on Bitcoin.....	39
5.4.1	Bitcoin & US Dollar.....	40
5.4.2	Impact of US Dollar on Global Foreign Exchanges	41
5.5	Reasons for fall of Bitcoin prices 2017 vs 2018	42
6.	Conclusions	44
6.1	Introduction	44
6.2	Conclusions	44
6.3	Limitations	45
6.4	Future Improvements	46
7.	References	47

List of Figures

Figure 1: Comparison US Dollar index with other global currencies.....	10
Figure 2:Global Foreign Exchange distribution.....	11
Figure 3: Bitcoin price chart for the past 1 year.....	12
Figure 4 : Bitcoin Price Prediction (as of 11/15/2018)	34
Figure 5: Bitcoin Price Prediction (as of 11/26/2018)	35
Figure 6 : Sentiment Data obtained from Text Mining.....	36
Figure 7 : Time Series Result (window size = 25).....	38
Figure 8 : Time Series Result (window size = 50).....	39
Figure 9 : US Dollar vs Bitcoin (YTD).....	40
Figure 10 : Impact on Global currencies post Midterm Elections (November 2018).....	41

1. Introduction

1.1 Introduction & Research Background

This section focuses on providing the information related to the multiple research objectives that I had developed before selecting this topic. The research topic is introduced followed by the research objectives and follow-up questions. The importance of those objectives and their outcomes are discussed.

1.2 Research Background

1.2.1 Bitcoin

Bitcoin, an idea about a form of peer to peer digital cash system was put forward by Satoshi Nakamoto during November 2008 when all stock markets and business establishments as well as banking establishments were reeling from a huge economic turmoil. Satoshi wanted to model Bitcoin as a decentralized digital currency which was away from the control of a single bank so as to prevent it from future effects of any such economic turmoil. In-order to ensure that Bitcoin is not spent recklessly or to further keep security and practicality of Bitcoin, a central ledger was planned & built that would keep track of all transactions taking place. The unique aspect of this ledger is that it is controlled by all the users who own bitcoins and not by a centralized entity such as banks or financial firms.

This ledger was named 'Blockchain'. Blockchain keeps track of all Bitcoin transactions occurring anywhere in the world. It is a cryptographic implementation that provides the highest security towards ensuring that misuse or leak of Bitcoin data does not occur. It is implemented as a chain of blocks where each block contains the hash of the previous block. A network of nodes keeps the network up & running as well as maintain the Blockchain.

A key feature of implementing Blockchain for keeping a check on Bitcoin transactions was for avoiding double-spending. Double-spending can be explained as a drawback of digital cash systems where a digital coin or token can be spent more than once if the details of the digital coin or token are duplicated. This majorly affects currency values since devaluations of such currency takes place.

Transactions in blockchain are confirmed using Proof of Work concept. Proof of work makes use of hash functions, mathematical computations and solving a puzzle in order to ensure the legitimacy of bitcoin transactions. Proof of work makes use of Merkle Trees, which store the hash of each data block. This technique allows efficient and secure verification of each data block inside the blockchain. The computations however in proof of work require time and resources in terms of graphical processing units and continuous power supply.

The advantage of using a proof of work algorithm is to ensure that the network is safe from Denial of Service attacks which have become rampant nowadays. This is because such an attack would require a huge amount of time & computation cost in order to be successful and even though bitcoin exchanges and wallets have been hacked due to poor security protocols, bitcoins stolen haven't been able to be cracked and used because of this important feature.

As we talk about blockchain and proof of work for ensuring security of Bitcoins, it is a matter of interest to further talk about how bitcoins are generated. The process of obtaining bitcoins is called as Mining. In simple term, bitcoins are awarded when miners are able to solve some puzzles which require heavy computations that take days or months even. Once they have solved the puzzle, a new block is formed and the transactions are confirmed. On successful confirmation, the miners are awarded bitcoins based on their work.

Bitcoin mining has been rampant in countries like Russia and China where old warehouses and industry units have been converted into mining farms. These farms have been setup with hundreds of CPU's and GPU's that work consume huge amounts of electricity and work 24 hours a day 7 days a week in order to mine bitcoins. This clearly shows the huge amount of investment that needs to be made for a return which could be a petty amount or turn the investor a millionaire by stroke of luck.

Bitcoin exchanges have come up in countries like South Korea, Japan, United States to name a few. These exchanges serve as fronts for helping people trade bitcoins. Bitcoins wallets, a form of digital wallets with greater security protocols built in have been developed for storing and using bitcoins. Even business establishments have started accepting bitcoins as payment for services & products and big banking firms like

Morgan Stanley have set up bitcoin trading desks to work more on exploring bitcoins and tending to clients who are interested in it.

Unlike traditional paper currency which can be printed as per market needs, Bitcoin has a limited supply. This is in order to ensure that printer inflation does not occur and the currency does not get devalued. However, due to the limited supply of Bitcoins and with 80% bitcoins already mined by mid-July 2018, it is anticipated that the remaining of the 21 million worth of Bitcoins will take a large amount of time to mine and in order to perform such large computations, relative infrastructure will also be required with the possibility of a low return in the future.

With a general acceptance and presence of Bitcoin since the last few years, other cryptocurrencies or digital currencies as we call them have also come to reality and become mainstream. Ethereum and Litecoin are two of the most commonly known cryptocurrencies which have been used for developing different applications and further researched by banking firms for future implementation.

1.2.2 Machine Learning

Tom Mitchell's widely quoted statement on algorithms used in machine learning as "A computer program is said to learn from some experience E from some class of tasks T and performance measure P if its performance in tasks T as measured by P improves with experience E ". This statement gives a fair idea about the type of processes that take place in machine learning.

The tasks that we have mentioned above can be classified into three major categories:

- **Supervised Learning:** Here, the datasets contains examples from which the computer can learn and produce results accordingly. The learning algorithm generalizes the results based on the training data provided to it.
- **Unsupervised Learning:** Here, the datasets are provided with no examples or patterns. It is expected of the algorithm to find patterns in the data and apply the results to the testing data which is provided to it subsequently.

- Reinforcement Learning: Here, datasets contain feedbacks which are fed to the training network in order to identify the patterns by itself and apply those results to testing data.

The tasks can be used for classifying data such as images or numbers in order to find a pattern or trends. For example, machine learning algorithms can be used to classify between objects on a street such as pedestrians, vehicles and street signs & other non-static objects. Similarly, machine learning algorithms can also be used for forecasting or predicting future data such as future stock prediction, predicting how a new product would be perceived in the market to name a few examples.

These days machine learning is being used in game development and on equipment's sent in probes to outer space. This tremendous progress in applications of machine learning are due to the fact that it improves the analytics and decision-making capabilities of applications.

We have plethora of machine learning frameworks available to use for depending on the type of application that needs to be created. A few of them are as follows:

- TensorFlow
- Keras
- Microsoft Cognitive Toolkit / CNTK
- DeepLearning

Tensorflow is an open source software library for high performance numerical computations. It can be easily deployed on a variety of platforms including cloud as well. Keras is a high- (TensorFlow, n.d.)level neural networks API to build and train deep learning models. Similar to Tensorflow, Keras is also written in Python. Keras is user friendly and supports convolutional networks as well as recurrent networks.

1.2.3 Sentiment Analysis

Merriam -Webster defines Sentiment as a specific view or judgement prompted by feeling. Using sentiments for gauging a particular event or reaction has been used widely in order to understand consumer thoughts and likewise used by businesses. We use the term Sentiment Analysis for determining the perception, reaction to a document, event or interaction. There are several different applications of sentiment analysis ranging from

looking at images & predicting the thought or feeling that the image conveys to using sentiment analysis for understanding consumers approach towards different products.

Sentiment analysis classifies this reaction as either positive, negative or neutral. It is based on analyzing the words that are present in a sentence to find the overall polarity of the sentence. We can develop techniques for analyzing sentiments using machine learning, natural language processing on large amounts of data including webpages, news articles, discussions, web blogs to name a few sources. As said above, sentiment analysis can also be performed on images and videos nowadays.

There are several different techniques to perform sentiment analysis. We can use several supervised learning techniques to train our data and then test the results on new data. Naïve Bayes is one of the most widely used supervised learning techniques for analyzing and finding the sentiment of a sentence. The twitter sentiment analysis library TextBlob is based on Naïve Bayes as well. With new algorithms and technologies being developed with every passing day, there have been new methods devised to find the sentiments. One such exciting technique is using TensorFlow which is a machine learning framework. TensorFlow uses a dictionary approach or a bag of words model to firstly develop a prediction result from the training data. After this, based on the test input that is provided, it accordingly predicts the sentiment of the test input. (Wikipedia, n.d.).

Our application of sentiment analysis is majorly directed towards understanding the reactions of investors and traders towards Bitcoin's volatile prices. We can further analyze sentiments towards US Dollar and the global economic scenario which has not been doing so well since the last few months. All this helps in trying to find ways to support our prediction.

1.3 Research Objectives

This research thesis focuses on the following major objectives:

- Is it possible to have Bitcoin as a publicly traded commodity?

The results for the above objective can be then applied to findings of the following objectives:

- Can Bitcoin be used as any other traditional currency like US Dollar, Euro, Pound Sterling and so on.
- Does Bitcoin have the ability to become a primary method of transaction, replacing US Dollar and other traditional currencies.

2. Literature Review

2.1 Introduction

The aim of this research is to examine whether the price of Bitcoin can be predicted similar to other stock market tickers. This will have a basis on whether we can further use it as a medium of payment. In order to arrive at this, we will look at factors such as if bitcoin can be used as an investment. This will help in providing support to our objectives. We will also analyze how sentiments affects bitcoin and whether it is similar to the manner in which stock markets are influenced by sentiments. This will lead to drawing similarities and differences between how bitcoin and stocks perform

2.2 How Market prices are affected by Sentiment

A company's financial evaluation is done on the basis of certain fundamentals which are qualitative and quantitative data. These factors are examined by Analysts and Traders in order to estimate whether an asset, company are performing well and whether investment can be made in it. Apart from these technical indicators, traders also rely on market sentiments or investor sentiments. These are sentiments that influence technical indicators as well since they can be used to make a profit by predicting short term price movements. Although these predictions are irrational investment decisions but since they yield a result, we can say that the decision was driven by sentiment rather than by facts.

Stock markets use different indexes to measure sentiments, some of them are CBOE Volatility index, 52 week High/Low Sentiment Ratio and 50-day moving average. After analyzing a trend, traders invest either in the direction of sentiment or at times opposite to the sentiment. The high/low sentiment ratio indicator compares number of stocks making 52-week highs with those making 52-week lows. They can either be bullish outlook or a bearish outlook depending on the ratio of highs and lows.

Market sentiments are also used to study and make decisions on how different sectors of the economy are performing. (Salzman, 2018) reported that market shares of Nike and Under Armor were reducing and the same time Adidas was experiencing

greater customer numbers. This was found from a survey floated sometime before the news was corroborated and made publicly. Such surveys and other types of studies such as following the number of users in a public place over a period of time by analyzing their cellular records helped in greater understanding of predicting profitability of a public place.

(Baker & Wurgler, 2007) extensively studies how investor let sentiments influence their investment decisions, developing a top-down approach that measures investor sentiments. The authors also provide a series of future challenges that can be overcome in order to better predict and understand investor sentiment. (Lemmon & Portniaguina, 2006) have used Michigan Consumer Confidence Index as a proxy for understanding investor sentiment. A time series framework was constructed, and the authors indicated that for equity portfolios, the time variation was based on consumer confidence. (Qiu & Welch, September 2004) have used Michigan Consumer Confidence with closed-end fund discount (CEFD) to ascertain that sentiment changes impact returns of stock markets. (Corredor, Ferrer, & Santamaria, 2013) have analyzed extensively the investor sentiment in four major European countries: Germany, France, U.K and Spain and the findings corroborate that sentiment has an influence on returns, with differing intensities across each market.

(Anjaria & Guddeti, Influence Factor Based Opinion Mining of Twitter Data Using Supervised Learning, 2014) have very well covered the influence of sentiments on prediction of electoral results for the US Presidential elections and Indian state elections. This paper guides home the point that the basis of an outcome of an important democratic activity like elections can have a huge bearing on the performance of stock markets. We have all seen several instances of this such as the stock markets rallying in green when a market friendly government is elected. (Watts, 2018) has provided a report speculating the impact of recently held Midterm Elections in November 2018 on U.S stock markets. The report covers excerpts from different banking directors and a thorough analysis of historical performances of stocks during election times which were directly impacted by the outcomes of the elections. However, a speculation is also drawn which tries to dispel these assumptions in light of the fact that stocks tend to sway more not due to the results of the winners which might have a little say but also due to the implications that arise from different policy decisions that are waiting to be implemented and might not be now due to a change in government rhetoric after election outcomes.

2.3 US Dollar leading world currency

During the times of the world wars and around, governments used to value their currency against gold. This was known as the *gold standard*. The Brenton woods agreement of 1944 however removed the gold standard completely. Based on the amount of gold each country held, the US government pushed forward for the adoption of US dollar as currency for exchanges since it held about three-fourths of the total world supply in gold. Now countries could set up their exchange rates for local currency versus US dollar.

With time the value of dollar started increasing and thereby its demand too. The International Monetary Fund and World Bank were set up as a result so that member countries could contribute their share as well borrow in order to maintain stability in local currencies and for economic development projects. But on account of massive inflation in the US during the 1970's and with the supply of dollar being very high related to the demand, people started redeeming their dollars for gold. This led to depletion of gold reserves at Fort Knox, the US government de-linked US dollar from gold which has led to the floating exchange rates that we see today.

After the events of 2008, the Feds had to step in to rescue businesses and banks. They printed a lot of dollars and supplied it to the markets in order to ease the situation. However, this excess supply of dollar found its way to emerging markets as well which meant a greater than supply of US dollars worldwide. This caused a drop-in dollar's value against Euro, Yen and other currencies. This measure taken by Feds called as quantitative easing program ended in 2014. Since then there had been speculation of economy coming back to a normal phase and this would lead to increase in interest rates which have been very low since 2008.

This speculation although increased the value of dollar had a negative effect on other currencies since investors and traders were willing to trade in dollars instead of other currencies. As the demand for dollars increase, it has a spiraling impact on the local currency since the exchange value or the amount of local currency to be spent for buying one-dollar increases. This spells a doom for the local country economy.

There have been 5 interest rate hikes in the last one year with the latest one being in September 2018. The graph below depicts how other global currencies such as Euro, Yen, Pound Sterling have fared since the start of 2018.



Figure 1: Comparison US Dollar index with other global currencies.

As we observe Figure 1, we can clearly see that with the increase in interest rates it's a signal that the US economy is improving. Lowering cost of prices, lowering unemployment ratios are some of the factors that are again improving investor confidence in US dollar bonds. However, as discussed above this also shifts the interests of investors from other global currencies to US dollar for security.

This has had a spiraling effect on economies of countries like Turkey, Argentina who have been seeing their currency devalued with every passing day. Other countries such as China have been struggling hard to ensure that Chinese yuan is still comparable to global currencies due to the ongoing trade wars. In order to do this, they have been devaluing their currency, but it has had a serious impact on countries that borrow from China.

The figure 2 below depicts the percentage distribution of different currencies globally. We can clearly see that US Dollar leads with 90% presence in the markets followed by 31% presence of Euro and other global currencies.

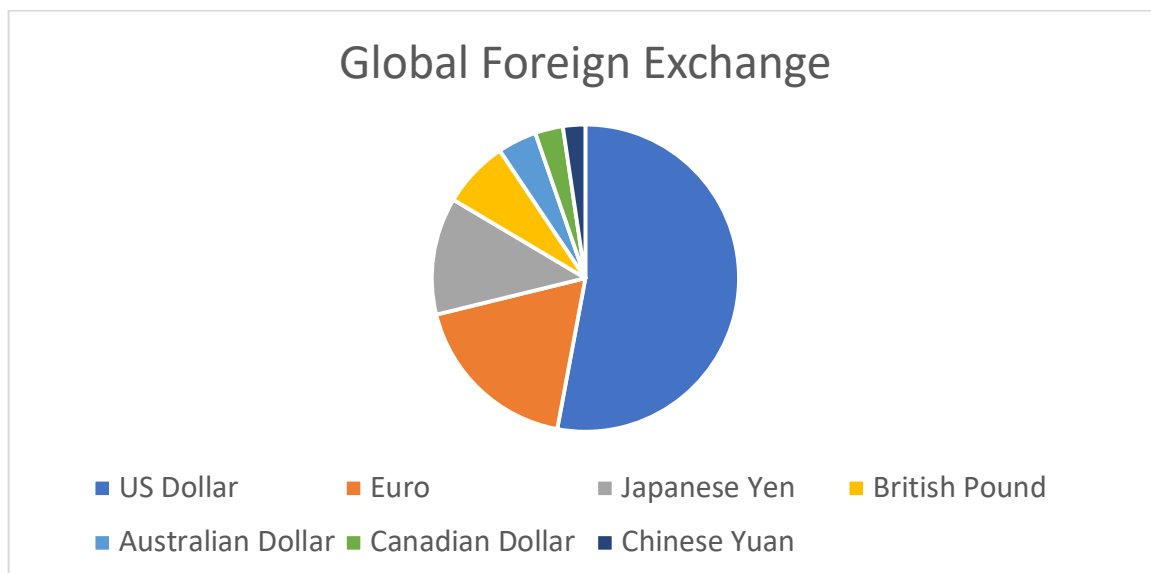


Figure 2: Global Foreign Exchange distribution

It wouldn't be right to say that dollar has a monopoly over world trade and any changes that affect the dollar have a cascading effect on other commodities too. However, this has been the way things were set up by the erstwhile governments. It's solely the author's view that there needs to be a new global currency which is immune from the impacts that a single currency causes globally. This global currency has its advantages in terms of not affecting developing countries' economy such as Argentina's pesos and Turkey's lira which are running in the red due to dollar becoming stronger of late.

2.4 Bitcoin as an investment

The volatility in the price of Bitcoin generated a lot of attention in the latter half of 2017 when the price of bitcoin was steadily rising and at one point one can say it was a steep rise as well. This rise not only attracted the attention of the general public which saw it as a way to increase their fortunes but also brought the focus of many hedge funds, banking corporations and even Wall Street and other world stock exchanges to it. Chicago Exchange Board did an in-depth study and introduced Bitcoin Futures trading with an eye to the future and to attract investors.

The media also has carried stories of different individuals becoming overnight millionaires due to the rapid increase in bitcoin value. Banking corporation CEO's have been speculating a lot on whether to open bitcoin

trading desks and support operations or not. Businesses have sprung up which solely use bitcoin for transactions such as Propy, a California based Real Estate agency that uses Bitcoin and other cryptocurrency for real estate transactions and is quite active on Twitter.

With all the attention on bitcoin, it seems quite an innovative idea to invest in Bitcoin and earn a fortune on your investment. However, before taking a step towards deciding and investing, let us first analyze the feasibility of investing in Bitcoin compared to investing in other stable stocks as from the above sections we do note that Bitcoin prices are volatile. This volatility is due to many factors. The primary of all these factors is the attention it is getting from individuals. We call this as the sentiment people have towards Bitcoin.

Bitcoin has been directly and indirectly influenced by sentiments that people have towards it, government policy decisions after gauging people's perception. This has caused a huge volatility in the prices as we see in the Figure 3. Such volatility only introduces a sentiment of apprehension in people about whether to invest or not. Bitcoin has been treated more like a property that can be purchased and kept with a hope of having an increase in its value rather than spending it. This affects the price of bitcoin as well since it was deemed to be a digital currency and not an investment property.



Figure 3: Bitcoin price chart for the past 1 year

An interesting fact to note is that Bitcoin has created an industry of itself. One of the mentions above was that of Propy. Apart from it, there are several other firms that's do business in bitcoin and other cryptocurrency, notably ethereum. There have been several bitcoin exchanges coming up in every country,

out of which major ones are located in United States, Japan and South Korea. These bitcoin exchanges serve as fronts for investors and clients who are interested in investing in bitcoin. They also serve as points of contact with governments and local income tax departments.

Bitcoin is based on blockchain technology, a secure platform which is being adopted in developing products by major banks worldwide as well as by companies in medical domain in order to provide clients with greater security for their data. In this manner, investment can be made in blockchain so as to not get tangled with Bitcoin and its associated volatility. Organizations such as IBM, Nvidia, Microsoft have been adopting blockchain technologies to fit in with their client demands.

Another manner of indirectly investing in Bitcoin is by mining Bitcoins. Companies like Nvidia develop processors that can perform complex computations as required by algorithms designed to produce bitcoins. It should be noted here though that mining bitcoins is a complex task in itself since it consumes a lot of electricity, time and resources in order to mine a single bitcoin. Greater analyses would be required before partaking this option which will include the amount of money that can be invested in setting such an equipment, understanding about the volatility of the returns and the legal aspect of declaring to the federal government of bitcoin ownership.

3. Method

3.1 Data Collection

The data for performing this research was collected from multiple sources which included Yahoo Finance, Twitter, Web blogs. This data was further validated in order to ensure that there were no garbage values present which could have affected the end results in any manner.

3.1.1 Bitcoin Price Data

Bitcoin price data (BTC-USD) was initially collected from Yahoo Finance. The data collected was cross-checked with that from Coinbase in order to avoid any unwanted discrepancies. Data collection in the beginning stage was manual since Yahoo Finance had deprecated their API Library which was earlier used on Python to seamlessly fetch data and save it directly by making an API call to the website.

However, in due course it was realized that the existing data had values which even though were compensated for, still affected the prediction and thereby produced wrong results. Hence, a shift was made to using REST API's from Quandl which proved to be significant since the results now obtained were more accurate and meaningful.

Three different Bitcoin datasets were collected from three different sources:

- Kraken
- Coinbase
- BitStamp

Quandl provided API's to fetch these datasets. The obtained datasets were then averaged into one dataset for consistency and in order to fill in the gaps created by missing data in either of the datasets. This obtained data was provided to Keras for prediction purposes.

3.1.2 Gathering Sentiment Data

To collect data for sentiment analysis from Twitter, tweets from different twitter handles were collected by using Twitter API's for python. Twitter API's use Tweepy, a library written in Python that enables the user to search for tweets based on specific keywords and fetches all the tweets related to the keyword. This by far is one of the most efficient unpaid services for gathering tweets from Twitter. For older data and better-quality tweets, there were also paid services whose API's easily fetched quality data.

Quality of tweets is essential for performing sentiment analysis since the results obtained affect other parameters taken into consideration. The keywords, for example "cryptocurrency" also included tweets on un-related topics as well. Hence, there was a need to filter the tweets being fetched and select appropriate ones. In absence of filter mechanisms provided by Twitter, new ones were created which used regular expressions library of python to remove unwanted letters and duplicated.

Apart from using Twitter tweets, as the work progressed web scraping news articles was also performed in order to have a greater accurate pool of data which could be helpful in analyzing and arriving to conclusions.

3.2 Bitcoin Prediction Process

The prediction process for Bitcoin involves using Machine Learning techniques for running a neural networks-based prediction model for predicting the next day price. The test data and training datasets were created keeping in mind the importance to produce an accurate or near-to-accurate prediction

3.2.1 Training and Testing Dataset Creation

A good training dataset and testing dataset can help to achieve an objective with greater accuracy compared to a poorly created dataset. With this in mind, multiple datasets were extensively created and tested multiple times in order to land up to the dataset which could be free from any discrepancies.

As discussed earlier in 3.1.1, we started predicting prices using datasets obtained from yahoo finance. However, with volatile changes in bitcoin prices and yahoo finance not keeping up with the updated values

in their datasets, we decided to switch to Quandl sourced datasets. Selecting a dataset is of utmost importance as said above and this decision aided in generating the results that we had been aiming for.

The training dataset ranges from the 20th of May 2014 until the 14th of January 2018. These values were selected because prior to 2014, the bitcoin price was very low to influence the prediction. From the late 2013 onwards, the price started increasing and by mid-May of 2014, it had been showing a trend in a stable price bracket. The testing dataset had a range from 15th of January, 2018 to 7th of November 2018. This is the period the author feels where the prices have been the most instable and volatile of all the years starting from 2014.

3.2.2 Building Neural Networks Model

Machine Learning is the most suitable technique which can be used here to predict Bitcoin prices. The model to be built had to achieve several goals in order to produce a near to accurate prediction. This included selecting the framework which could produce a good prediction accuracy, take in consideration of other parameters in its prediction algorithm and be trainable. Keeping these goals in mind, several different frameworks were tested and the author finally landed on using Keras, which is a neural networks API running on top of Tensorflow.

Keras was compared against Support Vector Machine. In comparison to SVM, Keras is less expensive in performance and computation. When we discuss about machine learning frameworks, depending on how we would want the learning to be, we can decide over the frameworks that can be used.

Next, there was the need to decide which layers had to be included and how many of them will be required along with deciding the epoch rates. The dataset used for training was standardized and transformed since this is more suitable for many activation functions.

3.3 Sentiment Analysis Process

Sentiment is commonly defined as a reaction to an event or thing. Every living organism expresses sentiments in one way or another. Human beings display reactions or sentiments in multiple ways such as facial expressions, change in tone, physical actions. This expression of sentiment is being used since a long time on different domains most common among them being products analysis and demand. Producers can tap into sentiments of the people to find out more about how their products are perceived and changes that could be made to make them more successful.

In a similar way, sentiment can be gauged as to how people react when there are tax cuts, changes in prices, stock prices changing. This sentiment can be analyzed to forecast or predict stock prices since if there is good news about a company performing well, people will be interested in buying stocks of that company or be a part of it so as to achieve success in the future. Similarly, in the event of a failure, people dis-associate themselves from poor performing companies.

The decision to use sentiment analysis was majorly to follow in the footsteps of existing works that had used sentiment analysis on twitter data with the help of different machine learning frameworks to analyze the sentiment of the data and build conclusions on it.

We will be covering multiple phases of using sentiment analysis apart from obtaining tweets from twitter and web scrapping data. We would be looking at how people perceive changes in government policies, economic changes, geopolitical factors to name a few.

3.3.1 Reducing noise from obtained data

As mentioned in section 3.1.2, content automatically obtained from online available datasets and API's has a lot of irrelevant data which is undesirable for analysis. This is because of the fact that such data can affect the results and produce a wrong outcome which could make the prediction inaccurate. In order to remove the impact of irrelevant data, python libraries were used to filter out such data.

The data was analyzed using regular expressions (Library in Python: 're') which removed all non-alphabetic symbols and retweet symbols (removing "RT", "@" & "#") that affected the analysis in terms of prediction negatively. The data was next checked for duplicity and any duplicate tweets found were removed. The data was also checked for tweets from bots which made irrelevant announcements or offered bitcoins and had been fetched with the remaining tweets. Manual analysis of the tweets was also performed in order to further weed out any tweets which had escaped the above filters. This phase was typically the cleaning phase since the data was cleaned and filtered.

The cleaning phase was also applied to data scrapped from different websites, news articles. However, it was noted by the author that web scrapped data was free of any hidden or unhidden bot messages and much easier to deal with. This was majorly because of certain libraries in python which only fetched data within targeted HTML tags.

3.3.2 Understanding Sentiment data

The next stage after thoroughly cleaning the data involved analyzing the data and deducing the sentiment associated with it. For this, Twitter API's provided a useful library in Python known as TextBlob. TextBlob was based on Naïve Bayes algorithm and easily predicted the sentiment scores of the data that was provided to it. The sentiment score was then compared with the cutoff values for classifying the tweets as positive, negative or neutral. Neutral tweets according to TextBlob were those which had equal positive & negative scores or could not be classified. The dataset was then updated with the sentiment score of each tweet after analysis.

However, it was found that TextBlob was unable to deduce the sentiment scores for tweets which could have been easily classified as positive or negative manually. This was also correlated with observation of more than 50% of tweet data being classified as neutral by TextBlob and only a small percentage of data was classified as positive or negative. This impacted the prediction scores as well. Observing the shortcomings of TextBlob library in predicting sentiments, a need was realized to have a better and more effective method for classifying tweets.

3.3.3 Developing Sentiment Analysis model

The amount of data that had been collected was difficult to be manually classified and TextBlob was also not performing well. Hence, machine learning techniques were used to classify the data as discussed in section 3.4.2. The data was classified into positive, negative and neutral sentiment and the entire classification process was supervised learning.

Due to the shortcomings of TextBlob library on classifying certain data, a bag of words technique was used and for this purpose, Keras was used to train the dataset in order to create a Bag of Words and testing dataset built from gathering data from Twitter, different news sources were tested. Compared to TextBlob library, the above technique yielded better results.

3.4 Text Mining Data

Web scraping or data scraping from web or text mining is a technique used for extracting data from websites using either web crawlers built personally or by web crawling service providers or by means of API's provided by websites. Some notable services providing web-scraping tools are Amazon, Google, Webhose.io etc.

Apart from fetching and classifying tweets obtained from Twitter API, the author realized a need to obtain data from sources other than twitter in order to improve the prediction algorithm and delve more deeper into finding the reasons for bitcoin's volatile price. For this purpose, it was decided to web scrape data from news articles which would serve as additional support apart from twitter obtained data. This web scraping of data was performed by writing code in Python and automating the feature of reading text from websites and running prediction algorithms on them.

3.4.1 News Articles

In order to have a much better quality of analysis data, news articles from different news outlets and countries were scraped legally using both personally built web crawlers as well as third party web scrapping tools. Some of the few websites from which news articles were obtained included Forbes & Reuters. These news articles written with opinions for and against were helpful in understanding sentiments of investors towards daily events such as volatile price changes of Bitcoin. Scraping helped in obtaining news articles which spoke about how business institutions were reacting to bitcoin and planning their moves in order to bring in more profits and clients interested in buying and dealing with bitcoins. Policy decisions made by governments worldwide to respond to growing popularity of bitcoin was also discusses extensively in multiple news articles.

Web scraping was done in Python using libraries such as BeautifulSoup, urllib, Requests which made the work of fetching the data easier since there was no cleaning of the fetched data involved here as discussed above in 3.3.1. BeautifulSoup, the library used here is useful for pulling data from HTML tags. Webpages have their textual data stored in “p” html tags which when passed to the beautiful soup parsing element fetches data within those tags filtering other garbage data such as advertisements or spam messages.

3.4.2 Reactions on Interest Rates, Foreign Exchange and US Dollar

As discussed above, web scraping data was helpful in understanding people’s perceptions apart from those obtained using twitter. News articles carried more opinions and clearly spoke of their apprehensions or incline towards bitcoin and other cryptocurrencies in focus.

It was exciting to come across news articles which spoke about topics that had a direct but hidden impact on Bitcoin prices. Primarily among these were the ones which discussed about recent policy changes implemented by the US government such as tax cuts and increase in federal interest rates. This news had a positive impact on investors and trader’s world over. However, with a strong US dollar other global currency have taken a hit and tumbled down.

3.5 Time Series Analysis

Time series analysis is a form of forecasting where sequential data is taken over a defined interval of time and analyzed. This analysis can be of predicting the next day price, or prices 10 days down the line. It is used to find a correlation or a trend among several data points in order to make a forecast for the future.

For predicting next day bitcoin prices, time series analysis plays an important role along with sentiment analysis and supervised learning techniques. One of the techniques of time analysis used here is Moving Average. Moving average calculates the average of a subset of data over a sliding window size and shifts forward instead of calculating the average of the entire dataset. This technique helps in easing out short term fluctuations that are present and is helpful in identifying long term trends & patterns.

One of the key mentions of time series analysis is the ARIMA model. ARIMA is the acronym for Autoregressive Integrated Moving Average. It is made up of three parts, AutoRegression model which uses the relationship between datapoints and calculates the difference between current datapoint and previous datapoint (Integrated) in order to make the series stationary. The moving average simply uses the dependency between datapoints and residual errors to find the lagged observations. It is a linear regression model.

However, we do see some discrepancies with using time series analysis. Foremost is that it does not use current data. This is attributed to the reason that since there is averaging of data points performed based on a rolling window size, a different result is obtained which may or may not be equal to the current data.

3.6 Conclusion

The method followed for research has been outlined in the above sections. The philosophy being highlighted through this study is that of positivity in the light of Bitcoin prices fluctuating by huge margins and investor confidence eroding with every passing day. The experiments use a machine learning approach to create results on which analysis can be performed and a near to accurate prediction can be obtained. The approach used to capture the data and datasets used in the process has been highlighted.

4. Working

4.1 Introduction

This chapter documents the programming languages and techniques that were used to conduct this research. The language used for writing code is primarily Python and PyCharm Integrated Development Environment was used for testing the code.

4.2 Dataset Creation

In this section, I cover the techniques by which datasets were obtained and created. There are multiple sources online for obtaining data. Prominent ones include Kaggle, Yahoo, third party paid services. For obtaining financial data related to stocks and in our case Bitcoin, we direct our attention towards providers like Yahoo Finance, Bloomberg, NASDAQ, CoinDesk, Quandl to name a few.

The datasets used for this research were obtained from CoinDesk, Yahoo Finance, Quandl, Kraken, Coinbase and BitStamp. A huge collection of Bitcoin data in order to ensure that any missing data in one of the datasets could be filled in with relative values. Bitcoin data from Yahoo Finance was obtained manually by saving the prices for required dates in a .csv file. This was due to the absence of libraries in Python which performed rest calls to fetch data directly from Yahoo Finance. These libraries had been discontinued since early 2017 by concerned developers.

The other source for Bitcoin data was from CoinDesk and Quandl. Quandl is a great source for obtaining financial data from different companies. We import the python libraries that would be required.

```
1. import quandl
2. import pandas
3. import numpy
4. import time
5. from matplotlib import pyplot as plt
6. import csv
```

The user first authenticates their access key when trying to obtaining data from Quandl. Once authentication is successful, the user can perform REST API calls to obtain the data.

```

1. quandl.ApiConfig.api_key = ''
2. data = ['BCHARTS/KRAKENUSD', 'BCHARTS/COINBASEUSD', 'BCHARTS/BITSTAMPUSD']
3. for i in range(index):
4.     train_dataframes[i] = quandl.get(data[i], start_date = '2014-05-
    20', end_date = '2018-01-14', returns=pandas)
5. for j in range(index):
6.     test_dataframes[j] = quandl.get(data[j], start_date='2018-01-
    15', end_date='2018-11-29', returns=pandas)

```

The data from three different bitcoin exchanges were obtained and then merged into a single dataset. The dataset also had additional columns which contained average of values from the above three data sources.

```

1. test_Open = merge_dataframes(list(test_datframes.values()), list(test_datframes.keys()), 'Open')
2. test_Open.replace(0, numpy.nan, inplace=True)
3. test_Open = test_Open.mean(axis=1)
4.
5. btc_test_dataframes = merge_dataframes(list(test_datframes.values()), list(test_datframes.keys()), 'Close')
6. btc_test_dataframes.replace(0, numpy.nan, inplace=True)
7. btc_test_dataframes['Avg Close'] = btc_test_dataframes.mean(axis=1)
8. btc_test_dataframes['Avg Open'] = test_Open

```

4.3 Prediction of Bitcoin Price

4.3.1 Keras

After having obtained the data from sources like Quandl, CoinDesk and Yahoo Finance, the next step was to train the data in order to build a prediction model. For developing a prediction model and deciding on the techniques for doing so, I analyzed the best approach which could be used. From different Deep Learning techniques, Keras and TensorFlow were selected for use here to develop a prediction model.

As mentioned in section 1.2.2, Keras is a high-level neural networks API written in Python. Since my coding ground was Python during the dataset creation, it was decided to move ahead with using Python. Python has a Keras library that can be used for working and developing Keras based programs.

```

1. import numpy as np
2. import pandas as pd
3. import requests
4. from sklearn.preprocessing import MinMaxScaler
5.
6. from keras.models import Sequential
7. from keras.layers import Dense
8. from keras.layers import LSTM
9. from matplotlib import pyplot as plt
10. import datetime as dt

```

Next, we read the data from the training dataset that is created and store it. The data is scaled and reshaped so that it can be used without any loss of precision for performing mathematical computations by Keras.

```
1. training_set = pd.read_csv('btc_train_data.csv') # reading csv file
2. scale = MinMaxScaler()
3. x_training = scale.fit_transform(training_set.iloc[:, 5:6])
4. y_training = scale.fit_transform(training_set.iloc[:, 4:5])
5.
6. '''x_training = scale.fit_transform(training_set.iloc[:, 1:2])
7. y_training = scale.fit_transform(training_set.iloc[:, 4:5])'''
8.
9. x_training = x_training.reshape((-1, 1, 1))
```

The data is then trained by using neural networks. I have used one Long-Short Term Memory Layer and three Dense layers with varying number of neurons. The number of neurons used for the training dataset was based on the results obtained after training the dataset with eight, sixteen and thirty-two neurons. It was observed that if more than sixteen neurons were used, the resultant data was overfitting. Hence, it was decided to use sixteen neurons maximum for any of the layers. A lstm layer was used keeping the amount of training data present and its usability in modelling long term data. The activation function used here is Rectified Linear Unit since it provided consistent results. This is because it is less expensive computationally compared to sigmoid function.

```
1. regressor = Sequential()
2. regressor.add(LSTM(units=4, activation='relu', input_shape=(None, 1)))
3. regressor.add(Dense(8, activation="relu"))
4. regressor.add(Dense(16, activation="relu"))
5. regressor.add(Dense(8, activation="relu"))
6. regressor.add(Dense(units=1))
7.
8. regressor.compile(optimizer='adam', loss='mean_squared_error')
```

Selecting a viable activation function is important since it will be helpful to approximate the function faster leading to faster training process (Sharma, 2017). A sigmoid function was not used here keeping in consideration the requirement of achieving precision.


```

1. predicted_price = regressor.predict(inputs)
2. predicted_price = scale.inverse_transform(predicted_price)
3. print('Predicted Price : ', predicted_price[:-1])
4. # Result on MatPlot
5. plt.plot(real_stock_price, color='blue', label='Real BTC Value' , marker = '.')

6. plt.plot(predicted_price, color='red', label='Predicted BTC Value', linestyle =
   'dashed')
7. plt.title('BTC Value Prediction')
8. plt.xlabel('Days')
9. plt.ylabel('BTC Value')
10. plt.legend()
11. #plt.show()
12. today = pd.Timestamp('today')
13. plt.savefig('kp._{:d%m%Y}.png'.format(today))
14.
15. today_open = predicted_price[-1]
16. print("Today Open Value : " , today_open)

```

Keras provides an inbuilt prediction function that automatically performs required computations based on the regressor model provided to it. I have used matplotlib library to display the result and further saved it as a .png file as well for image-based classification.

4.3.2 TensorFlow

TensorFlow is another open source library for high performance computations. A simple program was written using TensorFlow for the purpose of predicting. The central unit of data here is a tensor. It consists of a set of primitive values that are shaped into an array of any number of dimensions (TensorFlow, 2018). To evaluate tensors, a session is created. A session encapsulates the state of the TensorFlow runtime and runs its operations. We use placeholders which are inputs provided to the graph. Layers are used to package both the variables and the operations that act on them (TensorFlow, 2018).

```

1. def tensorflowModel():
2.     # Setting up the variable
3.     n_stocks = training_data.shape[1]
4.     n_neurons1 = 256
5.     n_neurons2 = 256
6.     #Setting the PlaceHolders
7.     X = tf.placeholder(tf.float32, shape=[-1, n_stocks])
8.     Y = tf.placeholder(tf.float32, shape=[-1, 1])

```

Initially, I started with having 4 set of neurons and 4 hidden layers for training the data. However, there was significant overfitting observed because of which the set of neurons was reduced to 2 and the hidden layers were also reduced to 2.

```

1.     sigma = 1
2.     weight_initializer = tf.variance_scaling_initializer(mode="fan_avg", distrib.
    ution="uniform", scale=sigma)
3.     bias_initializer = tf.zeros_initializer()
4.
5.     # Layer
6.     W_hidden1 = tf.Variable(weight_initializer([n_stocks, n_neurons1]))
7.     bias_hidden1 = tf.Variable(bias_initializer([n_neurons1]))
8.
9.     # Layer 2
10.    W_hidden2 = tf.Variable(weight_initializer([n_neurons1, n_neurons2]))
11.    bias_hidden2 = tf.Variable(bias_initializer([n_neurons2]))
12.
13.    # Output layer: Variables for output weights and biases
14.    W_out = tf.Variable(weight_initializer([n_neurons2, 1]))
15.    bias_out = tf.Variable(bias_initializer([1]))
16.
17.    # Hidden layer
18.    hidden_1 = tf.nn.relu(tf.add(tf.matmul(X, W_hidden1), bias_hidden1))
19.    hidden_2 = tf.nn.relu(tf.add(tf.matmul(hidden_1, W_hidden2), bias_hidden2))

```

Unlike Keras where the majority of operations are performed by it since it works on a TensorFlow backend, here we define these operations and can actually observe the working as well. TensorFlow provides optimizers that use standard optimizing functions such as adam, rmsprop and most commonly, gradient descent to optimize and understand the trained data.

```

1.     output = tf.transpose(tf.add(tf.matmul(hidden_2, W_out), bias_out))
2.     mse = tf.reduce_mean(tf.squared_difference(output, Y))
3.     opt = tf.train.AdamOptimizer().minimize(mse)
4.
5.     # Make Session
6.     net = tf.InteractiveSession()
7.
8.     # Run initializer
9.     net.run(tf.global_variables_initializer())

```

After setting the number of epochs for which the data has to be trained and the batch size, we run the prediction function and then test it on the testing dataset.

4.4 Time Series Analysis on Data

This section details the time series analysis that was performed on the dataset created above. Multiple window sizes were implemented in order to obtain an accurate result. There was also an attempt to average the results of different window sizes into one and obtain the final result.

```

1. def normal_rolling(data):
2.     df = data.sort_values('Date', ascending=True)
3.     open_pred = data[['Avg Open']]

```

```

4.     plot_open25 = open_pred.rolling(25).mean()    #for window size =100, 200, a huge
        difference in the moving average and actual open price was observed.
5.     write_to("Rolling Window : 25 ",plot_open25)
6.     print("Rolling Window : 25 ")
7.     print(plot_open25)
8.     print("Rolling Window : 15 ")
9.     plot_open15 = open_pred.rolling(15).mean()
10.    write_to("Rolling Window : 15 ", plot_open15)
11.    print("Rolling Window : 50 ")
12.    plot_open50 = open_pred.rolling(50).mean()
13.    write_to("Rolling Window : 50 ", plot_open50)
14.    print(plot_open50)
15.    print("Rolling Window : 100 ")
16.    open_pred100 = open_pred.rolling(100).mean()
17.    write_to("Rolling Window : 100 ", open_pred100)
18.    print(open_pred100)
19.    print("Rolling Window : Average ")
20.    avg_open_pred = (plot_open25+plot_open15+plot_open50+open_pred100)/4
21.    write_to("Rolling Window : Average ", avg_open_pred)
22.    print(avg_open_pred)
23.
24.    plt.plot(df['Date'],df['Avg Open'], label='Open Value')    #plot_open
25.    plt.plot(df['Date'],avg_open_pred, label='Rolling Mean Open Prediction')
26.    plt.xlabel('Timeline')
27.    plt.legend()
28.    plt.show()

```

The data was a pandas dataframe which was capable of being used for computations as required by Python.

Pandas is a library in Python which can be used for performing many mathematical and file operations.

Functions like reading a csv file, text files as well as certain mathematical operations are easily performed using pandas.

```

1. def adj_rolling(data):
2.     return pd.DataFrame({"Date": data["Date"] ,
3.                          "Avg Open": data["Avg Open"],
4.                          "Avg Close": data["Avg Close"]})
5.     '''return pd.DataFrame({"Date": data["Date"] ,
6.                             "Avg Open": data["Avg Open"] * data["Adj Close"] / data["Cl
7.                             ose"],
8.                             "High": data["High"] * data["Adj Close"] / data["Close"],
9.                             "Low": data["Low"] * data["Adj Close"] / data["Close"],
                             "Close": data["Adj Close"]}) '''

```

4.5 Web Scrapping Data

This section discusses about the method used for web-scrapping or obtaining data from different news blogs and testing them in order to find their polarity and use it for developing an improved prediction model. For obtaining data from websites, we use BeautifulSoup which is a Python library.

```

1. from bs4 import BeautifulSoup as soup
2. from urllib2 import urlopen

```

Beautiful Soup is used to obtain the data on the web urls that are passed to it as variable. Based on the HTML tag which is required by the user, BeautifulSoup filters out the text within the required tags and provides the result to the user. Sometimes it so happens that the text obtained may have different text encoding. In such situations, it becomes necessary to change the text encoding in order to make it readable.

```

1. plot_data =[]
2. class web_Scraper:
3.     def __init__(self,url):
4.         self.url = url
5.         global plot_data
6.
7.     def data_collection(self):
8.         uClient = urlopen(self.url)
9.         u_page = uClient.read()
10.        uClient.close()
11.        page = soup(u_page, "html.parser")
12.        pg = page.div
13.        data = []
14.        new_text = []
15.        for t in page.find_all('p'):
16.            data.append(t.text.encode("utf-8"))
17.        data = [str(i).decode('unicode-
escape').encode("ascii", "ignore") for i in data]
18.        print(data)

```

The data I was aiming to obtain from websites legally was placed within paragraph HTML tag or <p> tag as it is called using HTML notation on the concerned webpages. Having this information, the beautiful soup.find_all method was provided with the <p> tag as filter parameter so that all the text within the <p> tag from the webpage would be scrapped and any other text will be discarded or rejected. Apart from using HTML tags for scraping data, we can also make use of the class-name present in the HTML source code of the webpage for fetching text. BeautifulSoup further provides this functionality as well.

```

1. import requests
2. from bs4 import BeautifulSoup
3. url = "http://search.yahoo.com/search?p=%s"
4. query = "bitcoin"
5. r = requests.get(url % query)
6. soup = BeautifulSoup(r.text,"lxml")
7. soup.find_all(attrs={"class": "yscht1"})
8.
9. for link in soup.find_all(attrs={"class": "yscht1"}):
10.    print "%s (%s)" %(link.text, link.get('href'))

```

A different technique to obtain data from blogs and news articles was also attempted. This technique involved taking a screenshot of webpages that were difficult to access due to the failure of BeautifulSoup to return required text data. Such web sites had multiple advertisements on their page that loaded once the page was opened due to which BeautifulSoup was unable to fetch data. Selenium, a library used in Python and with support extended to all major web browsers was used to take screenshots of web pages. The web browser used here was Google Chrome and hence you will see a Google Chrome web-driver being used here.

```
1. #taking screenshot of twitter tweet webpage
2. from selenium import webdriver
3. Driver = "chromedriver"
4. drive = webdriver.Chrome(Driver)
5. drive.get('http://money.cnn.com/2018/06/21/investing/trade-war-stocks-trump-
   goldman-sachs/index.html')
6. screenshot = drive.save_screenshot('article1.png')
7. drive.quit()
```

After taking the screenshot, it was essential to read the text from the image. In order to do this, Python Imaging Library or PIL was used. PyTesseract, a most commonly used library that can read text from an image i.e. perform optical character recognition was also used along-with PIL. (Rosebrock, 2017).

```
1. from PIL import Image
2. from pytesseract import image_to_string
3. data = (image_to_string(Image.open('testing1.png'), lang='eng'))
4. print(data)
```

The textual data obtained was next analyzed for the sentiment or polarity which could be deducted from the text. TextBlob, a library frequently used for processing textual data was used in order to perform analysis on the data obtained from web-pages and provide a sentiment score for the same. It can classify text as positive, negative and neutral based on the polarity score discussed above. This polarity score is calculated based on a dictionary of words that are already trained by a Naïve Bayes classifier that forms the main part of TextBlob.

```
1. from textblob import TextBlob
2. analysis = TextBlob(data)
3. if analysis.sentiment.polarity > 0:
4.     print('positive')
5. elif analysis.sentiment.polarity == 0:
6.     print('neutral')
7. else:
8.     print('negative')
```

An example of a web scraped news article that has been analyzed by using textblob library. The library analyzes the entire text in order to find the total sentiment and polarity conveyed.

```
1. ['Bitcoin is back in the news as its price continues a meteoric rise, with the cost
   of a single coin breaking $10,000 yesterday. Today, it was reported by several maj
   or publications that NASDAQ, the second largest stock exchange in the world, plans
   to launch futures contracts for Bitcoin next year. That would allow investors to be
   t on the rise and fall of the cryptocurrency, profiting if they are correct about t
   he direction of its future price. ', 'NASDAQ is hardly the first player from the tr
   aditional finance world to jump on the Bitcoin bandwagon. At the start of this mont
   h, Chicagos CME Group said it would begin providing futures contracts on bitcoin as
   well. As The Verge reported this morning, while many.....'Predicting exactly when
   the Bitcoin bubble will burst, as it has many times before, is impossible. But pret
   ty soon, the worlds most venerable exchanges will let you start placing bets.', 'Co
   mmand Line delivers daily updates from the near-future.']]
2. Positive
3. ('Inside', [0.03866522366522367])
4. ('List : ', [0.03866522366522367])
```

4.6 Sentiment Analysis

In this section, I will discuss about the techniques used for performing sentiment analysis. I have used Twitter provided API's which fetch tweets from Twitter's database after authenticating the user. Tweepy is the library that provides access to tweets using the twitter api. The first step towards using this library for our purpose is to authenticate the user with the provided consumer and access keys.

```
1. def __init__(self):
2.     # keys and tokens from the Twitter Dev Console
3.     consumer_key = ''
4.     consumer_secret = ''
5.     access_token = ''
6.     access_token_secret = ''
7.
8.     # attempt authentication
9.     try:
10.        # create OAuthHandler object
11.        self.auth = OAuthHandler(consumer_key, consumer_secret)
12.        # set access token and secret
13.        self.auth.set_access_token(access_token, access_token_secret)
14.        # create tweepy API object to fetch tweets
15.        self.api = tweepy.API(self.auth)
16.    except:
17.        print("Error: Authentication Failed")
```

Once the user is successfully authenticated, we can proceed with passing arguments to the tweepy functions which perform the api calls to twitter and return the resultant tweets (Kumar).

```

1. def get_tweets(self, query, count):
2.     tweets = []          # empty list to store parsed tweets
3.     try:
4.         # call twitter api to fetch tweets
5.         fetched_tweets = self.api.search(q=query, count=count)
6.         count = 0
7.         # parsing tweets one by one
8.         for tweet in fetched_tweets:
9.             parsed_tweet = {}
10.
11.            # saving text of tweet
12.            parsed_tweet['text'] = tweet.text
13.            #print(type(parsed_tweet['text']))
14.
15.            # saving sentiment of tweet
16.            parsed_tweet['sentiment'] = self.get_tweet_sentiment(tweet.text)
17.            #print(type(parsed_tweet['sentiment']))
18.
19.            #writing tweet to file
20.            count = count+1
21.            self.writeToFile( count, (tweet.text).encode('utf-
22.            8') , parsed_tweet['sentiment'] )
23.
24.            # appending parsed tweet to tweets list
25.            if tweet.retweet_count > 0:
26.                # if tweet has retweets, ensure that it is appended only once
27.                if parsed_tweet not in tweets:
28.                    tweets.append(parsed_tweet)
29.            else:
30.                tweets.append(parsed_tweet)

```

The required tweets were fetched and saved to a .csv file. These tweets have a lot of unnecessary letters and symbols like '@', '#', numbers which need to be removed before the tweets can be analyzed. Python regular expressions were used to perform this function and all such letters and numbers were removed. The next thing to take care was of removing duplicate and garbage tweets from the fetched tweets. There was a high presence of duplicate and garbage tweets which if analyzed with the rest of the tweets impacted the sentiment greatly and distorted the results. Hence, fetched tweets were also analyzed to remove any duplicate tweets and unnecessary tweets.

```

1. def clean_tweets(filename):
2.     f = open(filename, 'r+')
3.     for line in f.readlines():
4.         #line = re.sub('[RT]', '', line)
5.         line = re.sub(r'[RT]+ @[A-Za-z0-9]+', '', line)
6.         line = re.sub('[:]', '', line)
7.         write_to_file(line.strip())
8.
9. def write_to_file(text):
10.    f = open('TweetsUp.txt', 'a')
11.    f.writelines(text + "\n")
12.    f.close()
13.
14. def remove_dups(filename):

```

```

15. dup = open(filename, 'r+')
16. newline = []
17. op = open('TweetsDup.txt', 'a')
18. for line in dup.readlines():
19.     if line not in newline:
20.         newline.append(line)
21.         op.write(line+"\n")

```

After running the fetched tweets through the cleanup code above, most of the unnecessary letters and numbers were removed. The following results were observed after this process.

```

1. The Blockchain could help make IPs smarter and easier! Allowing for records to be s
   tored on a secure, immutable bl... https://t.co/wWpmxDjVzS
2.
3. markfidelman Crypto scams and how to avoid them bitcoin cryptocurrency finance mon
   ey credit... https://t.co/yYKHov6y7o
4. Crypto scams and how to avoid them bitcoin cryptocurrency finance money credit http
   s//t.co/41x8PcSHvH
5.
6. I wonder what this potential solution to "blockchains problems" will entail. Cryptoc
   urrency certainly has potential... https://t.co/PFGCJil47m
7. Ax1 's Product Director Mark Shaw - Great meeting at Innosilicon - today looking fo
   rward to building a long term relationshi...
8.
9. Less than 2 days days left - Buy into AX1's ICO before you miss out
10. ax1mining ax1 cryptocurrency aheadofthegame tothemo...
11.
12.
13. In UK, a BMW Car dealership is now accepting Bitcoin as a payment option.
14. https://t.co/o6QYFkblp6
15.
16. Well what do we have here a conflict of Interest when it comes to bitcoin $BTC & Cr
   ytpocurrency CryptoCongress https://t.co/mb9rcQNd7V
17. Coinbase speculating on possible new assets..https://t.co/jd45czPpDU
18.
19. Ax1 is leaving China after a productive trip to meet with miner manufacturers - st
   ay tuned for updates coming next week
20.
21. Head and shoulders pattern on Bitcoin... could this finally be the bottom? bitcoin
   trading money cryptocurrency... https://t.co/WvFtrqf3VU
22. Digibyte on bitbox of the START on cryptocurrency exchange amount bitcoin Ethereum.
   Gaining more Adoption https...
23. The Fight Over Mt. Gox's Bitcoin Stash https://t.co/F0q0FH0Zgv MTGOX blockchain cryt
   pocurrency finTech law... https://t.co/7Ea0l0hXr9
24. We @ShemBooth1 & angry dentist do another bitcoin btc cryptocurrency show on @World
   CryptoNet - what would ulilke... https://t.co/KAhYUHVFTw
25. Ax1 is leaving China after a productive trip to meet with miner manufacturers - st
   ay tuned for updates coming next... https://t.co/taTzRyfo4G
26. AX1 set to purchase the most efficient miners - read about our trip to China
27. "CRYPTOCURRENCY BACKED BY HABANERO PEPPERS IS A THING IN MEXICO " ... A cryptocurren
   cy called Agrocoin is giving buyers a...

```

Below we can see garbage tweets and duplicate tweets that could not be filtered by using both the code written above as well as by a web-based filtering service provided by Twitter along-with their sentiment score. Inability to remove this garbage data affects the overall polarity score of the fetched tweets and

produces wrong results. The sentiments for some of these do not seem to match with the context of the tweet.

This has been observed frequently with tweets related to bitcoin and it is speculated that due to the volatility in the price of bitcoin, investors are reacting slowly than expected.

1. Hey everyone, this post about bitcoin PoW mining in r/environment is shedding other cryptos like Ethereum in a nega... <https://t.co/Gj3j82MFkH> , negative
2. Hawk or dove? Bitcoin is forcing central banks to take sides <https://t.co/pIuDADGgve> , neutral
3. RT @itsCryptoKermit: when you bought bitcoin at \$20k and it keeps dropping <https://t.co/4K8KMAvNwF> , neutral
4. Bisq adds 7 new payment methods to boost its US bitcoin trade <https://t.co/oOfrBHaeoJ> , positive
5. \$BARZ Updates <https://t.co/9N0GqXKnCP> #wsj #nytimes #reuters #bloomberg #thetstreet #jimmyfallon #forbes #nasdaq... <https://t.co/i1Wth8mwe7> , neutral
6. RT @BitfuryGeorge: Once the realisation of what Lightning will bring sinks in, we should see Bitcoin going to \$100,000 real fast ... , positive
7. News bitcoin daily! , neutral
8. RT @itsCryptoKermit: when you bought bitcoin at \$20k and it keeps dropping <https://t.co/4K8KMAvNwF> , neutral
9. RT @earthtokens: VIDEO: #EARTH #Token & the #Natural #Asset #Exchange introduction <https://t.co/yQQYX3h7rn> #ico #ethereum #bitcoin #crypto... , positive
10. RT @masterblack1406: <https://t.co/qh3INKBRTL> | Minería En La Nube De Bitcoin, Ethereum, Unit, Dolar: <https://t.co/L53h0jzwdi> vía @YouTube , neutral
11. RT @bigstonebtc: Segwit txの比率が25%にアップ。(CoinbaseとBitcoin core 0.16.0効果か?)
12. #ICYMI : #Crypto Markets Climb Towards Green, #Bitcoin Above \$10,000 <https://t.co/ZnX2xoBvic> #BTC #ETH #Crypto <https://t.co/QEoQExpb1H> , negative
13. RT @Crypticsup: The European Commission Announces Readiness to Regulate Cryptocurrencies

5. Findings and Analysis

5.1 Results from Bitcoin Prediction Process

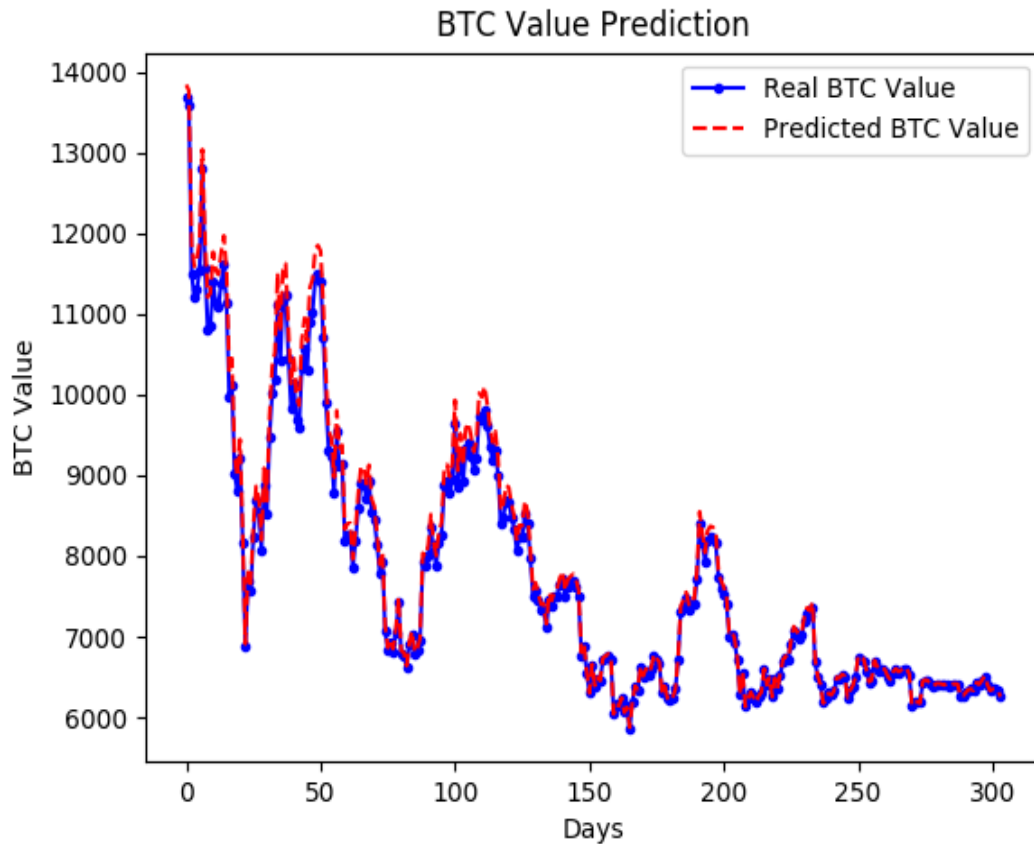


Figure 4 : Bitcoin Price Prediction (as of 11/15/2018)

From the Figure above, we can see that the predicted prices models on the lines of the real Bitcoin price. We have achieved a near to accurate price prediction using our algorithms which can be helpful in deciding whether to buy bitcoins or not. From the graph, we can observe the volatility of Bitcoin prices in the last 1 year itself since the sample testing space for this research was from November 2017 to November 2018.

As discussed above, this volatility in bitcoin prices has been due to several factors which we will be further discussing in detail below. We can attribute this fall of Bitcoin price due to the market sentiments which haven't been positive towards Bitcoin due to federal governments not readily accepting it. This coupled with stronger global currency indices may have been a reason in its fall.

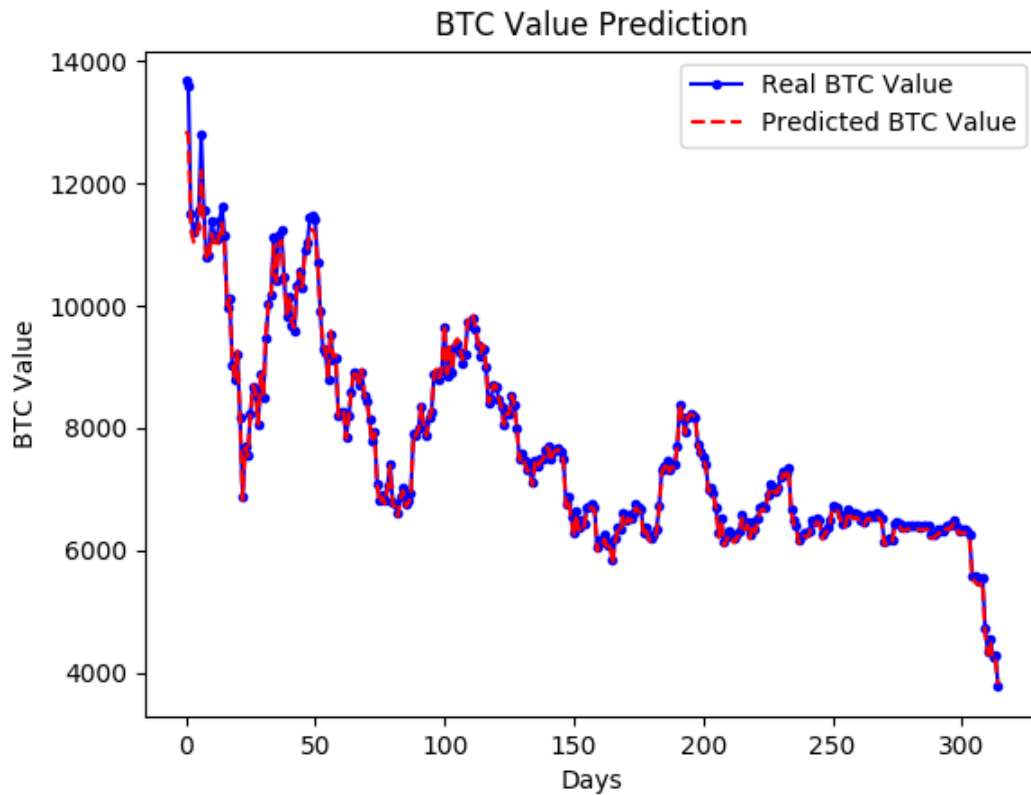


Figure 5: Bitcoin Price Prediction (as of 11/26/2018)

The above figure depicts the current Bitcoin price prediction as of 26th November 2018. During the last week of November 2018, Bitcoin prices slid down to all time low in the last 1 year. There are no specific reasons or events that point to the cause in the fall of Bitcoin unlike for stock markets, where a reason for the meltdown in a company's stock can be pinpointed to some frantic event. One reason attributed here is price correction of Bitcoin and other digital currencies which have brought down the exchange price by 10%.

(Saahil, 2015) in his paper on prediction stock market direction using Support Vector Machine discusses about the irregularities observed while using Support vector machine for predicting stock prices. The author stated that for short term results the prediction was similar to random guessing. However, for long term results it was observed that there would be more macro conditions related to the markets which would be helpful in giving strength to the prediction.

5.2 Sentiment Data Results

As discussed in 3.3, a variety of experiments were performed using Sentiment Analysis for predicting Bitcoin Prices which used Twitter data as the major source. However, there were discrepancies in the prediction of public sentiment due to the fact that given the volatility of Bitcoin Prices, public reaction was not quick and adequate enough to aid in predicting sentiments and correlating the results with results obtained from predicting Bitcoin prices using Machine Learning.

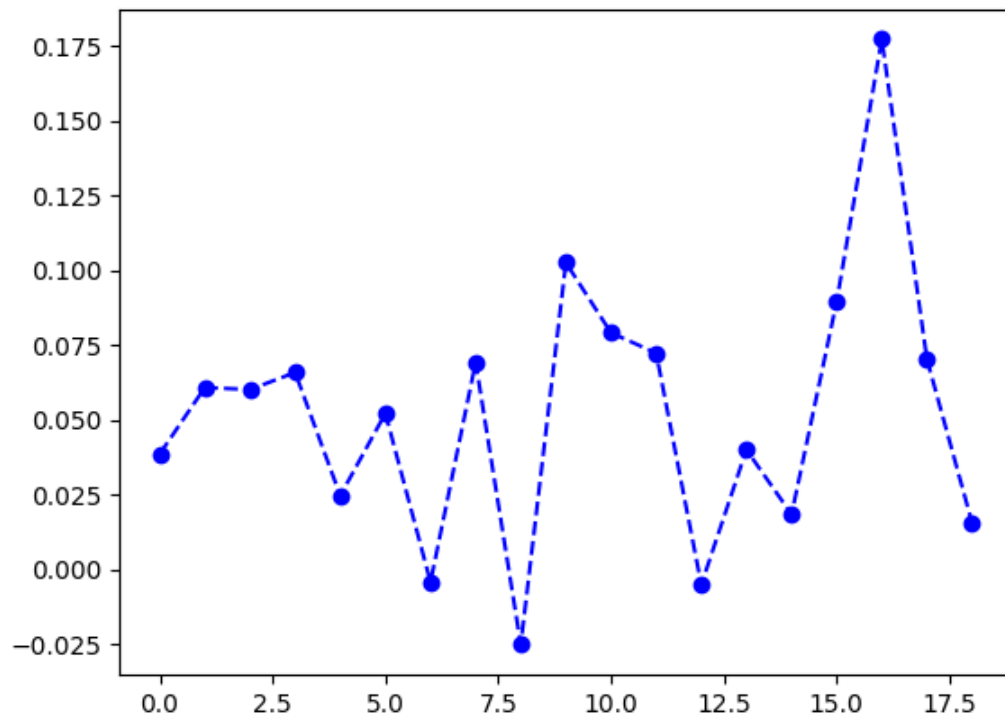


Figure 6 : Sentiment Data obtained from Text Mining

The figure above depicts results obtained from a sample set of web scrapped data. This web scrapped data consisted of news articles and blogs on Bitcoin from famous and widely read sources.

(McAteer, 2014) in his dissertation had evaluated a wide range of factors for analyzing correlation between Bitcoin and Twitter. The first approach was to compare volume of bitcoin transactions to that of volume of tweets that were posted over a duration of time. The results observed here highlighted greater dependence on using volume of tweets as a measure of the popularity of Bitcoin. However, I noted that there was a significant

difference in the environment during the time the research was performed and in the current situation in terms of increased volatility in the price of bitcoin and introduction of bitcoin futures which bore a significant impact as mentioned in a report by (Hale, Krishnamurthy, Kudlyak, & Shultz, 2018) for Federal Reserve Bank for San Francisco.

(Antweiler & Frank, 2004) have shown how message boards, online forums and Twitter can be used to predict market movements. For traditional stocks such as that of Apple, Facebook, Microsoft etc. we can agree that digital sources have a clear influence on investment decisions. However, it has been observed that bitcoin is not similar to a traditional stock with no regulation & oversight over its price fluctuations. The concept of insider trading is absent for bitcoin, however since bitcoins are mined, individuals having or obtaining a huge cache of bitcoins can rule the markets and lead it to an outcome that would suit them well. In this situation, sentiments against the bitcoin would have beneficial results for such entities and not for the remaining investors. A lop-sided scenario like this presents more challenges to the legality of bitcoin and calls for measures that would help it be sentiment influence free.

5.3 Time Series Results

As discussed in section 3.5, Time series analysis takes into account the rolling mean of a set of datapoints within a rolling window. Here, we analyzed it for two window sizes of 25 and 50. A rolling window of size n takes into account n days which are averaged, and a corresponding value is obtained and used. In this manner, the moving average shifts forward.

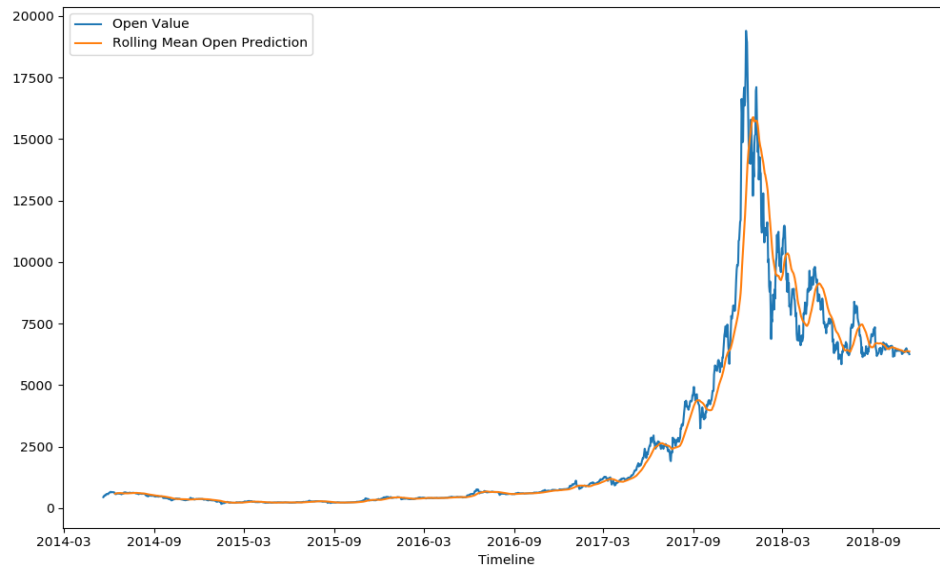


Figure 7 : Time Series Result (window size = 25)

The above figure displays the result obtained by applying a rolling window of size 25 on the dataset. We can see the relative proximity of the Rolling Mean Open Prediction with the actual Open Value. This result can be correlated with the prediction results obtained in Figure 5 to have a greater decision-making ability on predicting Bitcoin prices. However, there might be some loss of precision since the moving average starts with the first data point which is 25 after the actual data point in the dataset. We can see this in the figure above during the period between March 2014 and September 2014 when there is a difference in the points when the actual Open value starts, and the Rolling Mean Open Prediction starts.

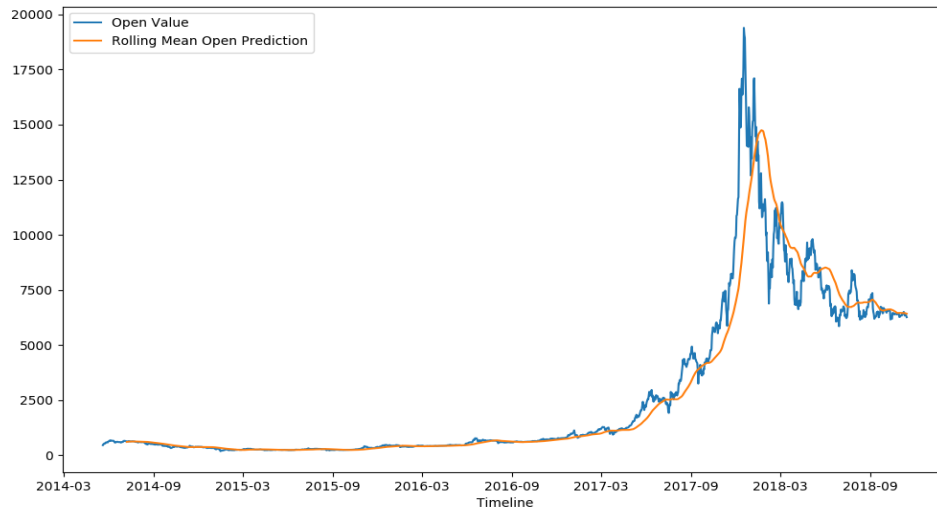


Figure 8 : Time Series Result (window size = 50)

The above figure displays the Time Series result obtained by using a Rolling Window of size 50. This essentially means that the moving average of 50 datapoints was calculated and hence we can observe the delayed start in the first datapoint of the actual Open value and the Rolling Mean Open Prediction.

When we compare the results of Figure 7 and Figure 8, we observe that for window size 25 there is a pattern better visible compared to that of window size 50. From this we can draw an inference that a smaller window size will aid in better forecasting since it has more precision and less number of lagged datapoints.

5.4 Impact of US Dollar & Federal Interest rates on Bitcoin

The reason for studying the impact of US dollar and other global currencies on Bitcoin is primarily because fiat currencies like US Dollar, Euro are used as the medium for purchasing Bitcoins since there are no other means of purchasing digital currencies apart from mining which has been discussed in 1.2.1.

As discussed earlier, since the last few years fiat currencies were not performing well globally. This led to rise of digital currencies like Bitcoins, Ethereum, Litecoin etc. However, since the past year there has been improving signs of stability in fiat currencies due to the long term measures which were set in place after

2008 and have been yielding results. This has led to a shift of focus back to US dollar and other global currencies from Bitcoin which was ruling the roost in late 2017.

5.4.1 Bitcoin & US Dollar



Figure 9 : US Dollar vs Bitcoin (YTD)

From the figure above, we see the performance of US Dollar and Bitcoin over the period of November 2017 to November 2018. The figure indicates a percentage of 7.53 for US Dollar which is the percentage of rise in US Dollar index price over this period compared to a fall of Bitcoin index price over the same period of time by a percentage of 8.68 after a spectacular performance during the last quarter of 2017. The reasons for the fall of Bitcoin are many and widely debated worldwide.

From reports of North Korea aiding crypto-hackers to attack Bitcoin exchanges and steal Bitcoins worth millions to governments of countries such as USA, China, South Korea to name a few raiding such exchanges and unwilling to support Bitcoin or other digital currencies, public sentiment has been largely influenced and their confidence diminishing in trading Bitcoin. Reports of big banking firm's CEO's speaking against Bitcoin has definitely hurt its credibility as well since getting a support from banks is an essential feature of a currency to be adopted. Although banks have voiced their opposition to Bitcoin, they have been spearheading the development of different products based on client demands using blockchain which is the backbone of Bitcoin as well. The security aspects of Blockchain are touted to be advantageous for multiple applications.

5.4.2 Impact of US Dollar on Global Foreign Exchanges

In the figure below, we see patterns of performance of global currencies here including US Dollar, Euro, Japanese Yen post the Midterm elections in United States of America during November 2018. A steady government at helm with a steady economic growth in during recent times has led to increasing investor and trader confidence in US Dollar. This has resulted in strengthening of US Dollar with traders preferring to buy US Dollar over other currencies as well their local currencies since US Dollar provides greater security. European Union has not been performing well economically as can be inferred from the latest results of firms' quarterly results which has led to fall in the Euro compared to US Dollar.



Figure 10 : Impact on Global currencies post Midterm Elections (November 2018)

The increase in federal interest rates 5 times in the current fiscal year has further supported this increasing trader confidence and a shift in global economic trends. The interest rate hikes were in accordance of improving economic scenario in United States of America which signal stability for the US Dollar and thereby other commodities connected with US Dollar such as Oil prices & Bullion prices.

Other major factors that have contributed towards increasing dependence on US Dollar are tax reforms in the United States in the past year as well as trade wars with China and increasing sanctions on Russia. The recent tax reforms have eased the tax burden on big corporations which leads to better profits and business development abilities. Similarly trade wars with China has led to a huge impact on Chinese Yuan with its devaluation being regularly undertaken by the People's Bank of China. This has led to traders jumping ship and demanding US Dollar in exchange for their products instead.

5.5 Reasons for fall of Bitcoin prices 2017 vs 2018

There has been a widespread crackdown on Bitcoin in the recent months following the magnanimous rise in the price of Bitcoin during the month of December 2017. This widespread crackdown although started since late last year, did not have a bearing on the prices up until January 2018 when the leading world economy USA's Supreme court ordered IRS to look into possible cases of tax evasions by early bitcoin investors in the years between 2009-2014 when the price of Bitcoin had risen tremendously. This news sent major ripples through the world which was already looking at news of successful/unsuccessful hacking attempts on different Bitcoin exchanges and silent crackdown on Bitcoin mining centers since late 2017. Below we analyze different reasons that have led or may have led to the fall of Bitcoin to its current value of \$6000 compared to \$20,000 in Dec'17.

Bitcoin along with Ethereum and other cryptocurrencies had been rising during the year of 2017. This led to an opening of Bitcoin Futures Trading at Chicago Mercantile Exchange Group (CME) with the aim of trading in bitcoin futures. In the past, futures trading has been introduced for other stock items as well such as Gold, Silver, Platinum to name a few. However, it had been noticed that when the either of above three mentioned futures were introduced in the market, their prices fell down by several points before regaining to their original value after a certain time period which varied from 4 weeks to 7 weeks easily. It is being speculated that even for Bitcoin, there must have been traders who might have bet on the fall of bitcoin after its introduction in the futures market. Negative speculation is a common feature observed in stock markets. It was also speculated that during the economic collapse in 2008, many traders had benefited from the fall since they had bet against the trend. This goes on to show that the stock market functions in unpredictable ways sometimes.

Bitcoin has been displaying signs of high volatility since the last few months. Its steady increase and subsequent huge fall lead one to speculate that maybe there may have been external forces trying to control the investor market by fluctuating bitcoin prices, thereby creating a demand supply situation. When the prices nearly touched 20,000 USD, these forces must have been the beneficiaries and sold their bitcoins at huge profits but at the cost of starting a huge turmoil in the market.

Countries like Russia, China, South Korea have been seriously cracking down on bitcoin mining centers and bitcoin exchanges banning the mining of bitcoins and any transactions involving them. South Korean Government has gone an extra mile by raiding financial institutions and exchanges dealing with bitcoins. This has set a massive negative and fearful sentiment among investors which can be attributed to the cause of bitcoin selling, thereby bringing down its price by over 13.7% in January itself

Governments have been leading the crackdown on bitcoin mining, exchanges, declaring them not legal in order to prevent their state currencies from getting affected. For instance, China's major bank People's Bank of China (PBOC) has been saying that bitcoin cannot be trusted since it's difficult to control a risky financial speculation, but maybe protecting yuan plays an important role.

The increase in federal interest rates (2%) by the Federal Reserve Board not only affects US Markets, but worldwide markets as well. For example, when the Federal reserve board increased the federal interest rates, it sent ripples through all major stock markets affecting the values of Nikkei, Sensex, Euro, GBP. Traders in these markets namely Japan Stock Exchange, Bombay Stock Exchange, London Stock Exchange started buying dollars which also affected the exchange rates. Now, when this happens naturally the leading banks of the respective countries have to step in to control any fall in their country's currency exchange value.

This change in federal interest rates can be seen impacting the prices of bitcoin. This is because investors purchase by exchanging one form of currency for 1 Bitcoin and there is a certain exchange rate setup. Now, when interest rates are increasing, people will naturally invest and buy USD since it's showing signs of stability and yield and not buy Bitcoin which has been severely volatile for the last several weeks.

6. Conclusions

6.1 Introduction

This chapter presents the conclusions from the research and demonstrate that the research objectives have been achieved. This chapter includes an overview of research results along with how the research performed can be improved in the future. Finally discussions on limitations and future improvements have also been covered in order to provide a possible future direction for research in this evergreen area.

6.2 Conclusions

The important findings from this research have been provided for achieving the research objectives envisaged in the beginning of this research. We will cover the objectives once again and an overview of the results that have been obtained.

1. Is it possible to have Bitcoin as a publicly traded commodity?

Result 1. Yes, it is entirely possible to have Bitcoin as a publicly traded commodity. From our research and the results that were obtained, we do observe that although volatile and having a degree of uncertainty to it, Bitcoin still behaves similar to traditional stocks. With further support from governments & financial institutions and oversight, Bitcoin can be traded. With the amount of activity and attention that Bitcoin has been garnering since the last few months, Chicago Board of Exchange has introduced Bitcoin Futures trading as of December 2017. This supports our conclusion very well that bitcoin can someday join mainstream stocks and be traded in global stock markets.

2. Can Bitcoin be used as any other traditional currency like US Dollar, Euro, Pound Sterling and so on.

Result 2. With support from our research results, we can say that Bitcoin does have the ability to be used for transactions. This would however require both the parties to setup a bitcoin wallet and have a secure payment method setup so as to prevent scrupulous elements from trying to hack into their wallets.

There has been more openness in accepting payments using Bitcoin by businesses. Most prominent among these are Microsoft, PayPal and Intuit. Big banking firms like Morgan Stanley, J.P Morgan have started a bitcoin trading desk in their respective companies so as to learn more about bitcoin and its impact.

When investors and traders see such global companies accepting bitcoin with open arms, it surely creates a feeling of positivity and optimism towards this digital currency and helps in further cementing its cause and use throughout.

3. Does Bitcoin have the ability to become a primary method of transaction, replacing US Dollar and other traditional currencies.

Result 3. For any new product to become global, it needs to be accepted by people and business firms alike. When we look at traditional methods of transactions such as US Dollar or Euro or any paper currency for that matter on which the global world depends, we observe that it has been accepted and implemented at the root level. Access to paper currency is the easiest and only requires barter of physical objects in order to earn it or give it away. Everything from oil prices, trading to commodities are purchased and sold using paper currency that is controlled by institutions and supported globally.

When we talk of Bitcoin which is a digital currency to start with, we realize that it not only requires people to have access to a digital device such as a smartphone but also requires them to be connected to wireless internet in order to support validation and authentication of bitcoin transactions. For countries where people have access to such technology there is a high possibility that bitcoin transactions can be adopted over traditional currency. However, in countries still developing and where people do not access to technology, this would prove to be a disaster.

Hence, although we are optimistic about Bitcoin becoming a medium of transaction in the present, it cannot be made as a primary method of transaction.

6.3 Limitations

There are many ways in which this research could be extended. Having a more continuous stream of bitcoin price data would have helped in developing an improved prediction model. This was observed when on days when the data was missing, the average values had to be filled in. These averages depending on how huge

the dataset was indeed had an influence on the prediction. A possible solution for this could be to implement pipelines in cloud using Kafka which allow continuous streaming of bitcoin price data from verified sources.

A major limitation that was encountered while studying Bitcoin was the huge volatility and fluctuation in its prices over a period of this research. Bitcoin prices steadily rose to 20,000 US Dollar from the course of September to December 2017. However, with news of IRS looking into Coinbase and South Korean & Japanese governments cracking down on Bitcoin exchanges, hacking of bitcoin exchanges in the US brought down the price by a huge percentage in the month of January 2018 itself. Since then the prices have been declining or we can say slowly becoming stable again.

While using sentiment analysis on Twitter data, it was observed that 40% of tweets obtained from Twitter were either garbage or duplicate tweets. Tweets by automated bots, tweets which have a keyword in them but no relation with the context are treated as garbage tweets. Hence, even though Twitter was a huge treasure trove of data for sentiment analysis and previously used for similar research by other researchers, it was realized that a more dependable and accurate source of information would be required for obtaining results from sentiment analysis. However, there is still optimism that Twitter improves their API's for fetching tweets and adds filters which would make it easier to collect the data automatically.

6.4 Future Improvements

Bitcoin has a huge potential of answering questions that have been lingering on in the capital markets about the advent of a new currency medium of transaction. There has been ongoing research on this and many papers have started appearing which are trying to find correlation between Bitcoin and its use in markets. This seems to be an exciting time for Bitcoin and other cryptocurrencies even though they have been inconsistent performers.

From sentiment analysis point of view, there are improvements to be made. Although this research made use of twitter data as well as news articles and web blogs, there is still possibility of mining more data from social media websites like Facebook and LinkedIn which could provide more clarity in results. (Best, 2018)

7. References

- [1]. Nakamoto, S. (2008). Bitcoin: A Peer-to-Peer Electronic Cash System.
- [2]. Bahloulou, W., & Bourib, A. (2016). The impact of investor sentiment on returns and conditional volatility in U.S. futures markets. *Journal of Multinational Financial Management*, 89-102.
- [3]. Anjaria, M., & Guddeti, R. (2014). Influence Factor Based Opinion Mining of Twitter Data Using Supervised Learning. *2014 Sixth International Conference on Communication Systems and Networks (COMSNETS)*. Mangalore: IEEE.
- [4]. Baker, M., & Wurgler, J. (2007). INVESTOR SENTIMENT IN THE STOCK MARKET. *Journal of Economic Perspectives, American Economic Association*, vol. 21(2), 129-152.
- [5]. Corredor, P., Ferrer, E., & Santamaria, R. (2013). Investor sentiment effect in stock markets: Stock characteristics or country-specific factors? *International Review of Economics & Finance Volume 27*, 572-591.
- [6]. Lemmon, M., & Portniaguina, E. (2006). Consumer Confidence and Asset Prices: Some Empirical Evidence. *The Review of Financial Studies, Volume 19, Issue 4*, 1499–1529.
- [7]. Moseki, K., & Rao, K. (2017). Analysing stock market data—Market sentiment approach and its measures. *Cogent Economics & Finance*.
- [8]. Qiu, L., & Welch, I. (September 2004). Investor Sentiment Measures. *NBER Working Paper No. 10794*.
- [9]. Saahil, M. (2015). Predicting Stock Price Direction using Support Vector Machine.
- [10]. Salzman, A. (2018, November 30). *Your Personal Data Is Being Used by Investors. Here's the Potential — and the Risks*. Retrieved from Barron's: <https://www.barrons.com/articles/how-big-investors-use-your-personal-data-to-play-the-stock-market-1543627499>.
- [11]. Uygun, U., & Taş, O. (2014). The impacts of investor sentiment on different economic sectors: Evidence from Istanbul Stock Exchange. *Borsa Istanbul Review*.
- [12]. Watts, W. (2018, September 8). *Will midterm elections sink the stock market? Here's what history says*. Retrieved from MarketWatch: <https://www.marketwatch.com/story/midterm-elections-usually-give-stocks-pause-in-september-but-pave-way-for-later-gains-2018-09-01>
- [13]. Bitcoin. Retrieved from : <https://en.wikipedia.org/wiki/Bitcoin>
- [14]. Sentiment Analysis. Retrieved from : https://en.wikipedia.org/wiki/Sentiment_analysis
- [15]. Best, R. (2018, October 1). *How the U.S. Dollar Became the World's Reserve Currency*. Retrieved from Investopedia: <https://www.investopedia.com/articles/forex-currencies/092316/how-us-dollar-became-worlds-reserve-currency.asp>
- [16]. Dwyer, G. P. (2015). The economics of Bitcoin and similar private digital currencies. *Journal of Financial Stability*, 81-91.
- [17]. Atsalakis, G. S., & Valavanis, K. P. (2009). Surveying stock market forecasting techniques – Part II: Soft computing methods. *Expert Systems with Applications*, 5932-5941 .
- [18]. Iuhasz, G., Tirea, M., & Negru, V. (2012). Neural Network Predictions of Stock Price Fluctuations. *2012 14th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing*, 505-512.

- [19]. Pring, M. J. (2014). *Technical Analysis Explained: The Successful Investor's Guide to Spotting Investment Trends and Turning Points*. New York: McGraw-Hill Education.
- [20]. Amadeo, K. (2018, September 05). *Bretton Woods System and 1944 Agreement*. Retrieved from the balance: <https://www.thebalance.com/bretton-woods-system-and-1944-agreement-3306133>
- [21]. Amadeo, K. (2018, October 19). *How the Fed Rate Hike Affects You*. Retrieved from the balance: <https://www.thebalance.com/how-fed-rate-hike-affects-you-4119767>
- [22]. Hur, Y., Jeon, S., & Yoo, B. (2015). Investigation into the Market Viability of Bitcoin: Measuring the Digital Currency's Speculative Nature. *SSRN 25999953*.
- [23]. Hayes, A. (2015). A cost of production model for Bitcoin. *SSRN 2580904*.
- [24]. Brandvold, M., Molnár, P., Vagstad, K., & Valstad, O. (2015). Price Discovery on Bitcoin Exchanges. *Journal of International Financial Markets, Institutions & Money*, vol. 36, 18-35.
- [25]. Vo, N. N., & Xu, G. (2017). The volatility of Bitcoin returns and its correlation to financial markets. *2017 International Conference on Behavioral, Economic, Socio-cultural Computing (BESCom)*, 1-6.
- [26]. Dyhrberg, A. H. (2016). Hedging capabilities of bitcoin. Is it the virtual gold? *Finance Research Letters*, 139-144.
- [27]. Jang, H., & Lee, J. (2017). An Empirical Study on Modeling and Prediction of Bitcoin Prices With Bayesian Neural Networks Based on Blockchain Information. *IEEE Access (Volume: 6)*, 5427 - 5437.
- [28]. Radityo, A., Munajat, Q., & Budi, I. (2017). Prediction of Bitcoin exchange rate to American dollar using artificial neural network methods. *2017 International Conference on Advanced Computer Science and Information Systems (ICACSIS)* (pp. 433 - 438). Bali, Indonesia: IEEE.
- [29]. Kelleher, K. (2018, July 02). *Bitcoin Futures Killed the Bitcoin Rally (and Will Keep It from Returning), Economists Say*. Retrieved from Fortune: <http://fortune.com/2018/07/02/bitcoin-futures-killed-bitcoin-rally-will-keep-coming-back-economists/>
- [30]. Hale, G., Krishnamurthy, A., Kudlyak, M., & Shultz, P. (2018). *How Futures Trading Changed Bitcoin Prices*. San Francisco: Research from Federal Reserve Bank of San Francisco.
- [31]. Parveen, H., & Pandey, S. (2016). Sentiment Analysis on Twitter Data-set using Naive Bayes Algorithm. *2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)*. Bangalore, India: IEEE.
- [32]. Ramadhani, A. M., & Goo, H. (2017). Twitter Sentiment Analysis using Deep Learning Methods. *2017 7th International Annual Engineering Seminar (InAES)* (pp. 1-4). Yogyakarta, Indonesia: IEEE.
- [33]. McAteer, C. (2014). *Twitter Sentiment Analysis to Predict Bitcoin Exchange Rate*. Dublin: University of Dublin.
- [34]. McNally, S. (2016). *Predicting the price of Bitcoin using Machine Learning*. School of Computing, National College of Ireland.
- [35]. Chakraborty, P., Priya, U., Rony, M., & Majumdar, M. (2017). Predicting stock movement using sentiment analysis of Twitter feed. *2017 6th International Conference on Informatics, Electronics and Vision & 2017 7th International Symposium in Computational Medical and Health Technology (ICIEV-ISCMT)* (pp. 1-6). Himeji, Japan: IEEE.

- [36]. Li, J., Bu, H., & Wu, J. (2017). Sentiment-aware stock market prediction: A deep learning method. *2017 International Conference on Service Systems and Service Management* (pp. 1-6). Dalian, China: IEEE.
- [37]. Deng, S., Mitsubuchi, T., Shioda, K., Shimada, T., & Sakurai, A. (2011). Combining Technical Analysis with Sentiment Analysis for Stock Price Prediction. *2011 IEEE Ninth International Conference on Dependable, Autonomic and Secure Computing* (pp. 800-807). Sydney, NSW, Australia: IEEE.
- [38]. Tar, A. (2018, January 17). *Proof-of-Work, Explained*. Retrieved from CoinTelegraph: <https://cointelegraph.com/explained/proof-of-work-explained>
- [39]. Nelson, E. (2018, August 16). *The biggest surprise in markets this year is the mighty US dollar*. Retrieved from Quartz: <https://qz.com/1356961/why-is-the-us-dollar-so-strong/>
- [40]. Knaub, J. R. (2014). *What are the weaknesses of time series model and how could it be improved?* Retrieved from ResearchGate: https://www.researchgate.net/post/What_are_the_weaknesses_of_time_series_model_and_how_could_it_be_improved
- [41]. Brownie, J. (2017, January 9). Retrieved from Machine Learning Mastery: <https://machinelearningmastery.com/arma-for-time-series-forecasting-with-python/>
- [42]. Bovaird, C. (2016, October 2). *Could Rising Interest Rates Threaten Bitcoin Prices?* Retrieved from Coindesk: <https://www.coindesk.com/rising-interest-rates-threaten-bitcoin-prices>
- [43]. Bovaird, C. (2018, March 21). *How Will Today's Fed Decision Impact Bitcoin And Gold?* Retrieved from Forbes: <https://www.forbes.com/sites/cbovaird/2018/03/21/how-will-todays-fed-decision-impact-bitcoin-and-gold/#43abdb913bc3>
- [44]. CCN. (2018, March 29). *Did Futures Markets Negatively Affect Bitcoin Price and the Entire Crypto Industry?* Retrieved from CCN: <https://www.ccn.com/did-the-futures-market-negatively-affect-bitcoin-price-and-the-entire-industry/>
- [45]. Davis, C. (2018, March 29). *Bitcoin price: Expert warns this 'key NEGATIVE' is likely to make bitcoin TUMBLE*. Retrieved from Express: <https://www.express.co.uk/finance/city/938440/Bitcoin-price-usd-news-stock-value-cryptocurrency-buy-latest-today-decline>
- [46]. Derousseau, R. (2018, January 12). *Why South Korea's Cryptocurrency Crackdown Could Pop the Bitcoin Bubble*. Retrieved from Money: <http://time.com/money/5100299/why-south-koreas-cryptocurrency-crackdown-could-pop-the-bitcoin-bubble/>
- [47]. Investopedia. (n.d.). *Fundamentals*. Retrieved from Investopedia: <https://www.investopedia.com/terms/f/fundamentals.asp>
- [48]. Kizil, A. (2018, July 10). *Understanding Market Sentiment and How It Affects Asset Prices*. Retrieved from Canstar: <https://www.canstar.com.au/investor-hub/understanding-market-sentiment/>
- [49]. Long, H. (2018, March). *Federal Reserve hikes interest rate to highest level in a decade amid stronger economy*. Retrieved from Washington Post: https://www.washingtonpost.com/news/wonk/wp/2018/03/21/federal-reserve-hikes-u-s-interest-rate-to-highest-level-in-a-decade-amid-stronger-economy/?noredirect=on&utm_term=.42b3f1c55607
- [50]. Murphy, C. (n.d.). *Essential Strategies for Trading Volume*. Retrieved from Investopedia: <https://www.investopedia.com/articles/active-trading/091514/essential-strategies-trading-volume.asp>
- [51]. Proof-of-work system. (n.d.). Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Proof-of-work_system#Reusable_proof-of-work_as_e-money

- [52]. Ritcher, W. (2017, February 16). *All that money printing is finally producing inflation*. Retrieved from Business Insider: <https://www.businessinsider.com/money-printing-is-finally-producing-inflation-2017-2>
- [53]. Sheikh, A. (2017). *How Will Bitcoin Futures Affect Bitcoin Prices? Here's What History Says*. Retrieved from masterthecrypto: <https://masterthecrypto.com/will-bitcoin-futures-affect-bitcoin-prices/>
- [54]. Supervised learning. (n.d.). Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Supervised_learning
- [55]. Heinz, S. (2017, November 9). *A simple deep learning model for stock price prediction using TensorFlow*. Retrieved from Medium: <https://medium.com/mlreview/a-simple-deep-learning-model-for-stock-price-prediction-using-tensorflow-30505541d877>
- [56]. Keras. (n.d.). *Keras : The Python Deep Learning Library*. Retrieved from Keras: <https://keras.io/>
- [57]. Kumar, N. (n.d.). *Twitter Sentiment Analysis using Python*. Retrieved from GeeksForgeeks: <https://www.geeksforgeeks.org/twitter-sentiment-analysis-using-python/>
- [58]. TensorFlow. (2018, 12 12). *Introduction*. Retrieved from tensorflow: https://www.tensorflow.org/guide/low_level_intro
- [59]. TensorFlow. (n.d.). *TensorFlow*. Retrieved from tensorflow: <https://www.tensorflow.org/>
- [60]. Rosebrock, A. (2017, July 10). *Using Tesseract OCR with Python*. Retrieved from pyimagesearch: <https://www.pyimagesearch.com/2017/07/10/using-tesseract-ocr-python/>
- [61]. Sharma, A. (2017, March 30). *Understanding Activation Functions in Neural Networks* . Retrieved from Medium: <https://medium.com/the-theory-of-everything/understanding-activation-functions-in-neural-networks-9491262884e0>