**A Dissertation Proposal on**

**Trading Strategy in Cryto Currency Market**

**Daniel Sapkota**

**ID: 16401960**

**BSc Computing**

**Supervisor: Himalayan Kakshyapati**

**NAMI College**

**University of Northampton**

**Jorpati, Kathmandu, Nepal**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Course: BSc Computing Module: CSY4010 | | | | | | |
| Assessment: Computing Project Proposal | | | | | | |
| Project Title: **Trading Strategy in Cryto Currency Market** | | | | | | |
| Date due in: | Agreed extension: | | | | | |
| Student Name: Daniel Sapkota  Student ID: 16401960 Signature:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | | | | | | |
| Module Tutor: Supervisor: | | | | | | |
| Assessment Feedback | | Please note the UON Policy Statement below | | | | |
| Marker's General view of the work: | | Mark/Grade: | | | | |
| Rating Scale | | A | B | C | D | F |
| 1. Introduction: Statement of problem (10%) | |  |  |  |  |  |
| 2. Background Research, Literature Review (20%) | |  |  |  |  |  |
| 3. Aims & Objectives (20%) - Intellectual Challenge | |  |  |  |  |  |
| 4. Proposed Methodology (20%) | |  |  |  |  |  |
| 5. Professional, Legal, Ethical Issues (5%). | |  |  |  |  |  |
| References & Bibliography (5%) | |  |  |  |  |  |
| Appendix 1. Gantt Chart (10%) | |  |  |  |  |  |
| Report Quality & Standard (10%) | |  |  |  |  |  |
| FEEDBACK/COMMENTS: | | | | | | |
| Supervisor’s Signature: Date: | | | | | | |

Table of Contents

[1 Introduction 1](#_Toc500262553)

[1.1 Introduction to the Topic 1](#_Toc500262554)

[1.2 The Problem Domain 1](#_Toc500262555)

[1.3 Hypothesis 1](#_Toc500262556)

[2 Literature Review 2](#_Toc500262557)

[2.1 Methods in the Stock Market 2](#_Toc500262558)

[2.2 Studies on Social Media Effects 2](#_Toc500262559)

[2.3 Studies focusing on Bitcoin Market 3](#_Toc500262560)

[3 Aim and objectives 3](#_Toc500262561)

[3.1 Aims 3](#_Toc500262562)

[3.2 Objectives 3](#_Toc500262563)

[4 Methodology 4](#_Toc500262564)

[4.1 Chart Data Collection 4](#_Toc500262565)

[4.2 Social Media Data 4](#_Toc500262566)

[4.3 Web Based Data 4](#_Toc500262567)

[4.4 General Pattern 5](#_Toc500262568)

[4.5 Transaction Based Data 5](#_Toc500262569)

[4.6 Machine Learning 5](#_Toc500262570)

[5 Implementation 5](#_Toc500262571)

[5.1 Chart Data Collection 5](#_Toc500262572)

[5.2 Web Based Data 5](#_Toc500262573)

[5.3 Chart Data Analysis 5](#_Toc500262574)

[5.4 Transaction Based Data 6](#_Toc500262575)

[5.5 Machine Learning 6](#_Toc500262576)

[6 Testing/ Evaluation 6](#_Toc500262577)

[7 Professional, social, economic and legal issues 6](#_Toc500262578)

[7.1 Legal Issues 6](#_Toc500262579)

[7.2 Economic Issues 7](#_Toc500262580)

[7.3 Professional Issues 7](#_Toc500262581)

[8 Resource Requirements 7](#_Toc500262582)

[8.1 Google Cloud 7](#_Toc500262583)

[9 References 7](#_Toc500262584)

[10 Appendix 1 – Project Gantt Chart, (project activities plan) A](#_Toc500262585)

[11 Appendix 2 – Concepts I](#_Toc500262586)

[11.1 Economics of Crypto Currency I](#_Toc500262587)

[11.2 Machine Learning II](#_Toc500262588)

[11.3 Social Media II](#_Toc500262589)

[11.4 Quants in the Stock Market II](#_Toc500262590)

[12 Appendix 3 - Additional Terminologies III](#_Toc500262591)

# Introduction

## Introduction to the Topic

In 2008, whitepaper about a decentralized digital currency that claimed to solve the problem of double spending was published. Double Spending had been the biggest constrain for the creation of a decentralized digital currency. Double Spending takes place when the same amount of currency is spent more than once. To solve this problem - Digital currencies used to rely on a server (Like Liberty Reserve, Perfect Money). Addition of a server added central point of failure. Bitcoin solved this problem by creating a public transaction ledger called blockchain and by using Proof of Work (PoW) system. A Bitcoin transaction occurs when calculations (work – thus PoW) is done to find a nonce (a random string) that results in a block (contains list of transactions) hash smaller than a given hash (called difficulty which is automatically calculated). When it is found, the block is added next to the previous blocks in a blockchain and distributed. Any blocks that contain a double spend transaction is rejected by the network when the blocks are distributed. The inner working of bitcoin is longer and beyond the scope of this thesis, so the cited materials should be considered for further information. (Antonopoulos, 2014; Nakamoto, 2009)

After the creation of Bitcoin, there has been several other implementations of Blockchain – with major as well as minor changes. These different coins are traded with each other like stocks. Details about the Market Operations is included in Appendix.

## The Problem Domain

Most of the profitable trading strategies are proprietary because there is no incentive to share them. Sharing them will make the methods unprofitable very fast. The Crypto Currency Market has been an extremely under explored topic. This is an attempt to create a trading system that is profitable in the crypto currency market by modifying strategies from the traditional market or by creating new ones.

Most Non-Academic research in this field has only focused on creating a trading algorithm. There hasn’t been focus on creating a portfolio that complies with the basic laws of economics and manages the risks well. Computer based trading dominates wall street (discussed more on Appendix) but it is not as common in this market – leaving an opportunity. Most of strategies in the market works when the market is bullish (rising). But a strategy that can profit from both bullish and bearish (falling) market and doesn’t have much risk is extremely uncommon.

## Hypothesis

This thesis attempts to prove the hypothesis that a programmatic trading system can be created on the crypto currency market to profit during any Market Trends, like in the Wall Street. To prove this hypothesis this thesis will work on proving/disproving several other hypotheses:

1. Patterns can be derived from past data to find how likely coins can perform in similar circumstances.
2. Data from Social Media, News Sites and Google and Wikipedia Trends can be used to determine general sentiment in a coin.
3. Blockchain based data like transactions can help determining interest in a coin.
4. Technical Analysis data can be combined with Statistical Functions and other data to make a system better.
5. Bots are used in Social Media to change general sentiment.
6. Low number of quants leave opportunity for arbitrage and correlation based trading.
7. Fundamentals are the highest determinant in a timeframe of years/decades. Some basics about the fundamentals can also be determined from charts
8. Machine Learning can be used with some of these data’s as features to create a profitable system.

# Literature Review

There has been attempts to “beat the market” as soon as the Financial Markets started. Speculative Investors have tried to beat the market for centuries. In initial days, speculation was most of the daily volume (Andersen, 2012). Jesse Livermore, a legendary stock investor used to remember pattern’s and use them before computers was invented in the 1920’s (Lefevre, 2004).

## Methods in the Stock Market

Time Anomaly like January Effect (Increase in stock prices during January) has been observed in previous works on stock markets. These Anomalies can exist in crypto markets too and because it is active across multiple time zones, there might be distinct characters for certain time zones. These Anomaly create trading opportunities.

Neglected Firm Effect (Lesser Known Firms producing abnormally high return) is another anomaly that can be easily imported to the crypto currency Market. Here too the smallest coins have gained the most in times (Bartos, 2015). Most of the public Machine Learning systems using these features in the US Stock Market have not achieved better returns but have achieved lower Standard Deviation (The measurement of Risk) than the S&P 500 Index. Using Standard Deviation as a measurement of risk is not always the best because it treats upside and downside the same. Other systems Like Value at Risk and Omega Ratio exist if the need is different. Because there are a lot of features, Kahn Processing Networks have been used in previous studies to perform parallel calculations. In most of the studies in the US Stock Market, SVN performed better than ANN (Andersen, 2012).

## Studies on Social Media Effects

Another Important place where features (to predict the Crypto Currency price) can be obtained is Social Media. The possibility that Social Media is being used to manipulate public opinion has been proved by various studies (Ratkiewicz, et al., 2011; Ferrara, 2015). For manipulations, bots (explained in Appendix 4) are typically used. BotOrNot is a framework developed for detecting if a twitter account is bot or not using 1000 features. It gives an account score between 0 and 1 where a score above 0.5 can indicate the account is a bot. The features with biggest influence were - Absence of geographical meta data, Activity statistics, Account creation date (bots had more recent accounts) and the Randomness of username. Bot or Not is around 95% accurate. (Davis, et al., 2016)

SentiStrength is a python module which assigns positive and negative polarity score between one and five for a given text. The sentiment score is difference between them where 0 indicates neutrality (Thelwall, et al., 2010). They can be used on reddit and twitter data to find out what mood bots are trying to spread and that can be used to profit. Some features for the model can be grabbed from here. A use of BotOrNot and SentiStrength in the 2016 Election Twitter Data showed that most of the positive bots were supporting Trump. It reviewed 20.7 million Tweets (Bessi & Ferrara, 2016). The value of Social Media can be seen by the fact that A Political Company in Mexico paid 80k $ in election campaign to create 22 trending topics in Twitter(Forelle, et al., 2015). During the 2016 US Election Campaign, The Trump Campaign worked with Cambridge Analytica, a firm whose investor is Robert Mercer - the CEO of Renaissance Technologies mentioned in Appendix 3 for Social Media boosting which probably involved the use of bots (Mayer, 2017).

## Studies focusing on Bitcoin Market

Studies using trivial features like Daily price change, Close Price, Volume, Average, SD, RSI (A technical indicators), Daily Return, % Oscillation, Average true range, Vortex and On Balance Volume have performed better than buying and holding in case of Bitcoin. A Machine Learning algorithm using these data had lesser draw down and greater return than buy and hold strategy (22.88 DD and 33.52 return compared to 4.86% return and 30.9 DD) (Żbikowski, 2016). Although the timeframe is small to conclude anything, the promises are bright. Another study used Block time, Difficulty, Hashrate, no of transactions (non market data), trade volume and transaction to trade ratio as features using Binomial Logical Regressing, SVN and Random walk to get considerable returns. Random walk turned out to bring the best results. (Madan, et al., 2015).

Studies have found significant correlation between google trend and bitcoin price and the volume of positive tweets (Matta, et al., 2015; Georgoula, et al., 2015). Statistical functions like Wavelet Coherence Analysis and Fourier Spectrum have also been used independently (Kristoufek, 2015; Bouoiyour, et al., 2016). Such result shows that addition of other variables like manipulation factors, social media sentiment, altcoin movement brings huge promises.

# Aim and objectives

## Aims

The Aim of this project is to create a trading system which can profit in both bullish and bearish markets.

## Objectives

The following objectives will be performed to reach the aim:

* Create a script to detect Sentiment in Reddit and Twitter for a coin.
* Find ways of determining the flow of news and the general sentiment conveyed by them.
* Track big movements across the blockchain
* Find Patterns from past data
* Find Correlation between the coins and find the optimal way to exploit them.
* Use basic data like transactions and search trends to find the change in interest
* Use Technical Analysis and Statistical Function as features for determining the Market Movement.
* Find time related correlation and how different markets react to an event
* Analyze News and the official forums to find out what is happening
* Combine some data and use Machine Learning to forecast the price change
* Create models to find the best diversification and mange risks
* Create a trading script based on the results above.
* Create separate scripts for each task for easy management and troubleshooting

The target audience of the deliverable is someone with basic knowledge about programming, mathematics and the Internet who is willing to learn other topics. To understand the report well and repeat the results, someone with knowledge of Machine Learning is needed. Although knowledge of Crypto Currencies and Financial Markets is a huge plus, it can be acquired while the report is being read.

# Methodology

The method revolves around collecting and analyzing data. These data are used as features in Machine Learning or by themselves to create a trading strategy.

## Chart Data Collection

Poloniex is a US based Crypto Currency exchanges which has the biggest volume for Altcoin to Bitcoin trade pairs. They provide API access to altcoin trading data history from the time it was started. Bittrex is another similar exchange which provides API access to all historical trade data.

For bitcoin margin and regular trading data, Bitfinex, Bithumb, Kraken, OKCoin and Coinbase API can be used. These are different exchanges popular on different countries. Bitfinex is more popular in Asian market (recently), Bithumb in the Korean (which is Asia but Korea has had high influence in the price), Coinbase is American. OKCoin is Chinese Market which was the initial driving force in Bitcoin. For options data, Deribit API will be used.

## Social Media Data

Reddit and Twitter provide API to access posts and other data their platform. Sub-Reddits (Discussion groups) for different coins should first be specified. Then data for keywords denoting a given coin will be searched in twitter. BotOrNot framework will be used to determine if the Twitter Account is a bot and what topic they are trying to spread if they are. For Reddit Account, methods like levenstien distance and other features used in BotOrNot relevant to it can be used. After the accounts are identified, the sentiment will be calculated using SentiStrength to find the sentiment normal people feel and how bots are trying to change it.

## Web Based Data

Google Trends Provide historic search data for a given keyword. Wikipediatrends provide search trends in Wikipedia to determine popularity. Google provides search API through Google Developers Console. Data from top level news sites like – Forbes, Business Insider and Crypto Currency based like Coin Desk, Coin Telegraph and Crypto Coin News can be found from the Google API and scrapped to analyze.

## General Pattern

Data derived from the charts can be used to create features based on basic technical analysis – like RSI (Relative Strength Index), Head and Shoulders and MACD (Moving Average Convergence Divergence). Features can also be made from trailing data. Similar past patterns can be studied and their occurrence can also be calculated.

## Transaction Based Data

Bitcoin Blockchain is a public data (along with blockchain of any other coin). These Data can be analyzed using SQL.

## Machine Learning

After the features are collected, Machine Learning will be done using SVM, NN and Random Walk. These methods will be used as most of the studies done in Bitcoin show high amount of success rate using these methods. Alternative Machine Learning Methods could be pursued but this method have had the highest success rate in similar studies. Instead of Machine Learning, Statistics can be used to determine correlations. Although using statistics might be faster, the results in similar studies using statistics have not been stronger. So, some statistical functions will only be used as features.

# Implementation

## Chart Data Collection

API of the exchanges mentioned in the Methodology section will be used in Python to collect Chart Data. Bitcoin Price data will be grabbed using Quandl for Python while web requests will be used to collect data from other exchanges. These data will be read as Numpy Arrays.

Some Data needs to be logged live. For this Amazon AWS instance with ubuntu will be used. Social Media Data.

## Web Based Data

Web Requests will be used to grab data from news sites in python while BeautifulSoup will be used to scrape the needed news data.

Twitter and Reddit API will be used to grab data from these platform via Python. For Social Media Data, BotOrNot and its modification will be used in python along with SentiStrength.

## Chart Data Analysis

For detecting General Patterns, Machine Learning Techniques will be employed. Past Data’s will be manually selected for some patterns.

## Transaction Based Data

Blockchain for any coin can be downloaded by downloading the official wallet. Then the Blockchain can be analyzed using SQL or any other Languages.

## Machine Learning

Machine Learning Modules in Python like scikit will be used for majority of the analysis. Feature calculation requires enormous computing power and for this GPU rigs provided by Google Cloud will probably be used.

# Testing/ Evaluation

The Project will be successful if it can create a trading strategy that can make profit in bull as well as bear markets in a timeframe of months. Back testing will be done to see if these criteria have been met. Different scripts will be in different files and they will be tested separately to test bugs. Also, Python Unit testing will be used to determine correct functionality.

Each feature will be carefully examined by testing the performance of the system with/without them.

Data will be divided into training data and testing data to see how well the strategy would have performed when long/short positions were opened on the testing data.

Charts will be made to show how much the factors affected the price. And then results will be shown in a table.

# Professional, social, economic and legal issues

## Legal Issues

There is no consensus on how Bitcoin and Crypto Currencies should be treated. It is being treated as Commodity (Canada), as Money Business Service (USA), as Financial Service (Finland) or the regulation is being worked on (Austria, Germany and others) or it is banned (Nepal, Bangladesh and others) (Unichange, n.d.).

Nepal, out rightly bans holding and trading Bitcoin by applying the law of Foreign Exchange (Redman, 2017). This thesis however does not require holding Bitcoin as it just considers hypothetical trading.

Poloniex is a US based Crypto Currency exchange. So, US laws may have to be considered when trading from here. But there hasn’t been any work done in this field, so it is unclear.

According to a SEC report: “U.S. federal securities law may apply to various activities, including distributed ledger technology, depending on the particular facts and circumstances, without regard to the form of the organization or technology used to effectuate a particular [cryptocurrency] offer or sale.” This line is unclear and does not show weather the security laws will apply in crypto. SEC has been silent in most stuffs regarding crypto currency. (Churchouse, 2017)

If Crypto’s are indeed Security, ample of laws will apply.

Twitter and Reddit and US based Social Network. They are private companies and have their own policies regarding bots and manipulation. However, breach of ToS is not breach of law.

## Economic Issues

This thesis aims to create models that can be further built upon to bring Economic returns. Crypto Currencies are the Internet of Money. It is currently a growing market (per year average growth of all crypto currencies). Creating Models and bringing in new money helps the space. The Economics of Crypto Currencies is explained in the Appendix 2.

Shorting strategies and increase in number of quants can decrease the profits made through manipulations so it a good thing for long term growth of the crypto currencies.

## Professional Issues

Crypto Currency do not belong to the developers who develop them. Currently (November 2017) they gained most of their popularity because of that reason of decentralization. Bitcoin is not a company with a CEO. So, there are no official bodies to grant recognitions to Bitcoin relate projects. FluffyPony, the lead developer of Monero (a crypto currency that focuses on anonymity) once pulled off a huge stunt via twitter that resulted in a big fall in price to show that there are no official bodies for a decentralized crypto currency.

However, Linux Foundation mailing lists was used by Satoshi Nakamoti to update changes in Bitcoin. It is still used to distribute information. There are big forums like bitcointalk and sub reddits on reddit where discussion takes place. Some crypto currencies have developed nonprofit foundations of linked groups to discuss. Demand for a trading system has been high among some developers and members of Bitcoin Community like Bitcoin Talk. However, bitcointalk and the /r/Bitcoin subreddit – biggest places for Bitcoin related discussion - in reddit are extremely censored. There are other Crypto Currency based forums but they are not as big as them.

# Resource Requirements

## Google Cloud

Machine Learning requires enormous computation power that cannot be used by normal PC’s. Due to this reason, calculations will be done on Google Cloud Platform.

# References

Andersen, A., 2012. *A Novel Algorithmic Trading Framework Applying Evolution and Machine Learning for Portfolio Optimization.* s.l.:(Doctoral dissertation, Doctoral dissertation, Masters Thesis, Faculty of Social Science and Technology Management, Department of Industrial Economics and Technology Management)..

Antonopoulos, A. M., 2014. *Mastering Bitcoin: Unlocking Digital Cryptocurrencies.* 1st ed. s.l.:O'Reilly Media.

Bartos, J., 2015. Does Bitcoin follow the hypothesis of efficient market?. *International Journal of Economic Sciences,* 4(2), pp. 10-23.

Bessi, A. & Ferrara, E., 2016. Social Bots Distort the 2016 US Presidential Election Online Discussion. *First Monday,* 21(11).

Bouoiyour, J., Selmi, R., Tiwari, A. K. & Olayeni, O. R., 2016. What drives Bitcoin price. *Economics Bulletin,* 36(2), pp. 843-850.

Burton, K., 2016. *Inside a Moneymaking Machine Like No Other.* [Online]   
Available at: https://www.bloomberg.com/news/articles/2016-11-21/how-renaissance-s-medallion-fund-became-finance-s-blackest-box  
[Accessed 4 11 2017].

Cecchetti, S., 2000. *Asset prices and central bank policy.* s.l.:Centre for Economic Policy Research.

Clarke, J., Jandik, T. & Mandelker, G., 2001. The efficient markets hypothesis. In: *Expert financial planning: Advice from industry leaders.* s.l.:s.n., pp. 126-141.

CoinMarketCap , 2017. *Total Market Capitalization.* [Online]   
Available at: https://coinmarketcap.com/charts/  
[Accessed 1 11 2017].

Davis, C. A. et al., 2016. *Botornot: A system to evaluate social bots.* s.l., 25th International Conference Companion on World Wide Web.

Donoho, S., 2004. *Early detection of insider trading in option markets.* s.l., ACM.

Fama, E. F. & Burton, M. G., 1970. Efficient capital markets: A review of theory and empirical work. *The journal of Finance,* 25(2), pp. 383-417.

Ferrara, E., 2015. Manipulation and abuse on social media by emilio ferrara with ching-man au yeung as coordinator. *ACM SIGWEB Newsletter,* p. 4.

Forelle, M. C., Howard, P. N., Monroy-Hernández, A. & Savage, S., 2015. *Political bots and the manipulation of public opinion in Venezuela.* [Online]   
Available at: https://ssrn.com/abstract=2635800  
[Accessed 4 11 2017].

Georgoula, I. et al., 2015. *Using Time-Series and Sentiment Analysis to Detect the Determinants of Bitcoin Prices.* [Online]   
Available at: https://ssrn.com/abstract=2607167  
[Accessed 11 4 2017].

Hankin, A., 2017. *Bitcoin Pizza Day: Celebrating the $20 Million Pizza Order.* [Online]   
Available at: http://www.investopedia.com/news/bitcoin-pizza-day-celebrating-20-million-pizza-order/  
[Accessed 1 11 2017].

Hopwood, D., Bowe, S., Hornby, T. & Wilcox, N., 2017. *Zcash Protocol Specication.* [Online]   
Available at: https://raw.githubusercontent.com/zcash/zips/master/protocol/protocol.pdf  
[Accessed 1 11 2017].

Howard, P. N., 2006. *New media campaigns and the managed citizen.* s.l.:Cambridge University Press.

Kishimoto , M., Tian , X. & Xu, L., n.d. *Detecting Insider Trading.* s.l.:MS&E444: Final Project Report .

Kristoufek, L., 2015. What are the main drivers of the Bitcoin price? Evidence from wavelet coherence analysis. *PloS one,* 10(4), p. e0123923.

Laffont, J.-J. & Maskin, E. S., 1990. The Efficient Market Hypothesis and Insider Trading on the Stock Market. *The Journal of Political Economy,* 98(1), pp. 70-93.

Lefevre, E., 2004. *Reminiscences of a stock operator.* s.l.:John Wiley & Sons.

Lin, T. C. W., 2017. The New Market Manipulation. *Emory Law Journal,* Volume 66, p. 1253.

Madan, I., Saluja, S. & Zhao, A., 2015. *Automated Bitcoin Trading via Machine Learning Algorithms.* [Online]   
Available at: https://pdfs.semanticscholar.org/e065/3631b4a476abf5276a264f6bbff40b132061.pdf  
[Accessed 11 4 2017].

Maidsafe Team, 2014. *MaidSafe.net announces project SAFE to the community.* [Online]   
Available at: MaidSafe.net announces project SAFE to the community  
[Accessed 1 11 2017].

Markham, J. W., 2013. *Law enforcement and the history of financial market manipulation.* s.l.:ME Sharpe.

Matta, M., Lunesu, I. & Marchesi, M., 2015. *Bitcoin Spread Prediction Using Social and Web Search Media.* s.l., UMAP Workshops.

Mayer, J., 2017. *The Reclusive Hedge-Fund Tycoon Behind the Trump Presidency.* [Online]   
Available at: https://www.newyorker.com/magazine/2017/03/27/the-reclusive-hedge-fund-tycoon-behind-the-trump-presidency  
[Accessed 4 11 2017].

Maziéres, D., 2015. *The Stellar Consensus Protocol: A Federated Model for Internet-level Consensus.* [Online]   
Available at: https://www.stellar.org/papers/stellar-consensus-protocol.pdf  
[Accessed 1 11 2017].

Nakamoto, S., 2009. *Bitcoin: A peer-to-peer electronic cash system.* [Online]   
Available at: https://bitcoin.org/bitcoin.pdf  
[Accessed 1 11 2017].

Noether, S., 2014. *REVIEW OF CRYPTONOTE WHITE PAPER.* [Online]   
Available at: https://downloads.getmonero.org/whitepaper\_review.pdf  
[Accessed 1 11 2017].

Patterson, S., 2011. *The quants: How a new breed of math whizzes conquered wall street and nearly destroyed it.* 1st ed. s.l.:Crown Business.

Ratkiewicz, J. et al., 2011. Detecting and Tracking Political Abuse in Social Media. *ICWSM,* Volume 11, pp. 297-304.

Ripple, n.d. *Solution Overview.* [Online]   
Available at: https://ripple.com/files/ripple\_solutions\_guide.pdf  
[Accessed 1 11 2017].

Shiller, R. J., 2015. *Irrational exuberance.* 3rd ed. s.l.:Princeton university press.

The Golem Project, 2016. *Crowdfunding Whitepaper.* [Online]   
Available at: http://golemproject.net/doc/DraftGolemProjectWhitepaper.pdf  
[Accessed 1 11 2017].

Thelwall, M. et al., 2010. Sentiment strength detection in short informal text. *Journal of the Association for Information Science and Technology,* 61(12), pp. 2544-2558.

Unichange, n.d. *The Legal Status of Bitcoin.* [Online]   
Available at: https://unichange.me/articles/legal\_status\_of\_bitcoin  
[Accessed 1 11 2017].

Vorick, D. & Champine, L., 2014. *Sia: Simple Decentralized Storage.* [Online]   
Available at: https://www.sia.tech/whitepaper.pdf  
[Accessed 1 11 2017].

White, L. H., 2015. The market for cryptocurrencies. *Cato J,* Volume 35, p. 383.

Wood, D. G., 2017. *ETHEREUM: A SECURE DECENTRALISED GENERALISED TRANSACTION LEDGER.* [Online]   
Available at: https://ethereum.github.io/yellowpaper/paper.pdf  
[Accessed 1 11 2017].

Żbikowski, K., 2016. Application of Machine Learning Algorithms for Bitcoin Automated Trading. *Machine Intelligence and Big Data in Industry 2016,* pp. 161-168.

# Appendix 1 – Project Gantt Chart, (project activities plan)

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Task** | **November** | | **December** | | **January** | | **February** | | **March** | | **April** | |
| Learn More About Machine Learning. |  |  |  |  |  |  |  |  |  |  |  |  |
| Start Logging Data from Exchanges |  |  |  |  |  |  |  |  |  |  |  |  |
| Read Papers and Books on Machine Learning Models |  |  |  |  |  |  |  |  |  |  |  |  |
| Start working on creating and testing Features |  |  |  |  |  |  |  |  |  |  |  |  |
| Grab the Reddit and Twitter Social Media Data and analyze them |  |  |  |  |  |  |  |  |  |  |  |  |
| Explore the Bitcoin Blockchain and grab features from it |  |  |  |  |  |  |  |  |  |  |  |  |
| Create the model |  |  |  |  |  |  |  |  |  |  |  |  |
| Improve the Model. |  |  |  |  |  |  |  |  |  |  |  |  |
| Fixing Bugs and Making the Thesis Better |  |  |  |  |  |  |  |  |  |  |  |  |

# Appendix 2 – Concepts

## Economics of Crypto Currency

Supply and Demand are the basic variables on defining the price of an asset. Traditionally Banks issue currency and thus it gains value. Banks control the amount of money in supply by changing the Interest Rate and by changing the supply. Too much Inflation and Deflation is considered bad for a national currency. To keep stable amount of money in circulation and cope with The Business Cycle banks modify the Interest Rate and change the rate of printing money. (Cecchetti, 2000)

In Bitcoin and most of the crypto currency – the supply is predefined. It is not variable like the fiat currency. The supply of Bitcoin is 21 Million. New Bitcoin gets issued when miners mine a block. Currently the block reward is 12.5 Bitcoin. The reward halves at specific time frames. Till now it has halved twice. The Reward for mining a block is the incentive for miners (along with transaction fee) (Antonopoulos, 2014). Till now 16 Million Bitcoin has been mined. The difficulty adjusts itself to make sure that new block is created once every 10 minutes. Because of this and the halving process, the increase in supply decreases over time. Most coins have borrowed this idea from bitcoin and they have a supply that diminishes over time.

Traditionally Currencies are issued by a Central Bank, and the Central Government backs it. However, these crypto currencies can be created by anyone and till now – all crypto currencies have been created by the Private Sector. Bitcoin has the largest value – and one of the reason for it is First Mover Advantage. However, there is no fear of Fear of First Mover creating a monopoly in the crypto currency space because bitcoin poses no danger of monopolistic output restriction as the supply is pre-defined. It has been argued that unless Bitcoin can substantially differentiate itself from other coins, its value will decrease (White, 2015). The differentiation currently is that – To enter most of the currencies one has to buy Bitcoin First.

The Crypto Currency space has increased its value tremendously from its creation leading many to speculate Bubble at different times. Bubble is valuation anchored by nothing but expectation of market value. Big Innovations have had Bubble Phases. Dot Com Bubble of 2000 is argued to be like the crypto currency rise. The Internet eventually became a way of life, but investors got slapped with reality in between where a lot of companies were closed or neared bankruptcy. There are over a thousand crypto currencies and data suggests that most will lose value and only a few will last over a timeframe of years (Shiller, 2015).

Announcement and news have had huge effect in the price of crypto currencies. There are various times when addition of new features, addition to new exchange or marketplaces have caused huge price bumps. And because of the unregulated nature of these currencies, there probably has been a lot of Insider Trading before the news.

There are Options, Futures and Margin Trading for Bitcoin like other traditional securities. Margin Trading is the process of taking loans to buy a coin (long) or taking a loan of that coin to sell it and profit from a drop-in price – called short. Options is used to buy an option to buy a coin or sell it at the present price in future. It will result in high profit or losses. Futures contract are leveraged to gain (or lose) from price rises and falls in future.

However, there are not a lot of these options for altcoins. Poloniex, GDAX and Bitfinex provide Margin Trading for a selected few altcoins and although BitMex provide Options and Future Contract for some selected altcoins, the liquidity is low.

## Machine Learning

Machine Learning is used to make automatic decision based on Empirical Data. It is a field of Artificial Intelligence.

Neural Network(NN), Support Vector Machine (SVM) and Evolutionary Deep Network (EDN) are the most used concepts inside Machine Learning for this purpose. To perform Machine Learning, a set of data that can affect the outcome are first collected. These data are called features and it is crucial that these data are selected carefully as having a bad set of features can cause issues when predicting in future like Overfitting and Underfitting. Features are tested to see if they affect the data and removed/added depending on if the problem is High Bias or High Variance.

## Social Media

The use of Social Media has been rising over time. The use of word bots in this thesis refers to account that automatically interact with other users. Bots generate one fifth of entire conversation in Twitter (Bessi & Ferrara, 2016).

The possibility of using Social Media to manipulate public opinion has been discussed since at least 2006 (Howard, 2006). The use of bots to manipulate Social Media has recently been a hot topic due to the 2016 US Election. In Crypto Currency space – Reddit, Twitter and Facebook are the most active Social Media. Behavioral Finance indicate that psychology has huge effect on determining price of an asset. So Social Media is an important indictor to determine the sentiment and a crude place to attack with bots for manipulators.

## Quants in the Stock Market

The Standard & Poor’s 500 (S&P 500) is Market Index based on the Market Capitalization of 500 large companies listed in NYSE or NASDAQ. The average annual return for S&P index over the past 90 years is 9.8%. Hedge Funds aims to gain bigger or similar return for a lower risk. So, a consistent return over 10% is considered to be a great accomplishment. (Patterson, 2011)

Historically, financial advisors and fundamental investors have had the best returns. But currently Renaissance fund is the biggest player in wall street. Founded by a cold war era code breaker, it is known for using Super Computers (Quants) and mathematical algorithms to trade (Patterson, 2011). From 1994 through mid-2014 its averaged a 71.8% annual return (Burton, 2016). Its Medallion Fund is well known for performing betting in Bear as well as Bull Market. In the 2008 crash, Medallion rose 80% (Patterson, 2011).

The rise of the quants in Wall Street indicate that quants can perform equally well in Markets where they are not the leaders yet.

# Appendix 3 - Additional Terminologies

**Blockchain:** Blockchain is a list of records (blocks) where data can be stored one after another after crypto graphic verification. The Bitcoin Blockchain contains a list of all transactions performed in the Bitcoin Network.

**Options:** Options are option to buy or sell a security at the current price in a future timeline. They expire at an expiry date. Call Options are options to buy while Put Options are options to sell.

**Futures Contract:** Futures Contract are contract to buy or sell a amount of a security in future at the current price. Futures Contracts are usually leveraged to increase the profit or loss.

**Technical Analysis (TA):** The way of forecasting price through past market data. Exponential Moving Average (EMA), Moving Average Convergence Divergence (MACD), Relative Strength Index (RSI) are some popular TA method.

**Bots:** Bots is a short form of Robots. The use of the word bots in this thesis refers to automatic scripts.

**Quants:** Short for quantitative analyst. Quants are people who analyze an event by means of mathematical or statistical modelling.

**Insider Trading:** The process of trading a security with nonpublic information about the security.

**Arbitrage:** The process of taking advantage of price difference in different markets.

**Whale:** Entity or a group of entity big enough to change the price of a security by making a move.

**Short:** The process of opening a negative position in a security. The security is loaned and sold to perform a short. The negative position profits when price falls and vice versa.

**Long:** The process of opening a positive position in a security. Long position profit when security rises and vice versa.

**Bullish Market:** A market in which price is rising and is also expected to rise. **Bearish Market:** A market in which price is falling and is also expected to fall.