NUI Galway
OÉ Gaillimh

Blockchain Data Analytics Tool

College of Engineering and Informatics

Bachelor of Science (Computer Science & Information Technology)

Project Definition Document

Author:
Jakub Wojtkowicz (17451684)

Academic Supervisor:
Dr Matthias Nickles

Table of Contents

List Of Figures

Chapter 1. Introduction

1.1 Purpose of project

The purpose of this project is to create a provocative, useful and innovative data analytics tool. The tool aims to visually present the cryptocurrency Bitcoin, its current transactions and prices then superimpose up to date Covid-19 data on to it.

This design aims to discover the correlations between the pandemic and Bitcoin. The discovered corelations will be processed by a machine learning algorithm that aims to correctly predict the future prices and the optimum time to invest. The Covid-19 data will be based on a plethora of different variables such as confirmed cases, fatalities, and economic halts due to world-wide lockdowns.

1.2 Project Topic Description

"A blockchain is a digital record of transactions. The name comes from its structure, in which individual records, called blocks, are linked together in single list, called a chain. Blockchains are used for recording transactions made with cryptocurrencies, such as Bitcoin, and have many other applications." – Christensson (2018) [16]

The Blockchain is stored on a peer-to-peer network of independent computers. Each computer has a copy of the transaction ledger and each transaction is verified using a cryptographic hash which means falsification is almost impossible [7]. This allows for a safe way to deal with currency and that is why Bitcoin and other cryptocurrencies were created.

When an individual record, or block is created it is linked to other blocks which results in a blockchain. This record is validated by many independent computers before it is added to the chain. The blocks are connected together using a cryptographic hash created from the content of the adjacent blocks. This makes is very difficult to alter any blocks as all subsequent blocks must be altered too [13].

To add a block to a chain you must go through a process called mining. This involves doing complex mathematics to find a suitable hash that begins with a specific amount of zeros. Once you find the correct hash that block is considered forever signed unless it is re-mined. The nodes or computers on the network accept this change in the ledger and update their ledgers, the miner then receives a financial reward [14].

If a change needs to be made to a block in the middle of a chain that block and all subsequent blocks need to be re-mined to be approved by the network. This is very difficult to do as each hash has about four billion possibilities and requires very powerful computers.

Blockchain analytics involves the analysis of these transactions and how the ledgers are used. The analysis aims to discover useful information about who is using them and why. It is possible to map the spending history specific entities on the blockchain [15].

1.3 Project Topic Relevance

Blockchain is a very relevant topic as its expansion has some promising benefits but also some worrying issues. The increase in the use of blockchain has created a vast decentralised peer-to-peer network which has in turn spawned a liberated currency in its cryptocurrencies as no central authority controls the flow of them. This is a very secure and transparent system but allows anonymity when used in a certain way which has sparked issues with criminal activity [20].

With the current Covid-19 pandemic crippling the world and its markets [21], It would be quite interesting to see how the virtual world and its currencies are reacting to the changes currently taking place such as inflation and stock market crashes in the real world.

1.4 Document Structure

The following document will discuss the preliminary research conducted prior to beginning this project, the goals and requirements that have been set out in order to achieve the end result, the software deliverable outline, the software tools that will be employed to complete the project and finally the planned steps and milestones that should keep this development on track.

Chapter 2. Preliminary Literature Review

2.1 Existing Approaches

Bitcoin Analysis tools such as Coindesk [3], Kitco [22] and Business Insider [23] are existing approaches to analyse blockchain transactions. They share some features such as graph representation of price of the currency over a selected period of time, customisation of the data dashboard and currency snapshot displayed. The data is live and provides current up to date information.

They allow for fine tuning of the data displayed such as trends over a specific period of time such as day, week or month and physical currency conversion type. A metric section sits below the graph and gives detailed information such as value transacted, transaction count and average transaction fee.

The tool does not indicate any patterns present, predict or suggest prime investing times. It also does not support any superimposing of other data onto the graph.

2.2 Research

The research involved prior to beginning the project included reading many informative articles and explanations of what is and how blockchain works such as the one on Investopedia.com [7]. These in dept explanations provided valuable information on how blockchain works, what its used for and its perks and pitfalls.

This Material was very provocative and inspired more research into the crypto currency aspect of blockchain. Bitcoin, the most popular of the current crypto currencies was the most prevalent in the research conducted. Great websites exist on explaining Bitcoin and its innerworkings such as bitcoin.org [8] and provide a myriad of examples and use cases of Bitcoin.

Investigations were conducted into price fluctuations of Bitcoin at a high level and a lot of movement was discovered before and during the current pandemic. This lead to a belief that there must be some correlation between the two and that more research at a low level needs to be done.

After this interesting revelation, machine learning that could build a model based on previous trends and identify the correlations was explored. This could then allow the model to predict and advise on future investment. This model could also possibly inversely predict pandemic data such as incoming lockdowns as a result of a Bitcoin trend.

From research of machine learning and trend analysis, it was discovered that linear regression or neural networks would be suitable to use for trend prediction. These approaches were explained on sites such as upgrad.com [9]. Long Short-Term Memory Networks [17] also have become a potential candidate for model generation.

2.3 Innovative Approach

The project aims at building on current functionality of common Bitcoin analysis sites. The added functionality would include an overlay of covid-19 data to discover any potential correlations, potentially an overlay of the count of popular social media key word appearances and the introduction of machine learning to build a model to predict trends in the world of Bitcoin.

Values such as total transactions and total value of transactions would be important to focus on along with the price of Bitcoin. These values would be juxtaposed with values such as Covid-19 cases, deaths and current amount of people who are restricted in terms of their movement or in full lockdown. The algorithm would then build a model with the data collected and predict future trends based on current ones.

Chapter 3. Project Goals and Requirements

3.1 Primary Goals and Requirements

Primary goals of this project will be any vital functionality needed to create a viable innovative data analytics tool. This will be divided as follows.

A webapp that displays Bitcoin data as a graph and by other means will be the first primary goal. The webapp at this point needs to visually display the data and have options of displaying data between different time periods.

The second primary goal will be to allow the overlay of Covid-19 data on the original Bitcoin graph. It will be updated to show data between a specific time period similarly to the Bitcoin data.

The third primary goal will be the implementation of machine learning to indicate Bitcoin trend behaviours and predict the Bitcoin data. This will allow the program to advise the user on ideal investing and selling times.

3.2 Secondary Goals and Requirements

The first secondary goal will be to allow the overlay of other data on the graph such as the count of key sords appearing on social media such as “Covid-19”. This will allow the tool to compare the Bitcoin and Covid-19 data with popularity of certain things such as the pandemic in this case.

The secondary goal is to make the machine learning component more dynamic so that more computations can be made on it such as using the keyword count instead of the Covid-19 data to train a model to predict trends.

The third secondary goal will be to allow the user to input their past investments into the program and it will show its predicted price at a future time of their choosing. It will also show the percentage gain or loss of their investment.

Chapter 4. Planned Software Outline

4.1 Planned Backend Software

The first component to complete will be a Java webapp that will serve as the base of the project. It will be the hub of the project as all the APIs and the UI will be accessed from here. A Maven [10] webapp will be implemented that employs the Spring Boot [5] framework.

The logic of the webapp will be based here, This will include API input conversion for machine learning purposes, preparation of data to be displayed for the frontend and any computation required to calculate values such as future value of specific investments. The backend will also connect to any APIs necessary to run the app such as the Bitcoin API [2].

4.2 Planned Frontend and UI Software

The UI will be built using the Vaadin [6] framework which allows the building of frontend components using Java. This will allow for a cleaner and faster webapp as one language can be used.

The UI will consist of a central graph with components such as buttons that support interaction. The UI will allow the user to also view data in a numerical way. An input area will be available for the user to input past investments to the show the current or future value of it along with the profit or loss experienced.

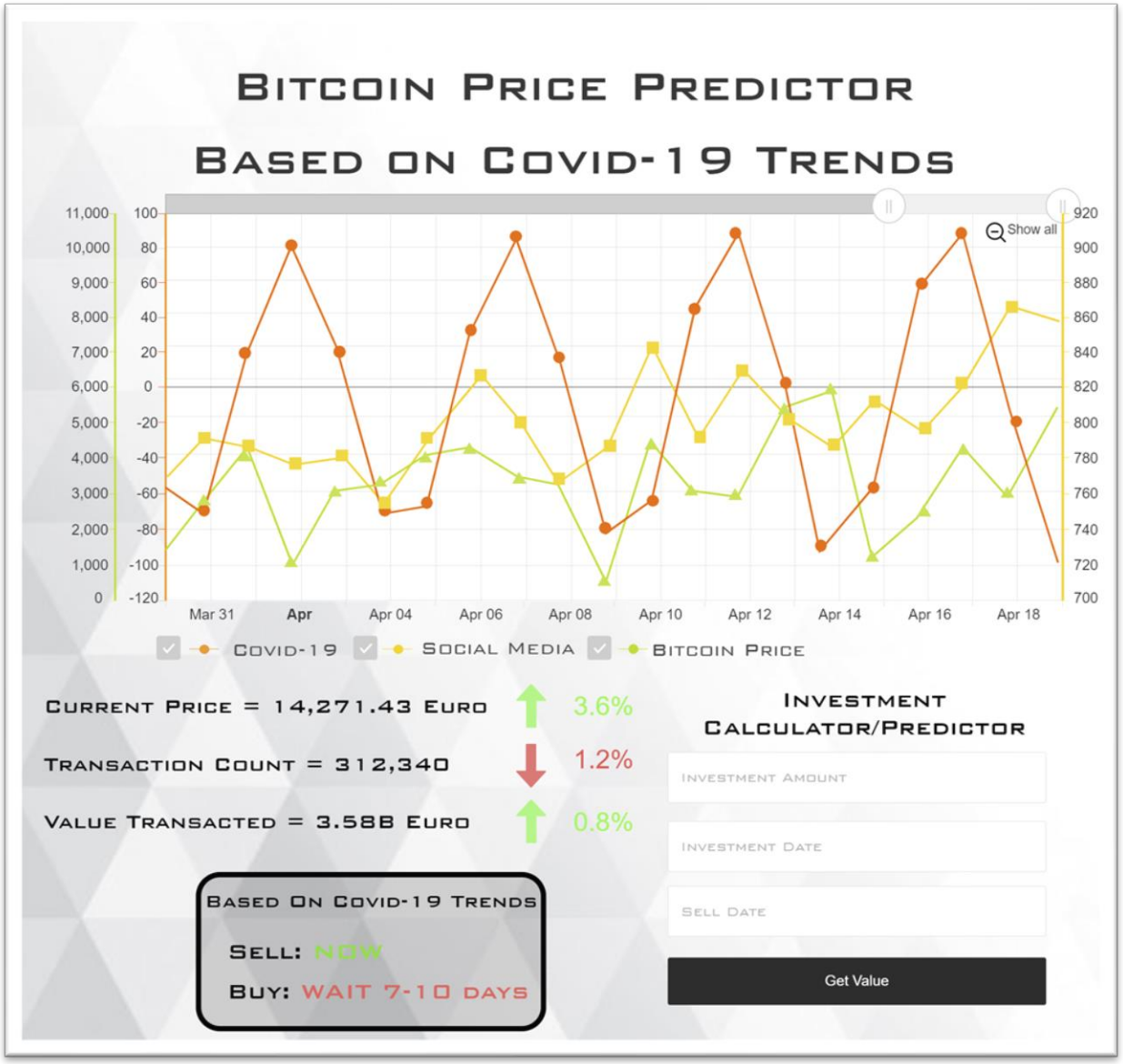

Figure 1: Initial Software User Interface draft

Chapter 5. Software Tools

5.1 Programming languages

The Software will be primarily written in Java. The UI and backend will accept Java APIs that may require other languages to operate such as python.

5.2 APIs

Live information will be retrieved using the Covid-19 API [1], Bitcoin API [2] and a to be decided API for web scraping key word searches and word appearances from social media. A machine learning API will be employed to create models and run data through them. In this project, the Smile [4] Tool will be used.

5.3 Planned Machine Learning Software

The machine learning will be done via an API. I will be using the Smile [4] machine learning tool. It provides a java API which will allow compatibility with my webapp. From initial research either linear regression [18] or a neural network [19] will be used to train my model. Long Short-Term Memory networks may also be an option to train the model. The approach will be decided once data flow is established from the main data APIs. [17]

5.4 Version Control

The project will utilise Github [11] to backup all project files and to keep a record of all code base changes. This will also aid in the recording of goals and milestones as they are reached.

5.5 Task Management

A task managing tool called Asana [12] will be used to set out tasks with due dates. This will keep the project on track and ensure that all primary goals and as many secondary ones as possible are reached.

Chapter 6. Planned Steps

6.1 Planned Steps & Milestones

1. Research Blockchain
2. Decide on project
3. Identify Tools and Technologies
4. Create Project Definition Document
5. Set up Github Repository
6. Create Webapp Skeleton
7. Create UI Skeleton
8. Connect Bitcoin API
9. Connect Covid-19 API
10. Display Basic Graph and data
11. Connect Machine Learning API
12. Set up Machine Learning Algorithm
13. Format data and send to Machine Learning API
14. Add Investment Calculator/Predictor UI
15. Add Ideal time to Buy/Sell Indicator
16. Add Customisation Currency/data displayed
17. Add Web Key Word Scraper
18. Adjust functionality to take in the new data

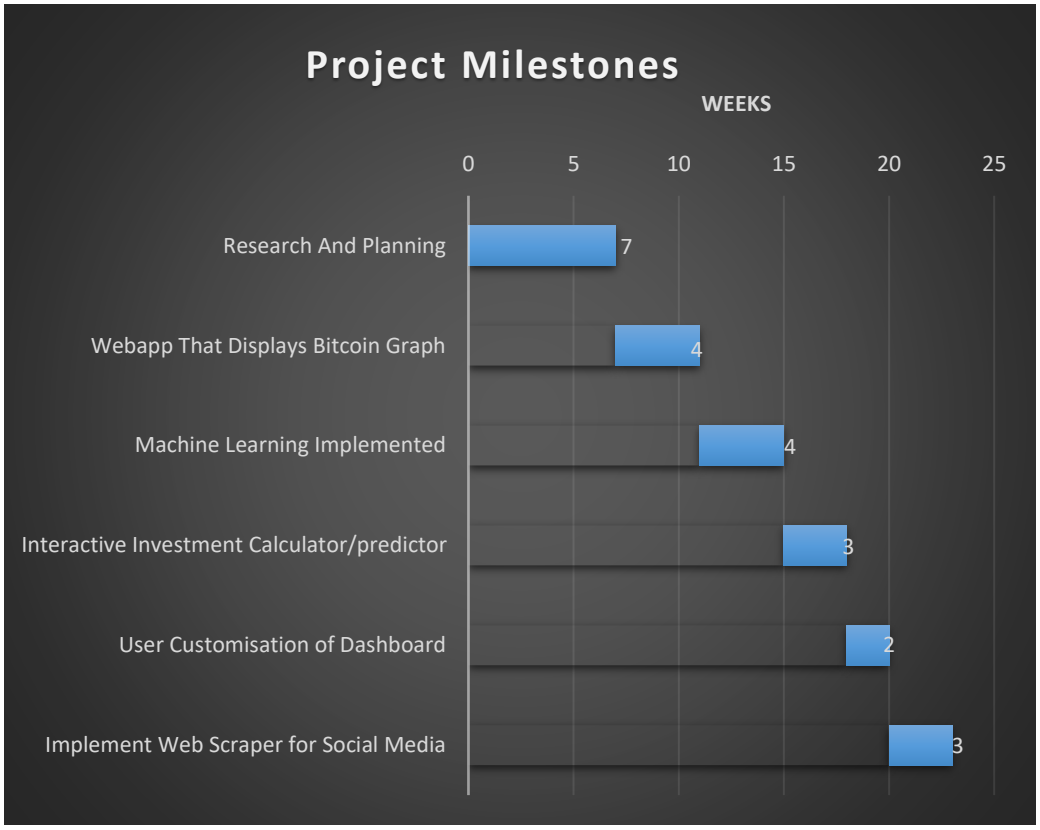

Figure 2. Project Milestone Gantt chart

References

[1] Covid-19 API. [Online] Available from https://covid19api.com/ Bitcoin price graph [Accessed November 2020]

[2] Bitcoin API (Coindesk). [Online] Available from https://www.coindesk.com/coindesk-api [Accessed November 2020]

[3] Bitcoin Analysis Tool (Coindesk). [Online] Available from https://www.coindesk.com/price/bitcoin [Accessed November 2020]

[4] Machine Learning API (Smile). [Online] Available from https://haifengl.github.io/quickstart.html [Accessed November 2020]

[5] Micro Service Framework (Spring Boot). [Online] Available from https://spring.io/projects/spring-boot [Accessed November 2020]

[6] Java UI (Vaadin). [Online] Available from https://spring.io/blog/2020/07/14/building-web-applications-with-spring-boot-and-vaadin [Accessed November 2020]

[7] Blockchain Explained (Investopedia). [Online] Available from https://www.investopedia.com/terms/b/blockchain.asp [Accessed November 2020]

[8] How does Bitcoin Work?. [Online] Available from https://bitcoin.org/en/how-it-works [Accessed November 2020]

[9] Machine Learning vs Neural Networks. [Online] Available from https://www.upgrad.com/blog/machine-learning-vs-neural-networks/#:~:text=Neural%20Networks%20are%20essentially%20a,in%20many%20fields%20of%20interest. [Accessed November 2020]

[10] Software Project Management (Maven). [Online] Available from https://maven.apache.org/ [Accessed November 2020]

[11] Version Control (Github). [Online] Available from https://github.com/ [Accessed November 2020]

[12] Task Management (Asana). [Online] Available from https://asana.com/campaign/fac/think?utm_campaign=Brand--UK--EN--Core--EX&utm_source=google&utm_medium=pd_cpc_br&gclid=Cj0KCQiAqdP9BRDVARIsAGSZ8An4W1ykUSQwC7rRpRdyKTU65eHcTMP6q4ekWi9gjylhQFokZz-gUDkaAkDJEALw_wcB&gclsrc=aw.ds [Accessed November 2020]

[13] Blockchain Technical Details (MIT). [Online] Available from http://blockchain.mit.edu/how-blockchain-works/ [Accessed November 2020]

[14] Bitcoin Mining (Investopedia). [Online] Available from https://www.investopedia.com/tech/how-does-bitcoin-mining-work/ [Accessed November 2020]

[15] Blockchain Analytics (Fintech News). [Online] Available from https://www.fintechnews.org/blockchain-analytics/ [Accessed November 2020]

[16] Blockchain Definition (TechTerms). [Online] Available from https://techterms.com/definition/blockchain [Accessed November 2020]

[17] Long Short-Term Memory Networks (Machine Learning Mastery). [Online] Available from https://machinelearningmastery.com/how-to-develop-lstm-models-for-time-series-forecasting/ [Accessed November 2020]

[18] Neural Networks (Wikipedia). [Online] Available from https://en.wikipedia.org/wiki/Artificial_neural_network [Accessed November 2020]

[19] Linear Regression (Yale). [Online] Available from http://www.stat.yale.edu/Courses/1997-98/101/linreg.htm [Accessed November 2020]

[20] Blockchain Pros and Cons (101 Blockchains). Available from https://101blockchains.com/pros-and-cons-of-blockchain/ [Accessed November 2020]

[21] Covid-19 and Stock Markets (Taylor & Francis Online). [Online] Available from https://www.tandfonline.com/doi/full/10.1080/20954816.2020.1757570 [Accessed November 2020]

[22] Bitcoin Price Tool (Kitco). [Online] Available from https://www.kitco.com/bitcoin-price-charts-usd/ [Accessed November 2020]

[23] Bitcoin Price Tool (Business Insider). [Online] Available from https://www.kitco.com/bitcoin-price-charts-usd/ [Accessed November 2020]