Extending the Possibility of Cryptocurrency Price Forecasting using a Mathematical Analysis of Predictive Modelling Tools

Trevor Kiggundu: 001001720
September 2021
MSc Computing and Information Systems
Supervisor: Dr Solomon Ebenuwa

A dissertation submitted in fulfilment of the requirements for the module, MSc Project, Computing and Information Systems Department, University of Greenwich.

ABSTRACT

Cryptocurrency has changed the digital payment market forever, and while digital currencies are not (yet) the mainstream way to exchange currency and complete transactions, it is clear to see where the market is headed, as most of our everyday processes have become digitalized and automated. It is only a matter of time. Despite their apparent popularity and relevance in the market, there seems to be no sure-fire method to accurately forecast future cryptocurrency prices, despite the existence of historical data. It is a very volatile market though, and many transactions and stock trades of this type are made purely by guessing, monitoring current data or simply, hoping to get lucky. However, as a mainstay of the current society, and certainly the future one, an entity as important as cryptocurrency, as well as its predictive abilities, should not be left up to chance.

ACKNOWLEDGEMENTS

I would like to thank everyone that was involved in the proposal, design, and eventual completion of this project. This includes, my supervisor Dr. Solomon Ebenuwa, using his expertise to support me throughout the difficulties I faced throughout the duration of the project: it was greatly appreciated. I would also like to thank my second marker, Dr. Ralph Barthel for his patience and guidance throughout the latter stages of the project cycle. Finally, I would like to thank all my loved ones for the continued support as I completed my final project.

Table of Contents

1 INTRODUCTION

The present global economy has been radically altered by various cryptocurrency technologies, such as blockchain for example. Never in the history of the financial sector has there been a paradigm shift comparable to that which blockchain and its associated digital currencies have brought to the world's economy. In January 2018, NASDAQ reported that Ethereum (ETH) prices had undergone a 13,500 % increase since the previous January, jumping from an average of 10 USD to 1,400 USD by the end of the year. The same phenomenon struck another popular cryptocurrency in Bitcoin (BTC) during the same period, as it ballooned from an average price of 1,150 USD to a justifiable 17,500 USD, a 1400% increase; however, as of September 2021, that price sits at a staggering 52,500 USD, and it only seems to be rising. These changes in price sent the stock market into a frenzy, and in turn started to beg a serious question: was this change in price and growth predictable? Digital currencies, although in their relative infancy, have repeatedly shown the capacity to challenge the financial status quo, and as a result, many of the world's economic powers are unable to envision what impact this new trend is going to have in the foreseeable future.

2 BACKGROUND RESEARCH

2.1 Problem Domain

Forecasting any mainstream form of currency is an extremely vital financial activity, even lauded by some as being the "most important task of quantitative finance" (Spilak, 2018). Tens of millions of stock market investors globally contribute to the market, all with the same goal in mind; to increase profits. Their combined efforts account for daily trading numbers within the region of "4.5 to 5 million" (Scutt, 2016) USD per day. This easily justifies the large amount of people that choose to spend hours monitoring these trends, waiting for the right moment to either buy or sell their shares, maximizing "their investment return" (Spilak, 2018). However, as much as stock management (cryptocurrency in particular) has become a mainstay in modern society, it also brings along a certain level of unpredictability that has been well documented throughout the years, with the most recent example involving GameStop (GME) and Dogecoin's (DOGE) boom in early to mid 2021.

This research is intended to produce a tangible and novel predictive modelling tool based on existing research, with an end goal of forecasting the future price dynamics of the major cryptocurrencies. To achieve this, the following aims have been identified and detailed below.

2.2 Project Objectives

The main objective of this project is to determine the best available machine learning method to predict cryptocurrency prices. This is much easier said than done, however. There are a variety of contributing factors that can influence the stock price of any given cryptocurrency, ranging from economic factors such as the quarter of the year that the market is in, to more trivial ones, such "social media factors" (Phillips et al, 2018) including memes and online forums such as Reddit. Identifying more of these factors is the second primary objective, as it will help to create a much more well-rounded analysis of the stock market.

Mathematical analysis is very important in this regard, as it helps to compute and provide statistics that can lead to much more accurate predictions about the future of influential entities such as the cryptocurrency market. The intended solution will operate with the price of the selected type of cryptocurrency being the dependent variable, and an undetermined value being the independent variable. A series of raw data will be used to encapsulate a sample of

the entire cryptocurrency market, and not just one cryptocurrency like ‘Bitcoin’ for example. This will help the consumer better understand the relationship between cryptocurrency price and the chosen variable and type of currency.

These primary objectives will also involve the analysis of various prediction methods based on different metrics, such as practicability and ease of use for the consumer. A comparative and explorative analysis of major cryptocurrency trends within the last 5 years will also be conducted, in order to determine whether the factors that affect normal currencies and stocks also affect cryptocurrencies. This will all be used in conjunction to help develop a web application that visually displays the raw data, time series data, and forecast data for every selected cryptocurrency. This application will also boast interactive user interface and experience features that make it easier for the user to view previous data and view future price predictions.

2.3 Projected Evaluation Procedures

Data collection is a crucial part of studies such as these, especially since the proposed application will be handling large amounts of real word data. Due to this, due diligence will be performed, and the appropriate legal, social, professional, and ethical guidelines will be implemented. After the completion of the product, a series of tests relevant to the software development process will be ran and compared to the projected objectives for the project. The final application will also be compared to an already available ‘market competitor’ to not only test its effectiveness numerically, but also to conclude whether the product developed is successful or not.

3 LITERATURE REVIEW

3.1 Introduction

This section aims to provide an extended look into the background research detailed in Chapter 2 of this document to make educated arguments about the topic, prior to the completion of the final product. All assumptions have been documented below and are supported by numerous academic journals that have detailed previous studies regarding the problem domain, and those that pose new questions for the future. There will also be an analysis of the academic journals that specifically focus on the various methods that have been implored to reach the most accurate forecast as possible, within the entire stock market. The topics covered in this section range from subjects such as economic factors that affect the stock market, to the deep and machine learning techniques used to remedy the problem. The academic journals found were collated from a variety of reputable sources, such as 'IEEE Xplore' and 'Google Scholar' etc.

3.2 Introduction to Cryptocurrencies

3.2.1 Cryptocurrency throughout the years

The term 'cryptocurrency' refers to any type of "virtual coinage system" that functions like standard everyday currencies, allowing users to provide "virtual payment" (Farell, 2015) of common goods and services. Cryptocurrencies have become extremely important in the modern financial market due to their flexibility, as their respective market does not operate within the binds of a 'central authority'. Instead, they created an unusual dent in the market in which some cryptocurrencies, such as Bitcoin, were able to completely decentralize the currency, freeing it from "hierarchical power structures" (Farell, 2015). The results of such actions are prevalent in the modern market, as Bitcoin is one of the highest grossing cryptocurrencies globally. This should not deter the consumer though, as the cryptocurrency market is also one of the most volatile in the world. As of today, there have been over 550 cryptocurrencies developed since their inception, though this study is only going to focus on a select five: Bitcoin, Bitcoin Cash, Litecoin, Ethereum, and Dogecoin.

3.2.2 Bitcoin (BTC) and Bitcoin Cash (BCH)

Bitcoin is the world's first "decentralized digital currency" (Frankenfield, 2021), first created in 2009 by an anonymous user going by the alias, 'Satoshi Nakamoto'. It is also the highest grossing cryptocurrency available on the market, with each 'bitcoin' commanding a staggering 52,500 USD per stock at the time of this report. Bitcoin's rules were designed with no apparent regard for "lawyers and influencers" (Böhme et al., 2015) as they strived to solve two challenges of digital currency at once: "controlling its creation, and avoiding its duplication" (Velde, 2013). Bitcoin's biggest selling point, however, can also be a cause of concern for the highly lucrative digital currency. Lack of standardization means that a lot of activity within the Bitcoin market is unregulated, and often taken advantage of by the black market, with many "individuals and organizations coming under attack" (Cheah et al., 2015) due to threats such as malicious software and hacking. It is also known to have a fork in the industry in terms of another very popular currency, Bitcoin Cash.

3.2.4 Litecoin (LTC)

Litecoin is another popular cryptocurrency in the market, first created in 2011 as "an early Bitcoin spinoff" (Nast, 2013) funnily enough. Though it has never seen its stock price rise to the astronomical numbers that its predecessor did, it still has found ways to mirror Bitcoin's success. Litecoin boasts a decentralized framework as well and has seen large price growth during the same period that Bitcoin did, with prices having "escalated by 7,291%" (Jana et al., 2019) throughout January 2017/18. Litecoin also faces the same "security issues" (Azad, 2017) that other major cryptocurrencies have encountered, and there are even arguments supporting the "inefficiency of Litecoin" (Jana et al., 2019). Other financial economists, however, believe that Litecoin might one day come to take over its 'bigger brother' based on a few different metrics, such as payment processing times.

3.2.5 Ethereum (ETH)

Ethereum is the fourth decentralized, open source blockchain cryptocurrency on this list, "second only to Bitcoin in market capitalization" (Vigna, 2021). Founded by Vitalik Buterin, a former Bitcoin employee (ironically), Ethereum is the only digital currency in the market that has come close to mirroring Bitcoin's heights in terms of stock price, sitting at just over 4,000 USD per stock. Ethereum is also comparable to the two of the three

currencies, as it too experienced a hike in prices within the 2017/18 calendar, seeing a price increase of over 18,000% in only 12 months.

3.2.6 Dogecoin (DOGE)

The newest cryptocurrency out of the five listed, Dogecoin also boasts a decentralized framework; this seems to be a mainstay within all modern digital currencies. It was originally founded as a joke by Billy Markus and Jackson Palmer, but neither of them could have predicted the growth that the currency would undertake due to its "sheer peculiarity" and "cultural relevance" (Young, 2019). Despite its 'cult favorite' status, Dogecoin did not reach its peak in terms of market capitalization until early to mid 2021, when it peaked and continues to grow past 85 million USD.

3.3 Economic Factors that affect Cryptocurrency

There are a variety of 'economic factors' that influence the cryptocurrency market. A study by Poyser (2017) regarding the determinants of Bitcoin's price found there to be four main factors separated into two main headings: internal and external factors.

Internal Factors		**External Factors**	
Supply and Demand	**Crypto Market**	**Macro-financial**	**Political**
-Transaction Cost -Reward System -Mining Difficulty (Hash Rate) -Coins Circulation -Forks (Rule Changes)	-Attractiveness (Popularity) -Market Trend -Speculations (Social Media Forums etc.)	-Stock Markets -Exchange Rate -Gold Price -Interest Rate	-Legalization (Adaptation) -Restrictions (Ban)

Table 1- Poyser's Factors that Influence Cryptocurrency Prices

3.4 Problems regarding Cryptocurrency Stock Price Prediction

Creating and forecasting predictive models for even the least popular of cryptocurrencies has become a challenge in the world of Finance, both in the industry and in academia, because of the abnormal behavior of stock prices. Stock prices do not follow the

conventional ‘time series’ sequence that is used for various other types of data analysis, splitting opinions between economists about whether or not stock prices are, at least, partially predictable based on historical trends. As a result, most day traders a taking a huge risk, as a study by the U.S. Securities and Exchange commission concluded that “70% of traders lose money every quarter on average”. The same study also found that most of those traders also lose “100% of their money within 12 months” (Benzinga et al., 2021). It is an extremely volatile market, that is potentially life changing, both in a positive and negative way.

3.5 Legal, Social, Professional and Ethical issues and considerations (LSPEi)

There is a large array of lingering legal, social, professional, and ethical issues that can arise whilst developing a prototype web application and creating financial modelling tools. This section aims to discuss some of these issues. Only the most relevant issues will be mentioned below.

3.5.1 Legal and Ethical Issues

Many of the legal, ethical, and professional issues regarding web application development are similar, funny enough. A common yet often overlooked is the act of copyright infringement, in which a user knowingly/unknowingly uses third party content without referencing it. This can especially cause issues if the content owner finds out that the work is being used “without their permission” (Copyright.gov, 2021). Special considerations, such as citing references and checking sources should be done to avoid these issues. There is also an ethical element involved regarding creating simulation tools for price forecasting. Regarding the cryptocurrency forecasting application, it is important to convey to the user of the system that this is just a simulation, and the tool should not be used to handle real world predictions and data as they could cause serious financial losses.

3.5.2 Social and Professional Issues

The majority of social issues regarding web elements surround the idea of accessibility and making sure that the software can be used by anyone. It is the responsibility of the developer to conduct themselves as a proper IT professional and ‘make IT for everyone’, per the BCS guidelines.

4 EXISTING SYSTEMS RESEARCH

4.1 Introduction

This chapter aims to outline and review some of the already existing systems that are already available to the public. This is important, as it not only provides a fair comparison between the existing systems and the one that is to be submitted, but it also allows for considerations to be made based on the design and functionality of these already existing products. A relevant quality assurance plan can also be created using these findings, as comparisons and contrasts between the 'competitors' and the proposed product can outline potential problems that might be otherwise overlooked. This will help to implement potential improvements in the final product. There was a wide range of existing products online, and considering the time constraints for the given project, three products were selected for review:

Product Name	Website URL
Yahoo! Finance	https://finance.yahoo.com
Polygon.io	https://polygon.io
Coin Dance	https://coin.dance
Digital Coin	https://digitalcoinprice.com

Table 2- Existing Systems

4.2 Yahoo Finance

Yahoo Finance is a website that provides financial news, data, and commentary via a variety of technologies online. It is the most practical of the 3 examples hidden, as it does not require the user to create an account, nor are any of its features hidden behind a paywall or subscription system. The website offers commentary about topics including, but not limited to, stock quotes, press releases, and even some original content from Yahoo itself. The website also offers tools and information regarding personal finance management for users that would find useful as well. Yahoo Finance provides an interactive experience for consumers, by not only listing historical data, market salaries and changes in trends in text, but by also providing a large array of interactive graphs, models, and charts. These are important, as they help the user better visualize the data, and allow for them to customize the content that they want to see without searching via numerous text/CSV files. The website

also allows the user to view financial data dating back as far as the inception of the stock/company. The most relevant aspect of the Yahoo Finance website, however, is the ability to view forecast data for the selected stock, or in this case cryptocurrency.

Appearance:

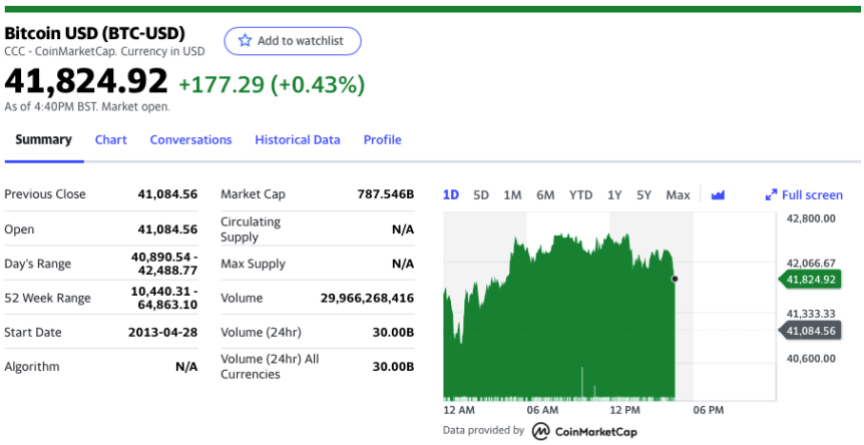


Figure 1 – Yahoo Finance

4.3 Polygon.io

Polygon.io is a company and web application that aims to provide consumers with the most up to date market data regarding stocks, currencies, forex, and cryptocurrencies. The most recent version of the application boasts features that allow the consumer to view API data from a variety of crypto, forex and stock exchanges. Polygon.io is not as user friendly as Yahoo Finance, as it does not provide graphs and/or interactive charts for users to view data. Instead, they provide all the data in a variety of formats so that the user can easily port and integrate it into the IDE of their choice.

Appearance:

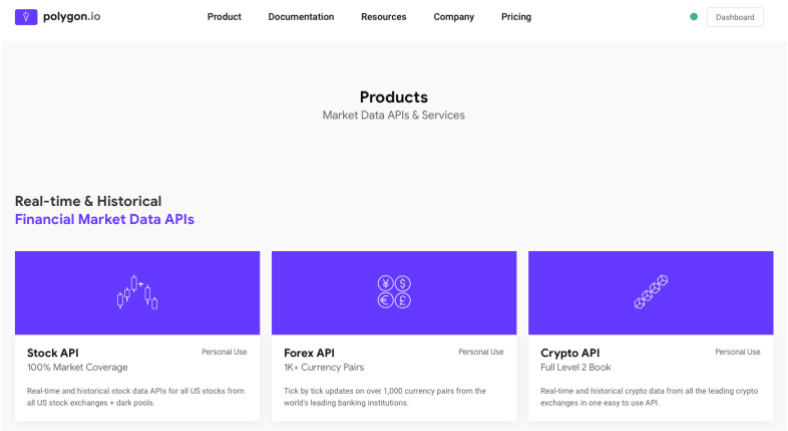


Figure 2 – Polygon.io

4.4 Coin Dance

Coin Dance is a third-party web application that aims to regularly provide both updated current and historical statistics about the cryptocurrency market. The Coin Dance website in run by a small community of ‘platform members’, of whom multiple have the responsibility of making certain that information such as trading volumes and blockchain statistics are available. Coin Dance boasts a large array of interactivity features in addition to the data it provides, including multiple visual aids such as graphs and charts that focus on specialized data, such as the ‘Hash rates per network’. Lastly, the website is free to use and updated frequently as part of the website’s goal to remain community driven.

Appearance:

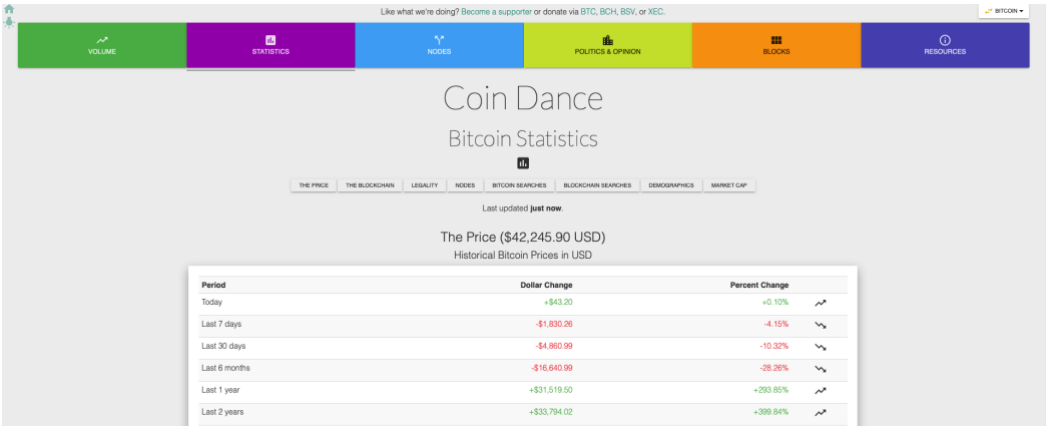


Figure 3 – Coin Dance

4.5 Digital Coin

Digital Coin is another third-party web application that aims to regularly provide both updated current and historical statistics about the cryptocurrency market, functioning very similarly to Coin Dance. There does not seem to be too much information about the service on the website, though they do claim to be one of the “trusted brands within the crypto community”. Digital Coin employs an extensive number of tools, calendars, and portfolios in addition to providing cryptocurrency advice and weekly press releases. Most importantly however, is Digital Coin’s ability to not only display historical data, but also forecast stock prices in the future.

Appearance:

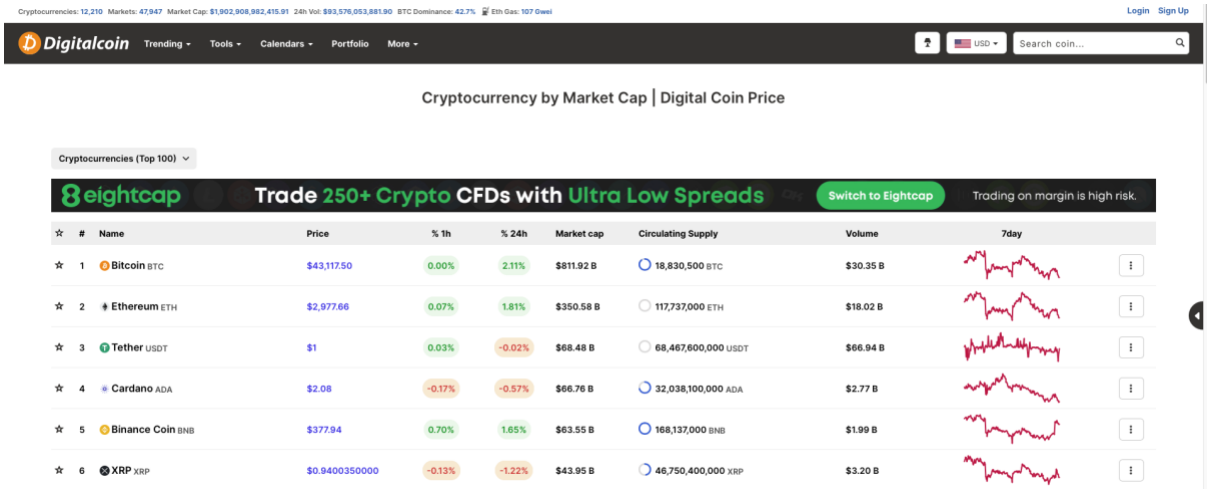


Figure 4 – Digital Coin

4.5 Adaptations from existing systems

The review of the products was critical to the project, as it helped to define the system requirements that are listed in chapter 5 of the paper. This also resulted in the creation of a web application that made sure to avoid the negative features of the existing systems, while also striving to implement the best features from all three websites. These features include but are not limited to: the accessibility of Yahoo Finance and Coin Dance, the high levels of normalized and specified data that Ploygon.io provides, the interactivity of the UI features used by Yahoo Finance and Coin Dance, and forecasting capabilities of Digital Coin.

4.5.1 Accessibility: Yahoo Finance, Coin Dance

Yahoo Finance and Coin Base are by far the two most convenient websites to use, as they both do not require user registration to access the market information displayed on the website. Users can visit the webpage, browse the information on the page, find and download the files they are looking for, and exit the website immediately without any restricted access or flags from the respective websites. This makes these websites much more accessible than some of the other products on the market, as the user can access useful information without user registration. This is also an attractive feature to users that might not want to create an account for whatever reason. Applications like Polygon.io and Digital Coin, however, do not provide this level of accessibility to its users, let alone the public. Registered users are capped at a minimum regarding the information they can access and are

subjected to ‘premium memberships’ behind a paywall to access information that ‘normal users’ cannot. This brings a bit of irony into the mix as well, as Polygon.io’s slogan is to “help developers build the future of FinTech by democratizing access to the world’s financial data”. This was seen as potential negative, and a decision was made to make the final product as accessible as possible, like Yahoo Finance and Coin Dance.

Example of premium memberships:

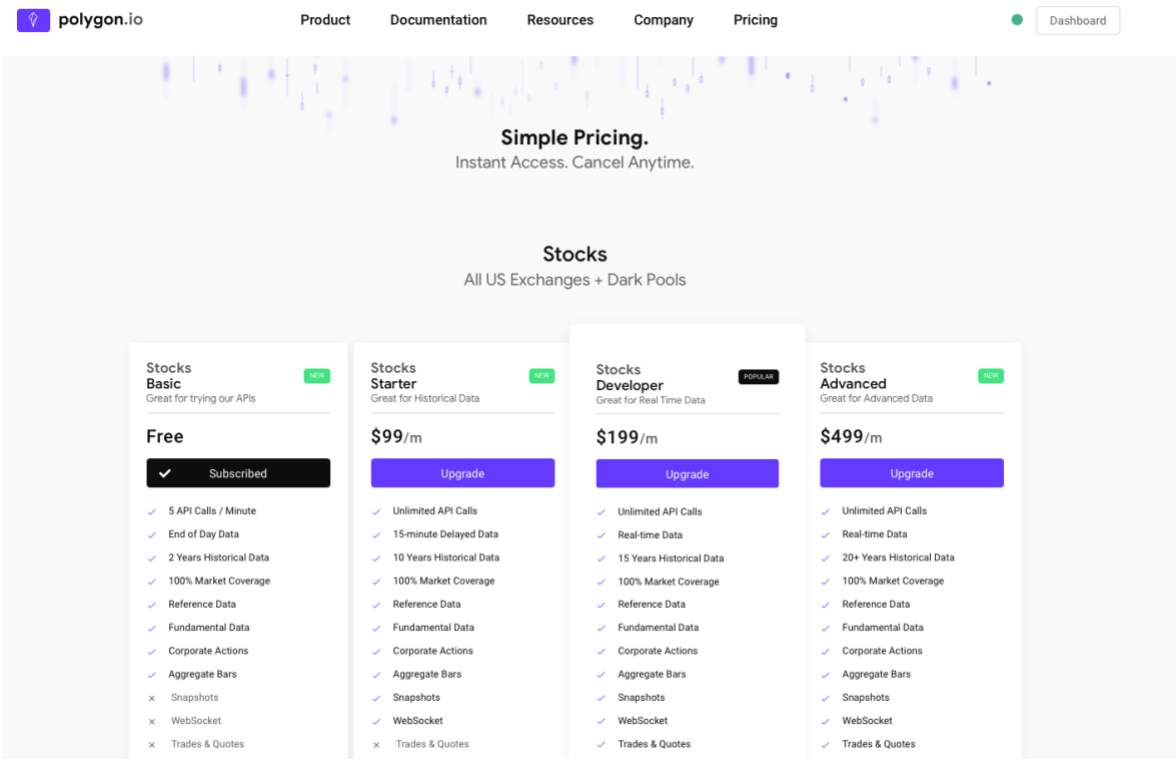


Figure 5 – Coin Dance

4.5.2 Interactivity: Yahoo Finance, Coin Dance

All the websites featured high levels of interactivity while keeping the user displaying the desired market data to the user. However, Polygon.io falls slightly behind its other two competitors as it does not provide graphs and charts for users to visually see the trends in market data in real time, opting for users to create their own applications. The additional of interactive charts that can be scaled up and down only separates the products even more. Due to this, a decision was made to create an application that mirrored a mixture of both Yahoo Finance and Coin Dance’s features, adding interactive graphs for the most important data, such as forecast data, and regular graphs and charts for the less relevant yet equally as intriguing values, such as component data. Yahoo Finance ended up being the

most relevant choice to model the final product after, as Coin Dance only shows Bitcoin values.

4.5.3 UX/UI designs: Yahoo Finance, Coin Dance, Polygon.io, Digital Coin

All the websites feature very visually pleasing user interface designs. However, once again, Polygon.io failed to live up to the user experience offered by the other two competitors. Then again, Polygon.io was unable to be fully explored due to the membership system, so an assumption can be made that these additional extra features may or may not be accessible behind the paywall. Yahoo Finance made a great effort to boost the UX by utilizing a consistent UI layout regardless of the webpage that the user was navigating on. Viewing stock data for different currencies had to be done on different web pages, although the new information was displayed in the exact same location as the previous currency. This was taken into consideration when designing the UI for the final product. Coin Dance took a slightly different approach, opting to display almost every single metric for the given currency available on a single page, complete with interactive charts and graphs. While this was greatly appreciated by the product developer, it did seem like there was too much data on a single Coin Dance page, a potential turn-off for some users. There was also a question regarding the relevance of some of the data shown; being able to view the different demographics of traders was interesting, thought it might not be relevant to a person trying to make an important decision regarding buying or selling their stocks. Based on these observations, a decision was made to only show the most relevant metrics on the final web application.

4.5.4 Cryptocurrency Prediction: Digital Coin

Though it is not as well-known and as trustworthy as the other sites (all sites should be used with digression), this website is one of the only ones on the market that lets the user view a large majority of its forecast data without subscription to content behind a paywall. This is important for the application, as one of the primary objectives of the product is not only to list and plot historical data, but also to forecast it in a way that is easily digestible for the user. A decision was made to also use this website as a comparison for the final values calculated within the final product.

5 SYSTEM REQUIREMENTS

5.1 Introduction

This chapter aims to discuss the considerations made for the proposed price prediction web application. Defining these requirements in crucial to the design process, as it allows for all components of the system to be accounted for and used efficiently. The requirements discussed below were curated from various relevant sources, including previous assignments, the software development process, academic journals and studies, and the existing systems reviewed in the previous chapter.

5.2 Analysis of Requirements

To follow the conventional software system design guidelines, the system requirements have been separated into two categories: functional and non-functional requirements. Functional requirements are critical "assets" (Easterbrook et al., 2008) and processes that the system should be able to undertake. the Non-functional requirements refer to the "important constraints upon the development and behavior of a software system" (Cleland-Huang et al., 2007), especially those defined by the functional requirements. These requirements are the direct correlation of the software development process, as they mirror the 'product plans' and 'process description'. The 'MoSCoW' method will also be used as a tool for task prioritization; it is an acronym that outlines the 'MUST' haves, 'SHOULD' haves, 'COULD' haves and 'WOULD' haves of the system. The MoSCoW method is important, as it helps to further define "the most important requirements, in what order to develop them, and what not to deliver" (Beltman, 2013) in case of any contingencies. Shown below are the functional and non-functional requirements for the product.

5.2.1 Functional Requirements

Functional Requirement		MoSCoW Category
1.	Interactive and user inclusive UX/UI design	SHOULD
2.	User ability to run the program via CMD/Terminal	MUST
3.	User ability to select cryptocurrency via dropdown menu	SHOULD
4.	User ability to view raw dataset	MUST
5.	User ability to view time series data on graph	SHOULD

6.	User ability to interact with time series graph	COULD
7.	User ability to view forecast data	MUST
8.	User ability to view forecast data on graph	MUST
9.	User ability to interact with time series graph	SHOULD
10.	User ability to view extended forecast components	COULD

Table 3- Functional Requirements

5.2.2 Non-functional Requirements

Non-functional Requirement		MoSCoW Category
1.	Usability: The application should have a modern user interface design, on par with similar websites.	SHOULD
2.	Ease of Access: The application should have a consistent layout, and all aspects should be labelled to increase user experience satisfaction.	SHOULD
3.	Interactivity: The application should include interactivity features to help the user better assess the data shown.	COULD
4.	Accessibility: The application must be accessible by all users, regardless of operating system or browser preference.	MUST
5.	Reliability: The application should be reliable and display data that is shown to be from a reliable source.	MUST
6.	Performance: The application should perform well and avoid flaws such as slow loading times.	SHOULD

Table 4- Non-Functional Requirements

6 SYSTEM DESIGN

6.1 Project Plan

Time management is an important aspect regarding the design and development of any project, let alone one involving software development. In order to highlight when each task in the process should take start and be completed, a GANTT chart is created. Considerations were made regarding the use of a PERT chart, but these were quickly debunked as they do not offer the same level of detail and timeline management as GANTT charts do.

6.2 Dataset

Two different sources of data were used as testbeds for the study, as there were two models created during the development of the final product. More on this topic will be detailed in chapter 7 of this report. There was a large array of datasets to choose from, as stock data is collected every minute of every day in the calendar year.

However, a decision was made to use a cryptocurrency dataset that detailed the 4 most popular cryptocurrencies in 2018; Bitcoin, Bitcoin Cash, Litecoin, Ethereum. A decision was made use the dataset from 2018 as it was the most recent full year data available before the COVID-19 pandemic. This dataset was used to create the first prototype. A second set of test data containing the cryptocurrency prices for the 4 previously mentioned currencies, as well as Dogecoin was used for the second model, with the data dating back to January 2016, as this allowed for a bigger sample size (1 year vs. 5 years) to be collected to make much more accurate estimates, assumptions, and predictions.

6.2.1 Raw Dataset 1: 2018 CSV file

This dataset is curated and produced by *pythonrprogramming.net,* by the user ‘SentDex’. This dataset is already formatted in a comma separated values (CSV) file, so there are no integration issues. The dataset includes 4 major cryptocurrencies (Bitcoin, Bitcoin Cash, Litecoin and Ethereum) and includes the following components from left to right for each currency: ‘Time, Low Price, High Price, Open Price, Close Price and Volume’. The time values are listed in seconds, and the remaining components are listed in USD.

BTC-USD

1528968660	6489.549805	6489.560059	6489.560059	6489.549805	0.5871
1528968720	6487.370117	6489.560059	6489.549805	6487.379883	7.706374
1528968780	6479.410156	6487.370117	6487.370117	6479.410156	3.088252
1528968840	6479.410156	6479.419922	6479.419922	6479.410156	1.4041
1528968900	6475.930176	6479.97998	6479.410156	6479.97998	0.753
1528968960	6477.959961	6480	6477.959961	6480	1.4909
1528969020	6477.220215	6480	6479.990234	6477.220215	2.73195
1528969080	6477.220215	6480	6477.220215	6480	2.17424
1528969140	6479.990234	6479.990234	6479.990234	6479.990234	0.9031
1528969200	6477.259766	6479.990234	6479.990234	6478.660156	3.258786
1528969260	6478.649902	6478.660156	6478.660156	6478.660156	1.970352

Figure 6 – Raw Dataset 1

6.2.2 Raw Dataset 2: Yahoo Finance

This dataset is curated, produced, and monitored every second by *Yahoo Finance*. This dataset is friendlier to use than the 2018 CSV file, as it is easily available to view online, includes statistics and analytics, and most importantly, labels the data and utilizes the traditional date time format, something that the other file did not do. The data can also be downloaded as a CSV if need be. The following components are shown from left to right for each currency: 'Date, Open Price, High Price, Low Price, Close Price, Adjustable Close Price and Volume'. The date is listed in yyyy/mm/dd format, and the remaining components are listed in USD.

BTC-USD

Date	Open	High	Low	Close	Adj Close	Volume
2020-09-29	10712.462891	10858.939453	10665.344727	10848.830078	10848.830078	20459870042
2020-09-30	10845.411133	10856.52832	10689.670898	10787.618164	10787.618164	20759622010
2020-10-01	10785.010742	10915.84375	10493.552734	10623.330078	10623.330078	27178227816
2020-10-02	10624.390625	10662.813477	10440.311523	10585.164063	10585.164063	23127839259
2020-10-03	10583.806641	10614.091797	10527.978516	10565.493164	10565.493164	17094010304
2020-10-04	10567.919922	10700.791016	10531.342773	10684.428711	10684.428711	17546792792
2020-10-05	10688.03418	10804.000977	10646.443359	10804.000977	10804.000977	19385191766
2020-10-06	10799.77832	10803.456055	10565.197266	10621.664063	10621.664063	22264958686
2020-10-07	10619.803711	10687.268555	10591.963867	10679.136719	10679.136719	17634388529
2020-10-08	10677.625	10939.799805	10569.823242	10923.62793	10923.62793	21962121001

Figure 7 – Raw Dataset 2

6.3 Technology

This section aims to provide an oversight of the different technologies, regarding both hardware and software, that were considered for the successful completion of the proposed product.

6.3.1 Programming languages

Initially, the interim system was proposed to be developed using the Java programming language, using tools such as the ‘Weka API’ for interactivity and data visualization on the graphs/charts. The developer also has previous experience performing linear regression analysis using Java the Java environment, within the NetBeans IDE. However, after extensive research into the market, and recommendations by the project supervisor, a decision was made to transition the programming language from Java to Python version 3.9.7, due to the vast majority of libraries available, such as ‘Numpy’, ‘Pandas’, and ‘SKLearn’. The user must download Python, as well as install these libraries via the command line or terminal for the project to successfully run on their desktop.

6.3.2 Integrated Development Environment

Python is part of a select group of languages that employs no specific integrated development environment (IDE) to run, relying on only the command/terminal line to install the appropriate libraries for design and development, though there are numerous IDE’s, in the form of text editors that boast better usability features, such as colour coding text, auto indenting code, and flagging syntax errors. The python script can be easily executed in whichever IDE the developer chooses.

7 SYSTEM DEVELOPMENT

7.1 Introduction

This chapter aims to discuss the system development procedures and methodologies that would best fit the creation of the product, as well as the step-by-step timeline of the development process from the beginning to the end of the project lifecycle. A good amount of time was spent researching the existing systems detailed in chapter 4 of the document. This was imperative, as one of the primary objectives of the study was to provide an interactive web application for the user to view the raw and forecasted data for each given cryptocurrency. These existing systems were used as the blueprint for the design of the final product. There were also a large array of methods and programming languages that could have been used for development, but after some deliberation, a final decision was made to develop the system using Python, while also using Sublime Text as the code editor and IDE.

7.2 Systems Development Methodology

This chapter aims to give a further insight into the system development methodologies that were considered during the planning of the project. All four developmental models (‘waterfall’, ‘extreme programing’, ‘rapid action development (RAD)’, ‘spiral’ and ‘agile development’) were analyzed and compared to the project requirements regarding their fitness for purpose. The spiral and extreme programming models were immediately eliminated from contention given the nature of the project. The waterfall model was also eliminated from contention, as it is better suited for longer, larger scaled, and more expensive projects. The two most suitable models for the development of the product are quite similar in the fact that they are a perfect fit for low costing, informal and last a “short duration” (Kyriakidou, 2018). The RAD model slightly edges the agile model in this instance however due to its lack of reliance on user feedback/collaboration during the development of the project. Given the time and interactivity constraints, and nature of the project, a conclusion was made that the best development model to follow would the rapid action development (RAD) model.

7.3 Implementation

Two prototypes were created for the final product, with both web applications using different sets of test data to attempt to solve the task at hand.

7.3.1 Prototype 1

Prototype 1 was the most time consuming, as it was not only created from scratch, but also involved the use of the 2018 CSV dataset, which was not formatted in the most ideal way to be fair. It was also the developers first time using Python for development in any instance. Prototype 1 was an attempt at predicting cryptocurrency prices using Recurrent Neural Networks, with TensorBaord and Keras being used for plotting the graph data. Sublime text was used at the IDE for development, and the first crucial step of the process was to install the latest version of Python and install the main library, which was ‘Pandas’. Installing this library was important, as it allowed for the CSV data to be ported and easily read in the Sublime Text IDE. As mentioned before, the 2018 CSV file was not normalized, nor was it balanced or labelled correctly. To remedy this issue, a classification function was created to define the names of each of the test data, as well as separate the most important data.

Code Snippet:

```
“ ratios = ["BTC-USD", "LTC-USD", "ETH-USD", "BCH-USD"]
for ratio in ratios:
    print(ratio)
    dataset = f'crypto_data/{ratio}.csv' ”

“ df = pd.read_csv(dataset, names=['time', 'low', 'high', 'open', 'close', 'volume']) ”
```

The next step was to create sequences based on the training data that had been extracted from the raw dataset, as it provided a much more realistic dataset to try and predict cryptocurrency prices with. To do this, the training data had to be normalized just as the initial raw data was, and the sequences were then created.

Code Snippet:

```
“ sequential_data = []
    prev_days = deque(maxlen=SEQ_LEN)
```

```
    for i in df.values:
        prev_days.append([n for n in i[:-1]])
        if len(prev_days) == SEQ_LEN:
            sequential_data.append([np.array(prev_days), i[-1]])

    random.shuffle(sequential_data) ”
```

The third step was to balance the data once again, this time having to normalize the sequential data to fit the recurring neural network. The training data was successfully normalized, scaled, and then finally pre-processed in preparation for building and training the final RNN model.

Code Snippet:

```
“ train_x, train_y = preprocess_df(main_df)
validation_x, validation_y = preprocess_df(validation_main_df) ”
```

The final step in the process was to build the model; the TensorBaord and Keras modelling tools were essential to this task, as they could easily be installed using the terminal and aided with the UX/UI design of the final product. The data was also trained for a period of 10 epochs via the terminal, before being plotted on a TensorBoard localhost web application.

Code Snippet:

```
“ tensorboard = TensorBoard(log_dir=f'logs/{NAME}')
filepath = "RNN_Final-{epoch:02d}-{val_accuracy:.3f}.hd5"
checkpoint = ModelCheckpoint("models/{}.model".format(filepath, monitor='val_accuracy',
verbose=1, save_best_only=True, mode='max')) ”
```

After going through the arduous process of balancing the different types of data and feeding it through the RNN, a decision was made that the first prototype was not a satisfactory product to interactively predict cryptocurrency data. The model was also only best used for predicting whether the price of the given cryptocurrency was going to increase or decrease, as it predicted the future price of the given cryptocurrency with only 59-60% accuracy after training for 10 epochs. The model was also unable to show long term results, as the training for forecasts even a couple of hours in the future would have taken up a considerable amount of time and GPU usage. Due to this, a decision was made that a second prototype should be

developed using a more readily available dataset, with the aim of creating a better product than the first one.

7.3.2 Prototype 2

Prototype 2 was the basis for the final product and was also somewhat challenging to program since it was programmed in a completely different operating system. This was the end of the difficulties however, as newer, more robust imported libraries such ‘Yahoo Finance, Facebook Prophet, and Plotly’ provided much more seamless integration for Financial data than more primitive Python libraries. A decision was also made to use Streamlit for data visualization, as it handled all the front-end code that would ordinarily have to be done separately from the backend financial data code. Python was reinstalled on the new operating system, and Sublime Text remained as the functioning IDE due to its previously mentioned capabilities. The relevant libraries were also installed and imported into the python file.

Code Snippet:

```
“ import streamlit as st
from datetime import date
import yfinance as yf
from fbprophet import Prophet
from fbprophet.plot import plot_plotly
from plotly import graph_objs as go ”
```

The top half of the GUI was them designed using Streamlit’s ‘selectbox’ and ‘slider’ elements, as well as the title of the page. Streamlit proved to be easier to use than TensorBoard, as it does not require nearly as much detail to create the application. Streamlit also employed a large array of interactivity features such as buttons and drop-down lists and did so in greater abundance than TensorBoard.

Code Snippet:

```
“ st.title("Cryptocurrency Price Prediction")
stocks = ("BTC-USD", "LTC-USD", "ETH-USD", "BCH-USD", "DOGE-USD" )
period = n_years * 365 “
```

The Yahoo Finance was then loaded into the Streamlit web application via the Python script using the ‘ticker’ object and was displayed in a table on the page. The ‘Plotly’ import was also used to graph the raw data and create a slider so that the user could easily browse through the chart.

Code Snippet:

“ fig.layout.update(title_text="Time Series Data", xaxis_rangeslider_visible=True)
st.plotly_chart(fig)
plot_raw_data() ”

Finally, the ‘Facebook Prophet’ was used to train the raw data and forecast the future closing prices for the given crypto currency. Once again ‘plotly’ was used plot the forecast data created by the FB Prophet model. A regular graph forecasting other components similar to those seen on Digital Coin was also created.

Code Snippet:

“ st.write('Forecast Data')
fig1 = plot_plotly(m, forecast)
st.plotly_chart(fig1)

st.write('Forecast Components')
fig2 = m.plot_components(forecast)
st.write(fig2) ”

8 SYSTEM TESTING

8.1 Introduction

This chapter aims to outline and define the testing guidelines and matrices for the completed stock prediction application. Testing is an important part of the system development process, as it allows for the developer to assess whether the system requirements for the final product have been met. A study by Durelli et al. (2019) found that system testing within a test-driven development model is as important as designing and programming the system itself, as it probes into the "behavior of software systems to uncover faults". This is also beneficial for the developer, as these faults can also lead to a change in project scope, cost or timeline if not handled correctly. The same study, however, also highlights that testing procedures are often not easy determine and fund, as "most testing activities are complex and costly" (Durelli et al., 2019). Due to these findings, a decision was made to develop testing procedures that were both relevant to the product, simple enough to understand, and cost effective to the developer.

8.2 Analysis of Testing Procedures

A series of white box and black box tests were selected as a reasonable metric to define the premature success of the product. White box testing is a testing methodology in which the "test sample" (Gao et al., 2018) is known to the tester. Black box testing refers to the opposite, in situations which the tester in none the wiser. These tests are based on the functional and non-functional requirements defined in chapter 5 of the document. Shown below are the test cases created for both types of testing.

8.2.1 White box testing

Test Case		Expected Result	Actual Result	Success?
1.	Run the application	Direction to localhost page	Direction to localhost page	Yes
2.	View default cryptocurrency data	Successful display of raw data in table format	Successful display of raw data in table format	Yes

3.	Select specific cryptocurrency from drop down menu	Redirection to localhost page with new data	Redirection to localhost page with new data	Yes
4.	Use slider to change the years of prediction	Page reloaded with change reflected on forecast graph	Page reloaded with change reflected on forecast graph	Yes
5.	View training time series data	Current open and close prices shown on graph since 2016	Current open and close prices shown on graph since 2016	Yes
6.	Interact with time series data	Successful interaction with a specific time period via use of mouse and slider	Successful interaction with a specific time period via use of mouse and slider	Yes
7.	View forecasted cryptocurrency data	Successful display of forecast data in table format	Successful display of forecast data in table format	Yes
8.	View forecast components data	Current and future daily weekly and yearly prices shown on graph since 2016	Current and future daily weekly and yearly prices shown on graph since 2016	Yes
9.	Interact with forecast data	Successful interaction with a specific time period via use of mouse and slider	Successful interaction with a specific time period via use of mouse and slider	Yes

Table 5- White box testing

8.2.2 Black box testing

Test Case		Expected Result	Actual Result	Success?
1.	Run the application	Direction to localhost page	Direction to localhost page	Yes

2.	View default cryptocurrency data	Successful display of raw data in table format	Successful display of raw data in table format	Yes
3.	Select specific cryptocurrency from drop down menu	Redirection to localhost page with new data	Redirection to localhost page with new data	Yes
4.	Use slider to change the years of prediction	Page reloaded with change reflected on forecast graph	Page reloaded with change reflected on forecast graph	Yes
5.	View training time series data	Current open and close prices shown on graph since 2016	Current open and close prices shown on graph since 2016	Yes
6.	Interact with time series data	Successful interaction with a specific time period via use of mouse and slider	Successful interaction with a specific time period via use of mouse and slider	Yes
7.	View forecasted cryptocurrency data	Successful display of forecast data in table format	Successful display of forecast data in table format	Yes
8.	View forecast components data	Current and future daily weekly and yearly prices shown on graph since 2016	Current and future daily weekly and yearly prices shown on graph since 2016	Yes
9.	Interact with forecast data	Successful interaction with a specific time period via use of mouse and slider	Successful interaction with a specific time period via use of mouse and slider	Yes

Table 6- Black box testing

9 EVALUATION

8.1 Introduction

It is important for any developed system to be properly evaluated, as it helps to provide feedback that critically discusses whether the goals and requirements for the system have been met or not. It also helps to further identify any problems for future developments, just like the system testing does, as software is everchanging and there will always be "accidental difficulties" (Brooks, 1997) that plague the software development process. Developers can, however, limit the impact that these essential and accidental difficulties bring by evaluating the software in-house efficiently and producing reports, so that the "effectiveness" (Merino et al., 2018) of the software is not questioned when it is released to consumers.

8.2 Process Evaluation

This section aims to critically evaluate the processes that were developed in order to complete the final deliverable. A concise list of the most important processes has been curated, as well as a few processes that would have aided in the production of a better-quality product. Shown below are the process positives and negatives of the final study.

8.2.1 Positives and Negatives: Process

Positives	Negatives
Clearly defined project objectives, with supporting documents supporting the problem domain	Poor time management: Better time management would have yielding a better product and study
Comprehensive literature review, spanning a large array of relevant studies/topics	Development of varying prototypes: all prototypes were built from scratch, rather than an extension of a previous version
Critical review of previously existing real-world systems, along with plans for adaptation to the final product.	Lack of extensive black box testing due to the nature of the study.
Concisely defined functional and non-functional diagrams, based on previous studies and industry standards	

Development of adequate system design and development matrices	
Analysis of existing industry standards aiding in the implementation of relevant testing guidelines.	

Table 7- Process positives and negatives

8.3 Product Evaluation

This section aims to critically evaluate the final product that was developed as a result of the extensive research, system design and system development procedures detailed in chapters 1-7 of the paper. The product evaluation aids the white and black box testing as an evaluation metric, as it allows for the system to be thoroughly critiqued via less defines matrices. Shown below are the product positives and negatives of the final product.

Positives	Negatives
Responsive web application with validated features such as showing when data has been loaded	Third-party financial data via Yahoo Finance and Facebook Prophet
Accessibility features: All application can be accessed from multiple browsers without a paywall.	Third-party UX/UI developer in Streamlit: website designs, visual aids
Ease of Access: All of the application features are displayed on a single concise webpage.	Third-party UX/UI developer in Plotly: charts, graphs etc.
Interactivity features such as dropdown menus, interactive graphs, and charts	Lack of specialized version control system (Git, GitHub etc.)
Modern and visually pleasing GUI design based on modern examples	Product cannot be accessed/ran without prior knowledge of Python and its libraries
Adequate and proper use of Python libraries to ensure that acceptable UX/UI designs are used for the GUI.	The product may be tedious to use for some users, as it requires scaling the data using the slider for a better view.
Final web application offers accurate and reliable historical data, as well as forecast data based on the real-world trends.	

Clearly labelled buttons, headings, and graphs to aid the user when browsing the application.	
Ability to show the user both the forecasted open and close price, as opposed to the close price that many applications show.	

Table 8- Product positives and negatives

8.4 Results

8.4.1 Results Tables

Bitcoin (LTC) Price Prediction						
	1 Month Prediction		**6 Month Prediction**		**1 Year Prediction**	
Model Name:	**Open (USD)**	**Close (USD)**	**Open (USD)**	**Close (USD)**	**Open (USD)**	**Close (USD)**
Prototype Application	55.84k	55.99k	75.52k	75.71k	87.07k	87.26k
Digital Coin	N/A	57.89k	N/A	80.25k	N/A	74.56k
Difference:	N/A	**3.39%**	N/A	**5.99%**	N/A	**14.55%**
Accuracy:	**96.61%**		**94.01%**		**85.45%**	

Table 9 – Bitcoin Price Prediction

Litecoin (LTC) Price Prediction						
	1 Month Prediction		**6 Month Prediction**		**1 Year Prediction**	
Model Name:	**Open (USD)**	**Close (USD)**	**Open (USD)**	**Close (USD)**	**Open (USD)**	**Close (USD)**
Prototype Application	208.03	203.67	303.64	306.73	329.12	331.38
Digital Coin	N/A	208.93	N/A	271.92	N/A	290.37
Difference:	N/A	**2.58%**	N/A	**11.35%**	N/A	**12.38%**
Accuracy:	**97.42%**		**88.65%**		**87.62%**	

Table 10 – Litecoin Price Prediction

Ethereum (ETH) Price Prediction						
	1 Month Prediction		6 Month Prediction		1 Year Prediction	
Model Name:	**Open (USD)**	**Close (USD)**	**Open (USD)**	**Close (USD)**	**Open (USD)**	**Close (USD)**
Prototype Application	3.35k	3.35k	4.60k	4.61k	5.91k	5.87k
Digital Coin	N/A	4.23k	N/A	5.17k	N/A	5.82k
Difference:	N/A	**26.27%**	N/A	**12.15%**	N/A	**0.85%**
Accuracy:	**73.73%**		**87.85%**		**99.15%**	

Table 11 – Ethereum Price Prediction

Bitcoin Cash (BCH) Price Prediction						
	1 Month Prediction		6 Month Prediction		1 Year Prediction	
Model Name:	**Open (USD)**	**Close (USD)**	**Open (USD)**	**Close (USD)**	**Open (USD)**	**Close (USD)**
Prototype Application	550.60	564.99	807.77	808.09	939.92	941.02
Digital Coin	N/A	679.09	N/A	870.79	N/A	915.76
Difference:	N/A	**2.58%**	N/A	**11.35%**	N/A	**12.38%**
Accuracy:	**97.42%**		**88.65%**		**87.62%**	

Table 12 – Bitcoin Cash Price Prediction

Dogecoin (DOGE) Price Prediction						
	1 Month Prediction		6 Month Prediction		1 Year Prediction	
Model Name:	**Open (USD)**	**Close (USD)**	**Open (USD)**	**Close (USD)**	**Open (USD)**	**Close (USD)**
Prototype Application	0.32	0.32	0.42	0.43	0.57	0.58
Digital Coin	N/A	0.33	N/A	0.39	N/A	0.38

Difference:	N/A	**3.13%**	N/A	**9.30%**	N/A	**34.48%**
Accuracy:	**96.87**		**90.70**		**65.52**	

Table 13 – Dogecoin Price Prediction

8.4.2 Results Discussion

Short Term Forecast – 1 Month

As expected, the algorithm and web application showed the highest accuracy rates when plotting the short-term forecast. The one-month prediction data produced by the prototype application yielded highly favourable percentages when compared to the predicted prices generated by the Digital Coin dataset, with all of the cryptocurrencies save for one (Ethereum) boasting accuracy numbers in the mid to high 90's. This proved that the application created was reliable enough to make short-term predictions, as its final numbers were only a small fraction off the numbers available to the public. The prototype application predicted Ethereum's 1 month price with an accuracy rate of 73.73%, which is a tad bit misleading as the USD values were only 1,500 apart. While this separation in price might not be inherently catastrophic, especially compared to prototype 1 and other models on the market, it could be the difference between trader lucking out, or losing all of their money, depending on how many stocks they have invested into the currency.

Medium Term Forecast – 6 Months

A drop in close price accuracy, and an increase in price difference was expected for the mid-term predictions, especially as the algorithm did not have any recent data to make trend lines from. However, the developer was pleasantly surprised to discover that the six-month prediction data had also yielded highly favorable percentages in comparison to the Digital Coin dataset. All the currencies boasted accuracy numbers within the mid 80's to the mid 90's, and there were no apparent outliers in the accuracy data. Ethereum was still the lowest scoring in terms of accuracy, but it had much improved from the previous dataset, rising from 73.73% to an impressive 87.85%, a 19% increase. Bitcoin had the highest-level of accuracy percentage for this round of testing, leapfrogging Bitcoin Cash and Litecoin to a high of 94.01%.

Long Term Forecast – 12 Months

The largest drop in close price accuracy was predicted to happen within the long-term testing round of data collection, as this set of data was 2 times less exposed to the actual training data than the mid-term was, and a considerable 12 times less exposed than the short-term prediction. The data, however, showed otherwise, as the application was able to still retain numbers close to the short and mid-term forecasts in various metrics. All the currencies save for one (Dogecoin) boasted accuracy numbers within the mid 80's to the high 90's, with some currencies recording their highest accuracy percentage yet compared the Digital Coin dataset. Despite starting with the lowest accuracy value (73.73%) in the arguably the easiest metric (short-term, 1 month forecasting) Ethereum was able to boast an astonishingly high 99.15% accuracy rate for the 12-month forecast, the highest for all 3 metrics so far. The same fate did not befall Dogecoin however, as its accuracy rate dropped from a high 90.70% to a disappointing 65.52% percent. This figure is also slightly misleading though, as the prototype applications actual close price only dropped 20 cents to the dollar at 0.38 USD compared to Digital Coin's 0.58 USD. Nonetheless, this drop in price/accuracy could prove catastrophic for a shareholder owning multiple amounts of the given stock.

8.4.3 Final Discussion

A conclusion can be made that the prototype application was a success, as it consistently proved to compete with a real-world application in Digital Coin, whist only using data dating back to 2016, unlike the latter that uses all available cryptocurrency data thanks to the community that monitors the web page. The prototype application consistently boasted percentage difference numbers within region of 1-8%, with a few outliers producing larger margins of error. These outlandish changes in price valuation are justified however, as the cryptocurrency market itself is very volatile, and is subject to changes based on various factors.

10 CONCLUSION

10.1 Final Discussion

The finished project was initially brought to light as a potential solution for the difficulties the technology industry faces regarding forecasting cryptocurrency prices. The volatility of the market not only encourages risky behaviours, but also has the capacity to send the stock market into a frenzy due to many differing, almost always unrelated factors. Gone are the days in which these currencies could just be ignored or left to float around the market as 'hobbies', as the digitalization of every human processes was bound to take over the finance market as well. After considering many modelling tools, IDE's, and data sources, a dataset featuring current and historical data from Yahoo Finance was selected to be forecasted using the Facebook Prophet libraries to forecast the data. Plotly was then used to graph and scale the data on a Streamlit web application. The accuracy that was achieved using these models to predict and compare to real life models was astounding, and while nothing is 100% predictable, there is light at the end of the tunnel that there will one day be a solution to issue regarding cryptocurrency price prediction.

10.1.1 What Was Learnt

A better understanding of the stock market, let alone the cryptocurrency market was achieved throughout the duration of the project. A new affinity for script languages was also developed, as the developer had only interacted with object-orientated languages before commencement of the study. They are much easier to run, and given the right Python libraries and imports, can create a large array of projects and use cases. There was originally a concern about data visualization, as traditional IDE's such as Visual Studio are known for making the process easier for the user. However, these fears were quelled after researching and discovering the large collection of free, open-source libraries available to the user.

10.1.2 What Went Well

The project successfully met its goal, which was to produce a web application that forecasted the future price for every given currency. The project was also able to be programmed using multiple imports from different sources, and it can also run on different operating systems and IDE's. Another successful part of the project was choosing the correct tools for modelling, predicting and data visualization.

10.1.3 Room for Improvement

Future implementations of the application could include studying a larger set of test data, as the dataset used for the application only dated back to January 1st, 2016. It is possible that implementing older data into the system could have yielded better results, especially when considering the sharp price hikes that befell a large array of digital currencies between 2017 and 2018; these prices continue to rise until this day. Although the application was highly successful, it is possible that implementing a Recurrent Neural Network (RNN) such as the Long-Short Term Memory (LSTM) that was used for prototype 1 could have helped to yield better results, as the training data would have been more balanced and normalized as a result. Lastly, the same study and application could be used to analyse the more lesser known and grossing currencies, as these might show a higher level of volatility compared to the more sought-after ones.

11 REFERENCES

- Anon (n.d.) *Python Programming Tutorials*, [online] Available at: https://pythonprogramming.net/introduction-deep-learning-python-tensorflow-keras/ (Accessed September 30, 2021).

- Azad, 2019. Security Issues in Blockchain based Applications. *International Journal of Engineering and Advanced Technology*, 8(6S3), pp.890-896.

- Beltman, 2013. *Novel issues in user requirements analysis for a feedback system on energy consumption : a qualitative study into microgeneration, enterprises and the robustness of the MoSCoW-method - University of Twente Student Theses*. [online] Essay.utwente.nl. Available at: <https://essay.utwente.nl/64586/> [Accessed 30 September 2021].

- Benzinga, 2021. *If You're Day Trading, You Will Probably Lose Money: Here's Why*. [online] markets.businessinsider.com. Available at: <https://markets.businessinsider.com/news/stocks/if-you-re-day-trading-you-will-probably-lose-money-here-s-why-1030667770> [Accessed 30 September 2021].

- Böhme, R., Christin, N., Edelman, B. and Moore, T. (n.d.) Bitcoin: Economics, Technology, and Governance, *Journal of Economic Perspectives*, [online] Available at: https://www.aeaweb.org/articles?id=10.1257/jep.29.2.213 (Accessed September 30, 2021).

- Brooks. (1987). No Silver Bullet Essence and Accidents of Software Engineering.

- Cheah, E. and Fry, J., 2015. Speculative bubbles in Bitcoin markets? An empirical investigation into the fundamental value of Bitcoin. *Economics Letters*, 130, pp.32-36.

- Cleland-Huang, J., Settimi, R., Zou, X. and Solc, P., 2007. Automated classification of non-functional requirements. *Requirements Engineering*, 12(2), pp.103-120.

- Computer, 20(4), 10-19. doi:10.1109/mc.1987.1663532

- Durelli, V., Durelli, R., Borges, S., Endo, A., Eler, M., Dias, D. and Guimaraes, M., 2019. Machine Learning Applied to Software Testing: A Systematic Mapping Study. *IEEE Transactions on Reliability*, 68(3), pp.1189-1212.

- Frankenfield, J. (2021) Bitcoin Definition: How Does Bitcoin Work?, *Investopedia*, Investopedia, [online] Available at: https://www.investopedia.com/terms/b/bitcoin.asp (Accessed September 30, 2021).

- Haq, I., Khan, Z., Ahmad, A., Hayat, B., Khan, A., Lee, Y. and Kim, K., 2021. Evaluating and Enhancing the Robustness of Sustainable Neural Relationship Classifiers Using Query-Efficient Black-Box Adversarial Attacks. *Sustainability*, 13(11), p.5892.

- Jana, R., Tiwari, A. and Hammoudeh, S., 2018. The Inefficiency of Litecoin: A Dynamic Analysis. *Journal of Quantitative Economics*, 17(2), pp.447-457.

- Kyriakidou, A. (2018) COMP1632: Systems Development Project, COMP1632:

- Merino, L., Ghafari, M., Anslow, C. and Nierstrasz, O. (2018) A systematic literature review of software visualization evaluation, *Journal of Systems and Software*, Elsevier, [online] Available at: https://www.sciencedirect.com/science/article/pii/S0164121218301237?casa_token=hU3GNq75qKAAAAAA:IVpV6J-1sehiL8PkWSZkXHKfGNd5xK0uFs0r4gyD1vIfbpJchbwRz0t5voIV4Fvw2I5EuF9ijPQ (Accessed September 30, 2021).

- Nast, C., 2013. *Ex-Googler Gives the World a Better Bitcoin*. [online] Wired. Available at: <https://www.wired.com/2013/08/litecoin/> [Accessed 30 September 2021].

- Phillips, R. and Gorse, D., 2018. Cryptocurrency price drivers: Wavelet coherence analysis revisited. *PLOS ONE*, 13(4), p.e0195200.

- Poyser, O. (2017). Exploring the determinants of Bitcoin's price: an application of Bayesian Structural Time Series. Dissertation.

- Python-Engineer (n.d.) python-engineer/python-fun: Some fun and useful projects with Python, *GitHub*, [online] Available at: https://github.com/python-engineer/python-fun (Accessed September 30, 2021).

- Scutt, 2016. *Here's how much currency is traded every day*. [online] Business Insider. Available at: <https://www.businessinsider.com/heres-how-much-currency-is-traded-every-day-2016-9> [Accessed 30 September 2021].

- Soni, N., 2020. An Analysis of Cryptocurrency and Their Functioning. *SSRN Electronic Journal*,.

- Velde, 2013. *Bitcoin: A primer*. [online] Blog.philippe-poisse.eu. Available at: <http://blog.philippe-poisse.eu/public/monnaie_locale/bitcoin/cfldecember2013_317.pdf> [Accessed 30 September 2021].

- Vigna, P., 2021. *DeFi Is Helping to Fuel the Crypto Market Boom—and Its Recent Volatility*. [online] WSJ. Available at: <https://www.wsj.com/articles/defi-is-helping-to-fuel-the-crypto-market-boomand-its-recent-volatility-11622712602> [Accessed 30 September 2021].

- Young, I., 2018. Dogecoin: A Brief Overview & Survey. *SSRN Electronic Journal*,.

APPENDIX A: Similar Products

1 Yahoo Finance

Home Page

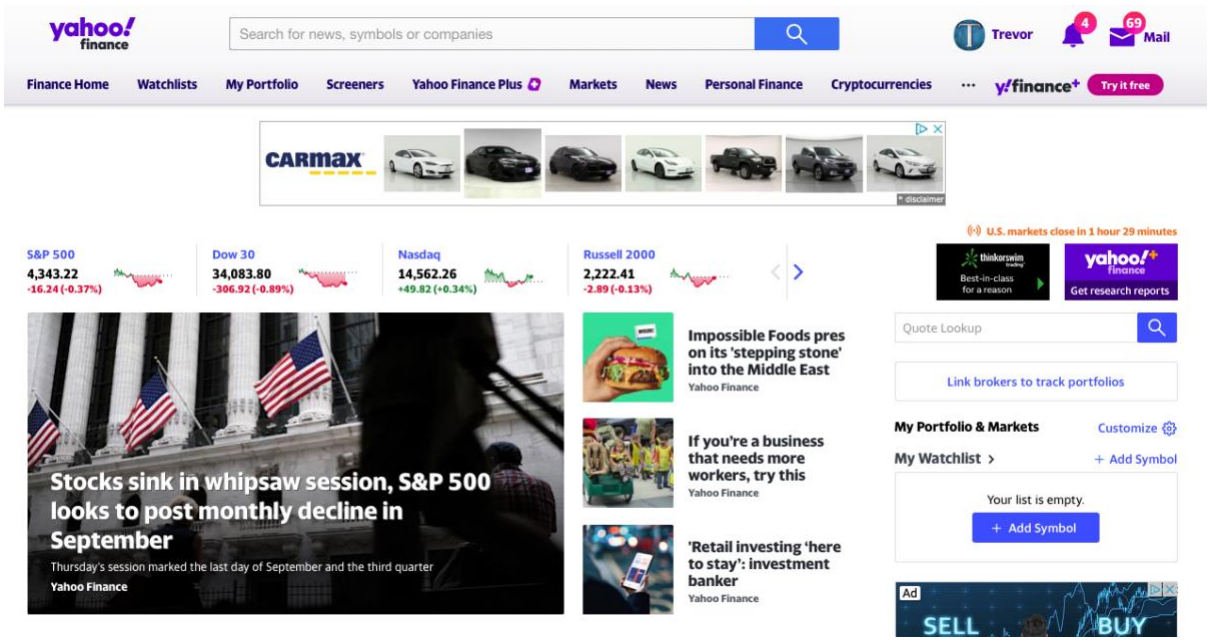
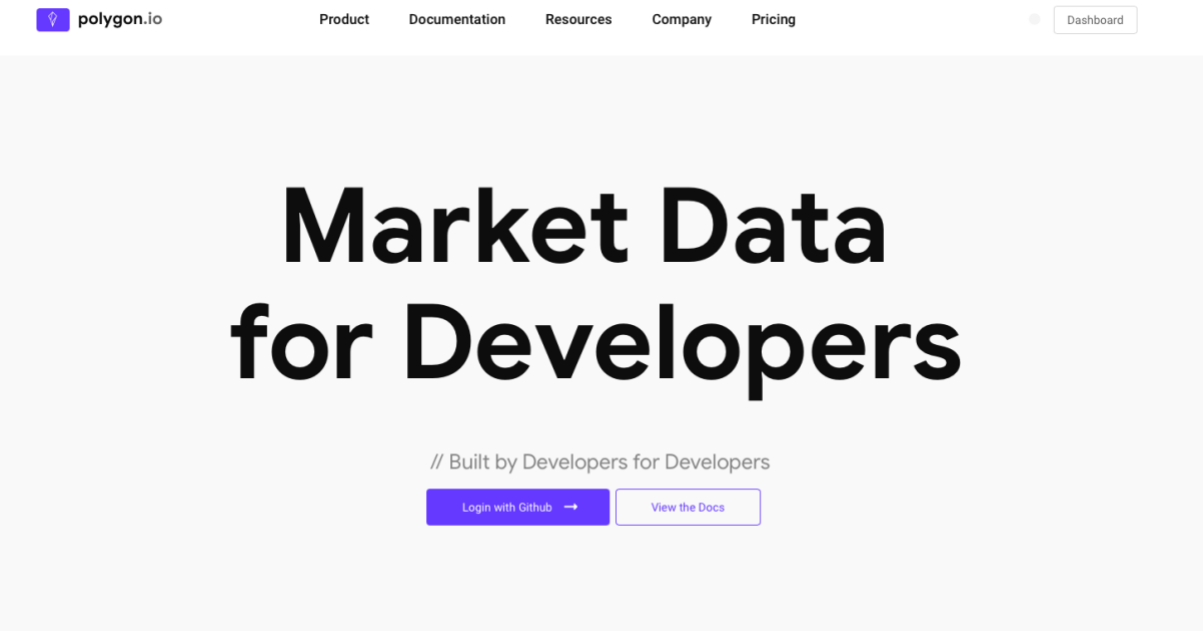


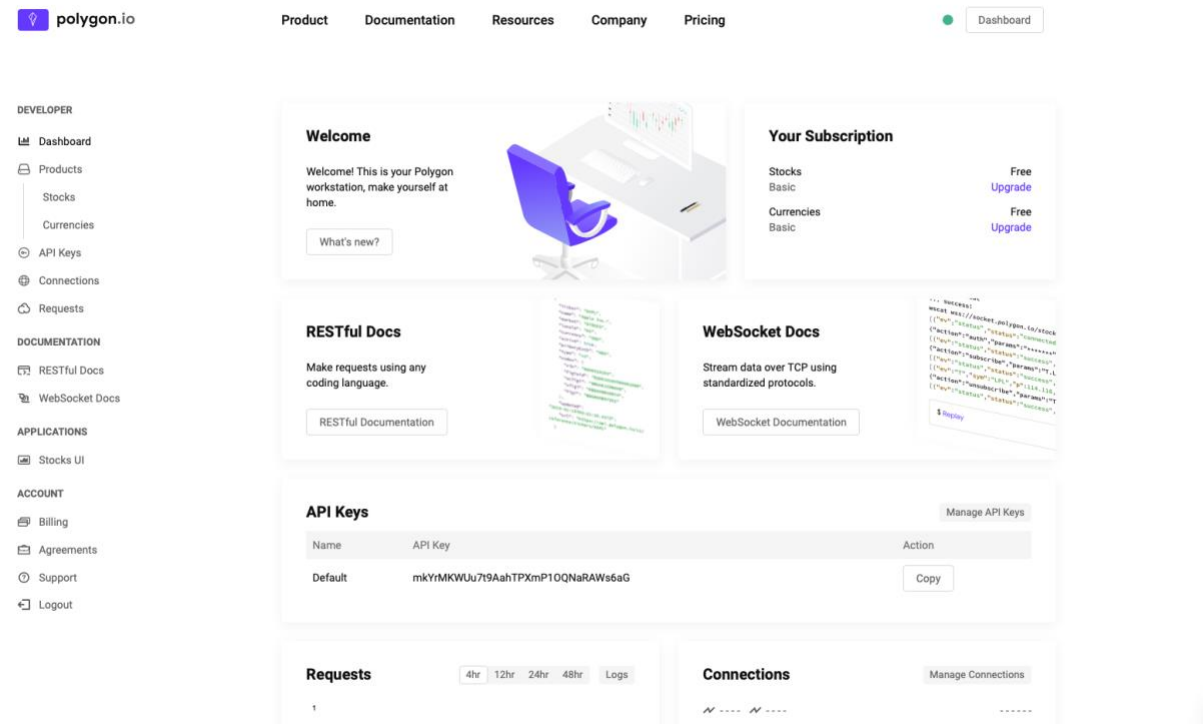
Chart Data

2 Polygon.io

Home Page

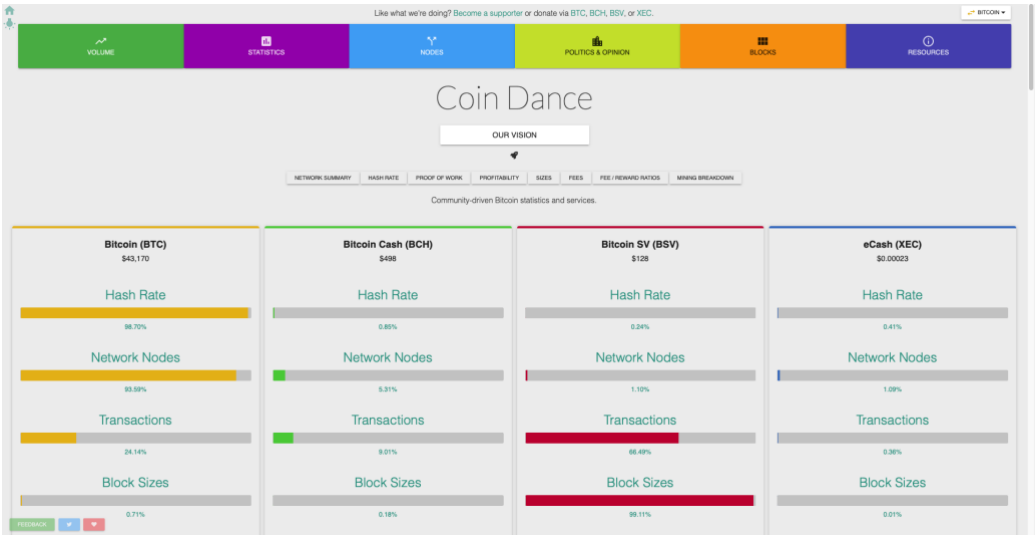


Dashboard

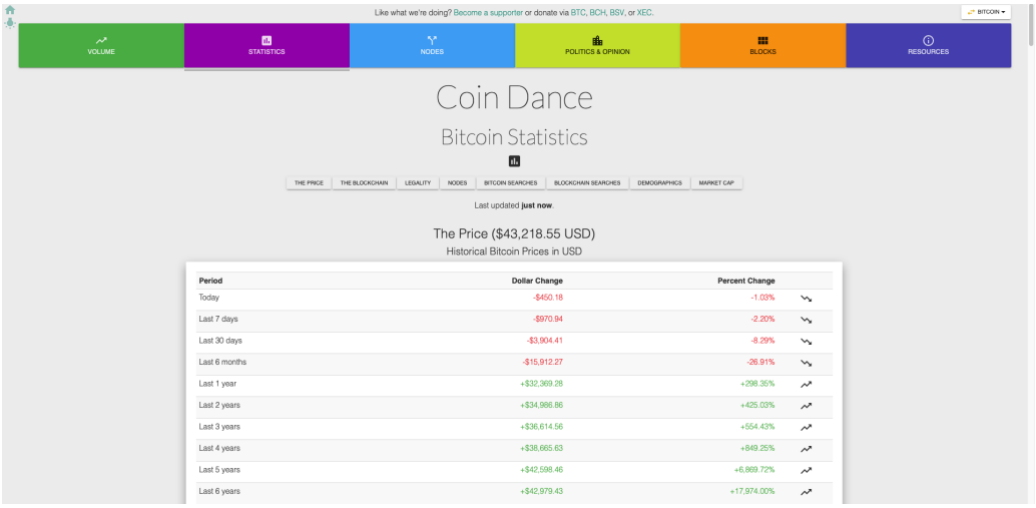


3 Coin Dance

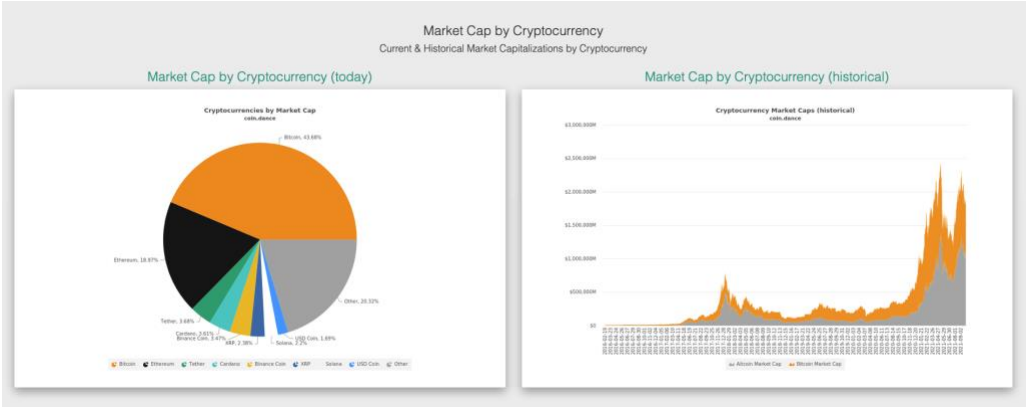
Home Page

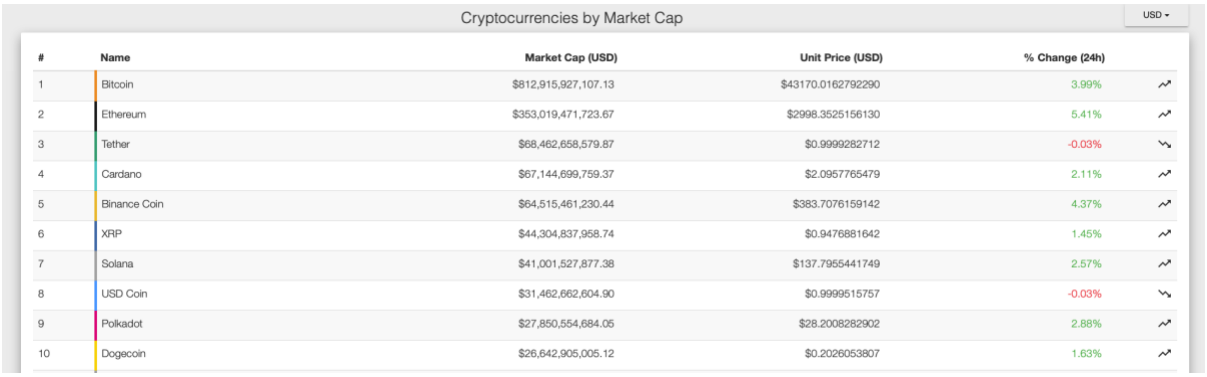


Dashboard



Additional Metrics





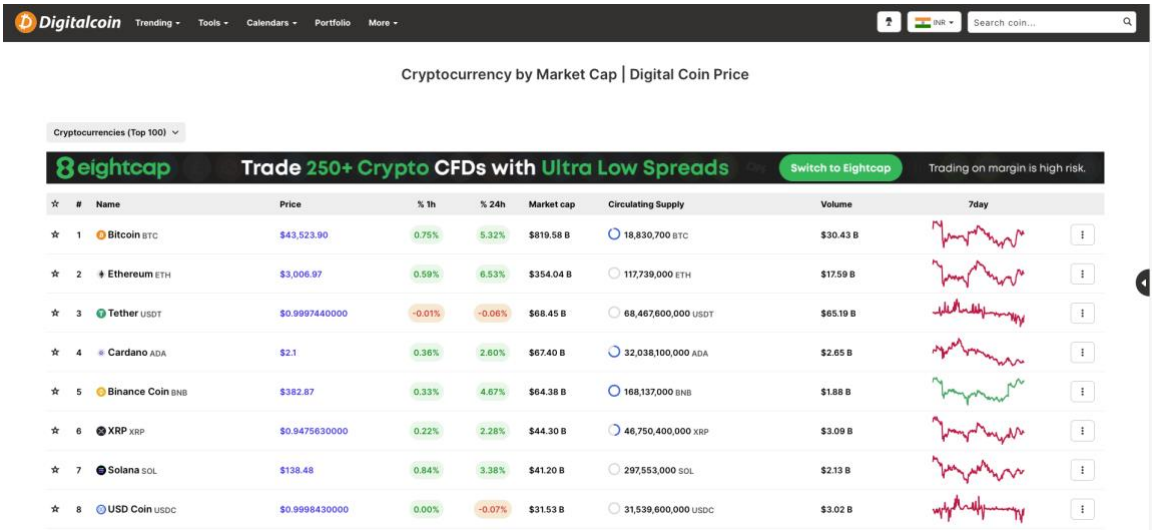
Cryptocurrencies by Market Cap

USD

#	Name	Market Cap (USD)	Unit Price (USD)	% Change (24h)
1	Bitcoin	$812,915,927,107.13	$43170.0162792290	3.99%
2	Ethereum	$353,019,471,723.67	$2998.3525156130	5.41%
3	Tether	$68,462,658,579.87	$0.9999282712	-0.03%
4	Cardano	$67,144,699,759.37	$2.0957765479	2.11%
5	Binance Coin	$64,515,461,230.44	$383.7076159142	4.37%
6	XRP	$44,304,837,958.74	$0.9476881642	1.45%
7	Solana	$41,001,527,877.38	$137.7955441749	2.57%
8	USD Coin	$31,462,662,604.90	$0.9999515757	-0.03%
9	Polkadot	$27,850,554,684.05	$28.2008282902	2.88%
10	Dogecoin	$26,642,905,005.12	$0.2026053807	1.63%

4 Digital Coin

Home Page



Digitalcoin Trending Tools Calendars Portfolio More

Search coin...

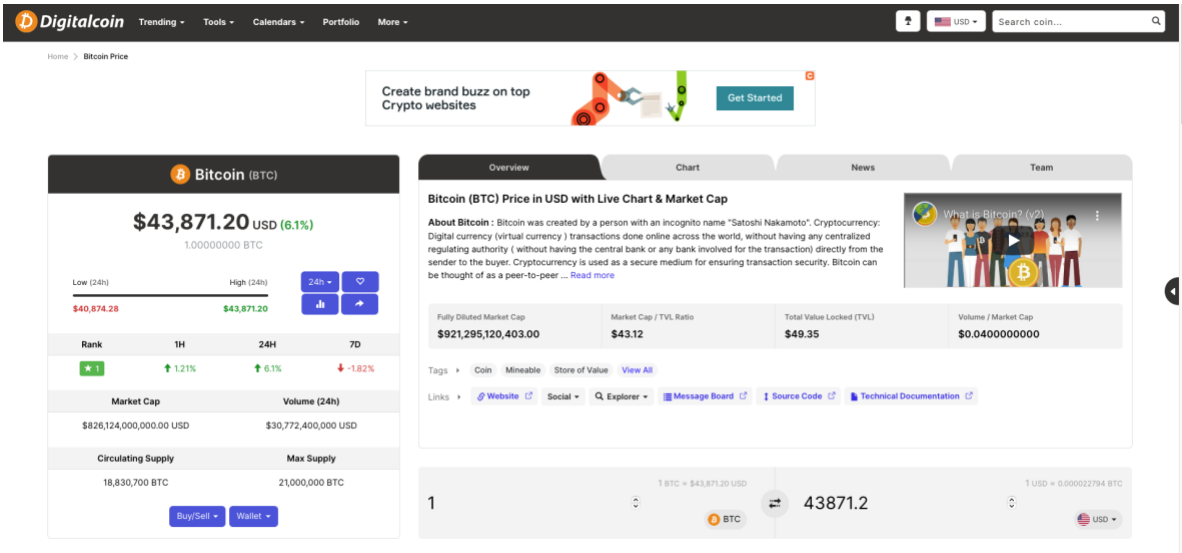
Cryptocurrency by Market Cap | Digital Coin Price

Cryptocurrencies (Top 100)

eightcap Trade 250+ Crypto CFDs with Ultra Low Spreads Switch to Eightcap Trading on margin is high risk.

#	Name	Price	% 1h	% 24h	Market cap	Circulating Supply	Volume	7day
1	Bitcoin BTC	$43,523.90	0.75%	5.32%	$819.58 B	18,830,700 BTC	$30.43 B	
2	Ethereum ETH	$3,006.97	0.59%	6.53%	$354.04 B	117,739,000 ETH	$17.59 B	
3	Tether USDT	$0.9997440000	-0.01%	-0.06%	$68.45 B	68,467,600,000 USDT	$65.19 B	
4	Cardano ADA	$2.1	0.36%	2.60%	$67.40 B	32,038,100,000 ADA	$2.65 B	
5	Binance Coin BNB	$382.87	0.33%	4.67%	$64.38 B	168,137,000 BNB	$1.88 B	
6	XRP XRP	$0.9475630000	0.22%	2.28%	$44.30 B	46,750,400,000 XRP	$3.09 B	
7	Solana SOL	$138.48	0.84%	3.38%	$41.20 B	297,553,000 SOL	$2.13 B	
8	USD Coin USDC	$0.9998430000	0.00%	-0.07%	$31.53 B	31,539,600,000 USDC	$3.02 B	

Dashboard



Digitalcoin Trending Tools Calendars Portfolio More

USD

Search coin...

Home > Bitcoin Price

Create brand buzz on top Crypto websites

Get Started

Bitcoin (BTC)

$43,871.20 USD (6.1%)

1.00000000 BTC

Low (24h) $40,874.28

High (24h) $43,871.20

24h

Rank	1H	24H	7D
1	1.21%	6.1%	-1.82%

Market Cap	Volume (24h)
$826,124,000,000.00 USD	$30,772,400,000 USD

Circulating Supply	Max Supply
18,830,700 BTC	21,000,000 BTC

Buy/Sell Wallet

Overview Chart News Team

Bitcoin (BTC) Price in USD with Live Chart & Market Cap

About Bitcoin : Bitcoin was created by a person with an incognito name "Satoshi Nakamoto". Cryptocurrency: Digital currency (virtual currency) transactions done online across the world, without having any centralized regulating authority (without having the central bank or any bank involved for the transaction) directly from the sender to the buyer. Cryptocurrency is used as a secure medium for ensuring transaction security. Bitcoin can be thought of as a peer-to-peer ... Read more

What is Bitcoin? (v2)

Fully Diluted Market Cap	Market Cap / TVL Ratio	Total Value Locked (TVL)	Volume / Market Cap
$921,295,120,403.00	$43.12	$49.35	$0.0400000000

Tags Coin Mineable Store of Value View All

Links Website Social Explorer Message Board Source Code Technical Documentation

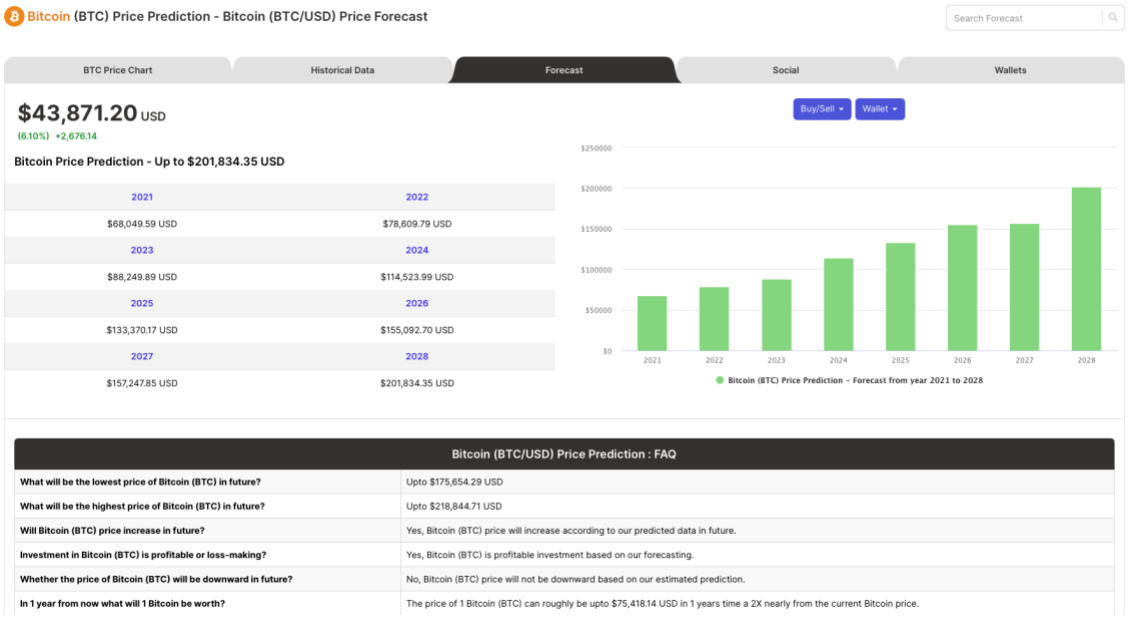
1 BTC = $43,871.20 USD

1 USD = 0.000022794 BTC

1 BTC 43871.2 USD

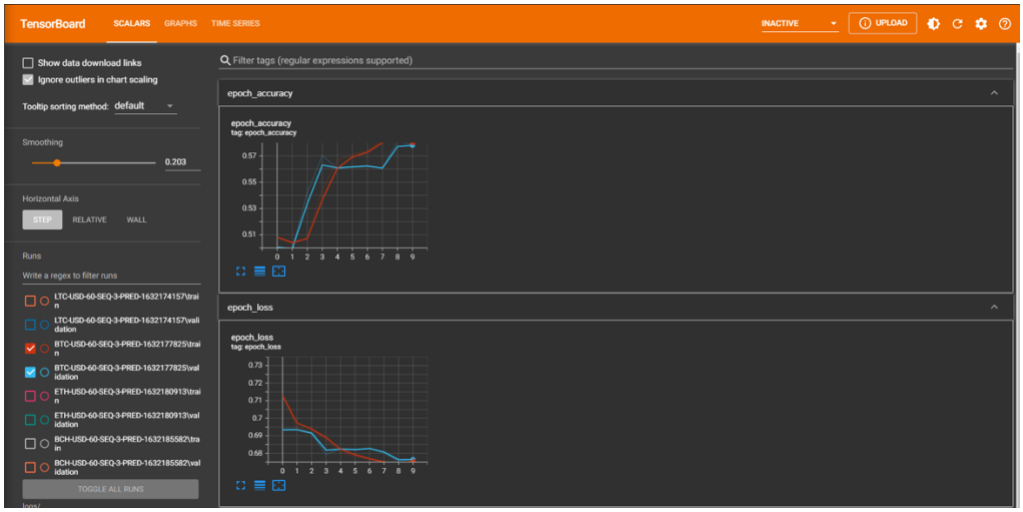
Additional Metrics

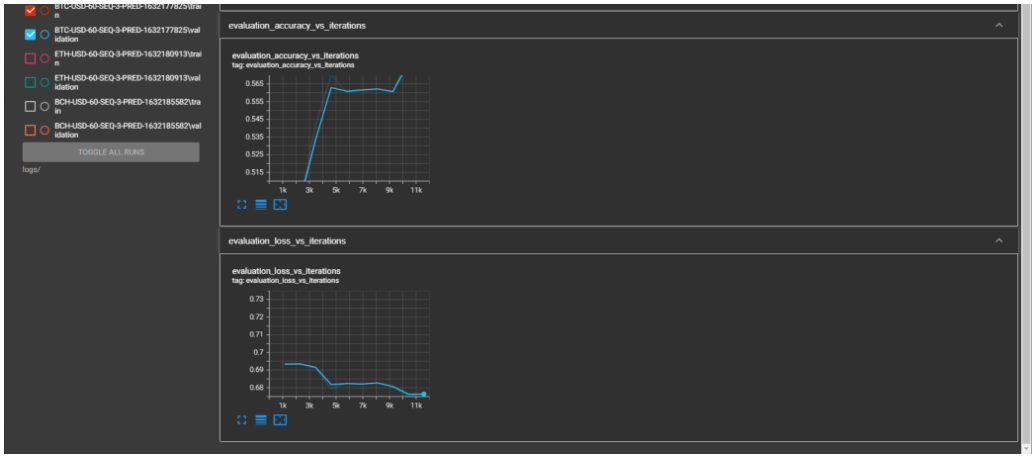
Forecast Data

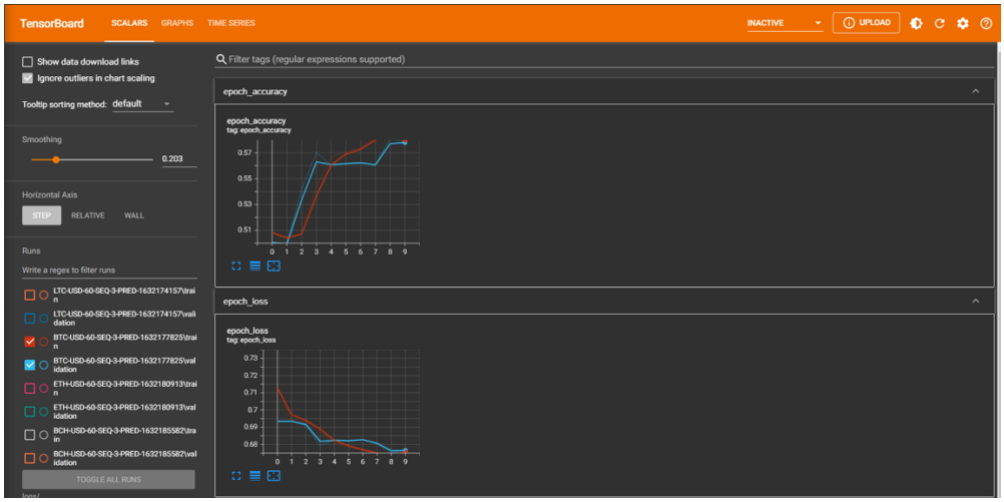


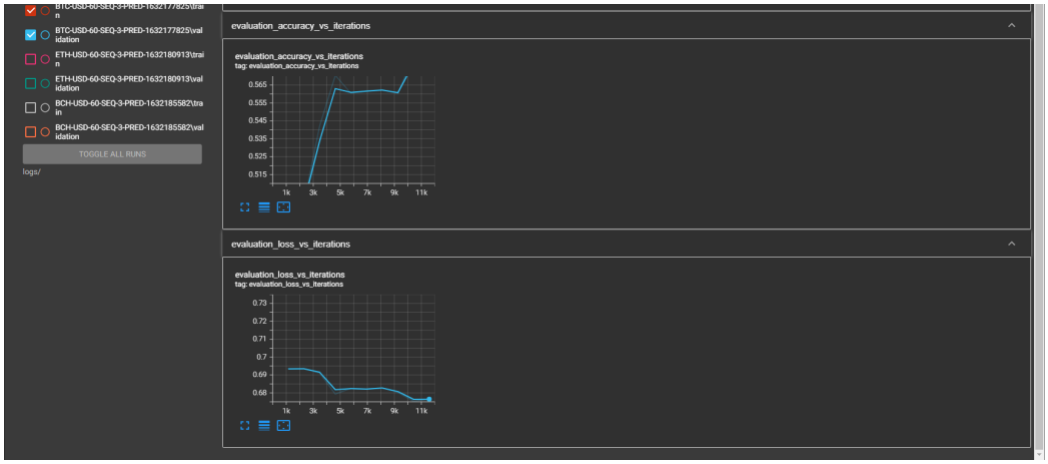
APPENDIX B: Prototypes and Design

Original Design (TensorBoard):

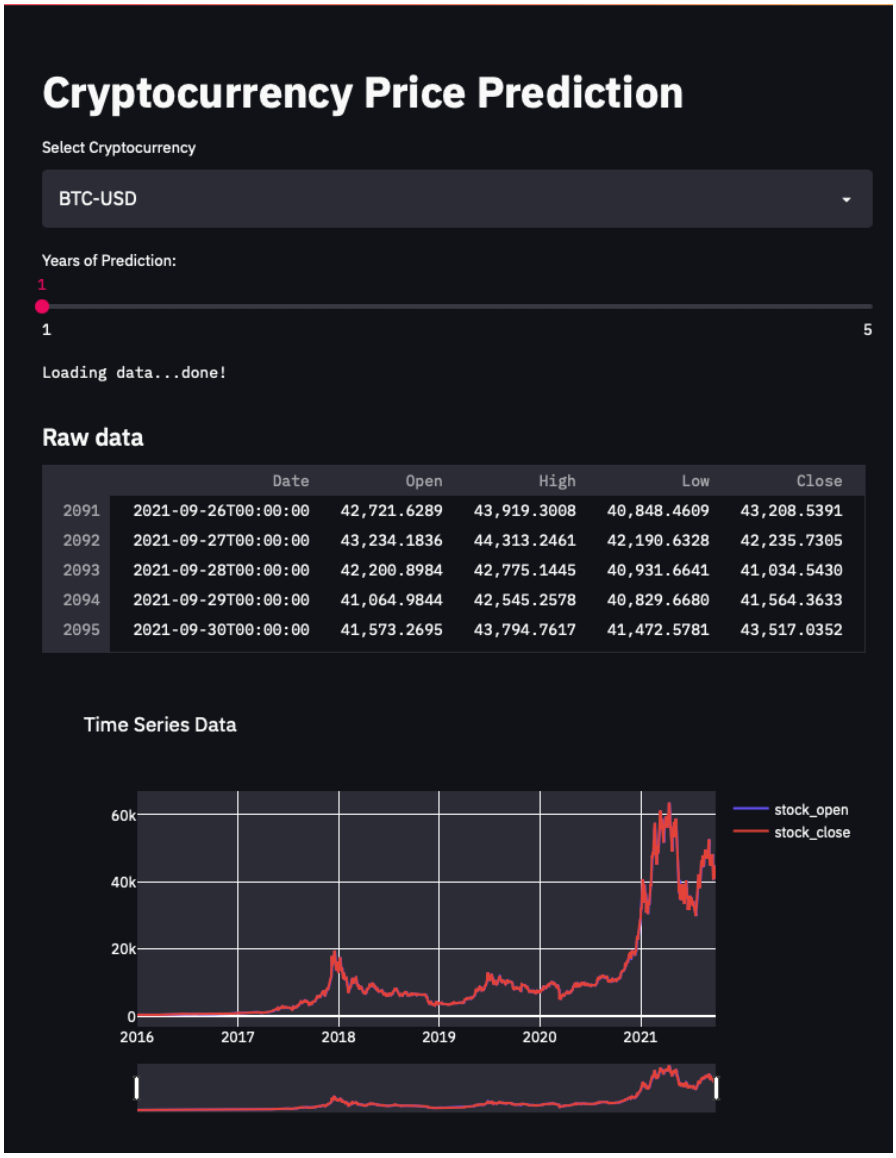




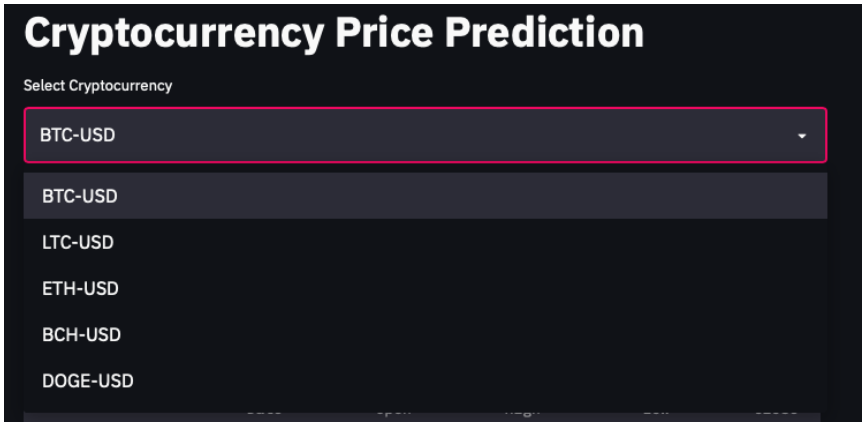




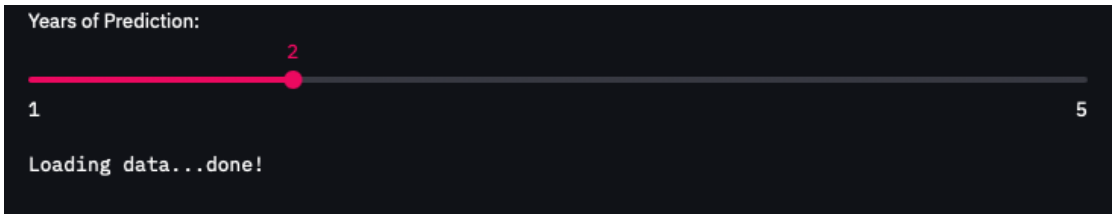
Final Prototype:



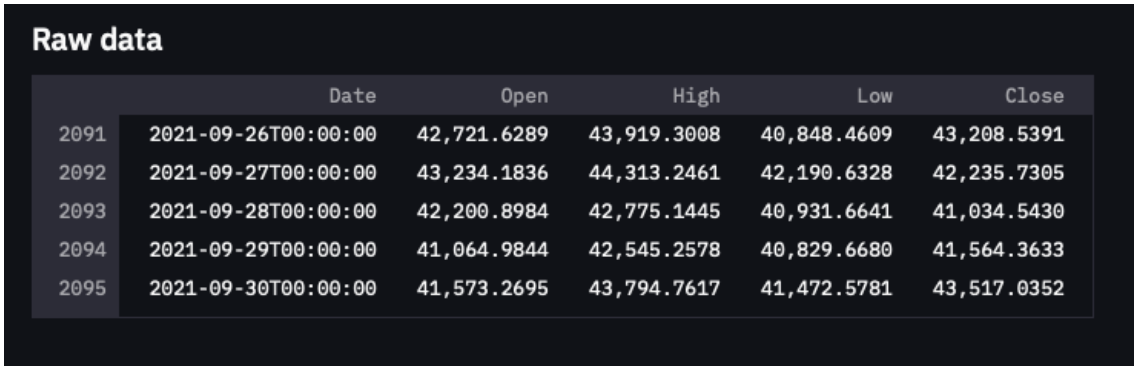
Dropdown Menu:



Prediction Slider and 'Loading Data' Validation:



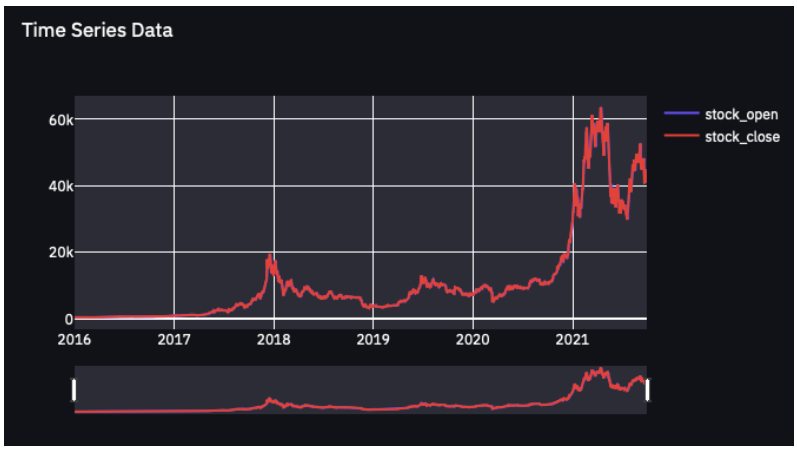
Raw Data Table:



Raw data

	Date	Open	High	Low	Close
2091	2021-09-26T00:00:00	42,721.6289	43,919.3008	40,848.4609	43,208.5391
2092	2021-09-27T00:00:00	43,234.1836	44,313.2461	42,190.6328	42,235.7305
2093	2021-09-28T00:00:00	42,200.8984	42,775.1445	40,931.6641	41,034.5430
2094	2021-09-29T00:00:00	41,064.9844	42,545.2578	40,829.6680	41,564.3633
2095	2021-09-30T00:00:00	41,573.2695	43,794.7617	41,472.5781	43,517.0352

Plotted Training Data with Slider:

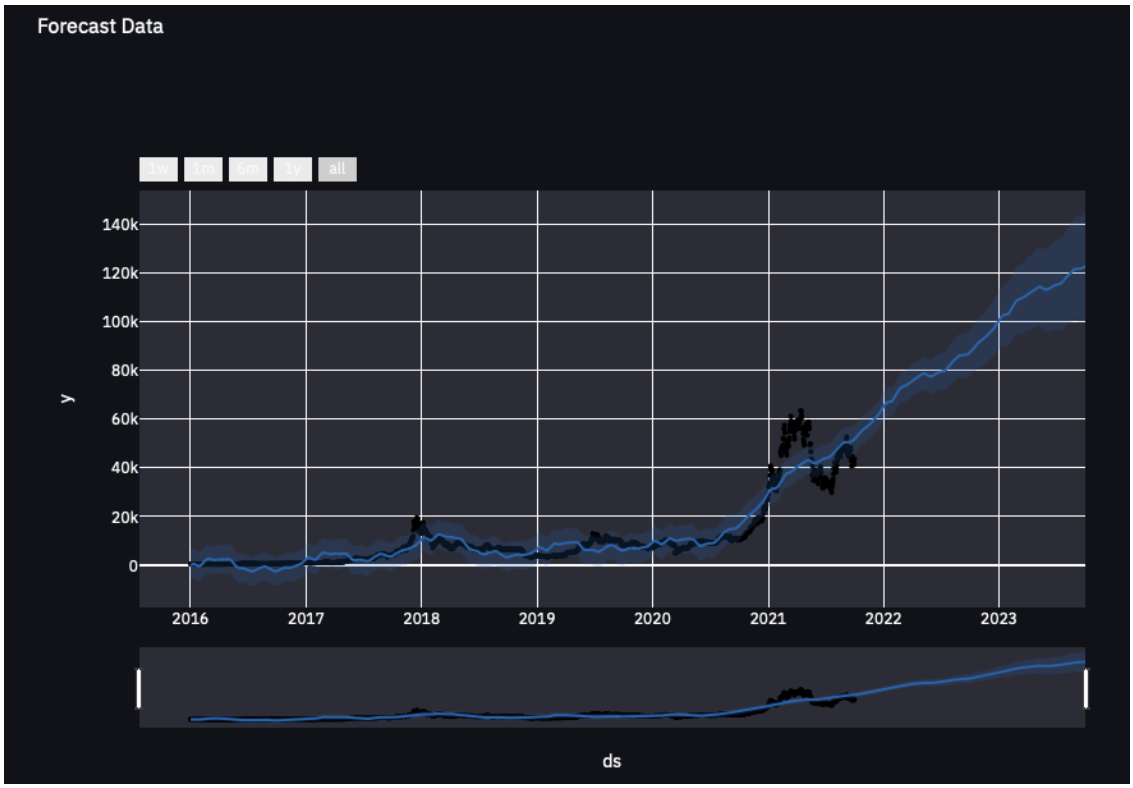


Forecast Data Table:

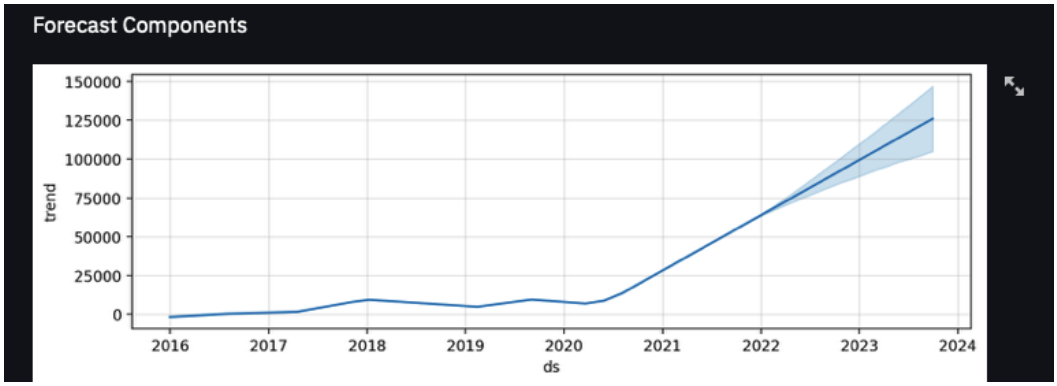
Forecast Data

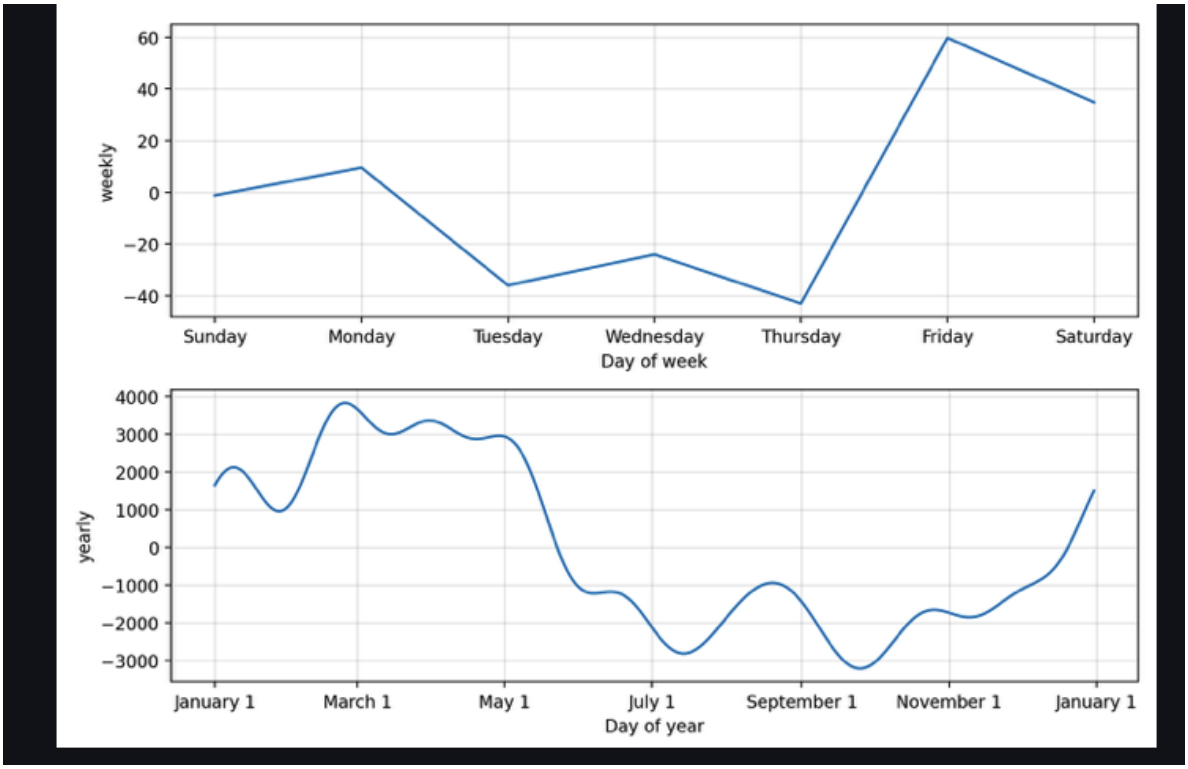
	ds	trend	yhat_lower	yhat_upper	trend_low
2821	2023-09-26T00:00:00	125,522.8900	99,332.8708	144,513.2505	104,654.55
2822	2023-09-27T00:00:00	125,620.4036	99,782.6328	144,563.7049	104,695.28
2823	2023-09-28T00:00:00	125,717.9172	100,172.3515	144,484.9616	104,736.01
2824	2023-09-29T00:00:00	125,815.4307	100,322.6283	145,214.8918	104,776.74
2825	2023-09-30T00:00:00	125,912.9443	100,767.0082	144,172.6647	104,817.47

Plotted Forecast Data with Slider:



Forecast Components:

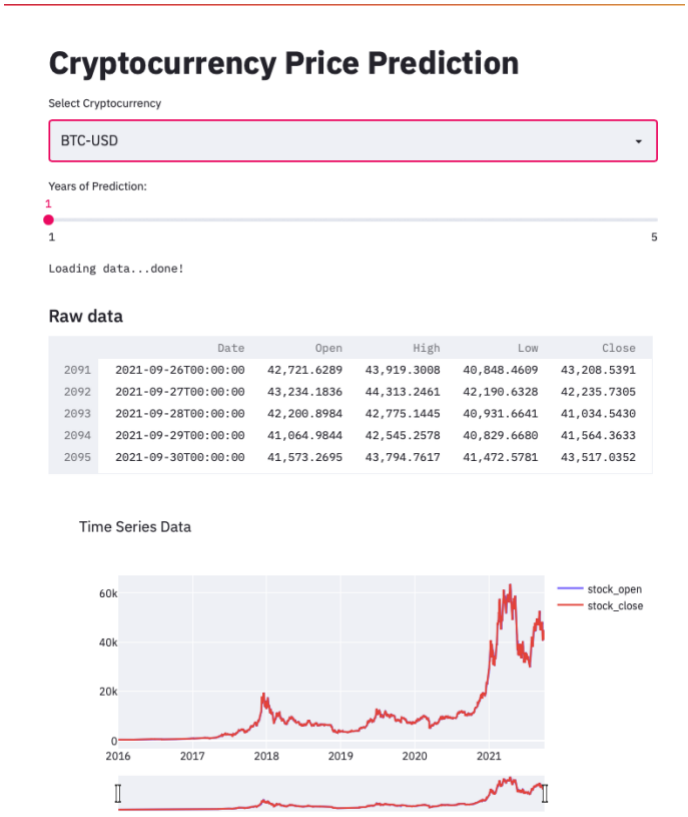



weekly
60
40
20
0
−20
−40
Sunday
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Day of week
yearly
4000
3000
2000
1000
0
−1000
−2000
−3000
January 1
March 1
May 1
July 1
September 1
November 1
January 1
Day of year

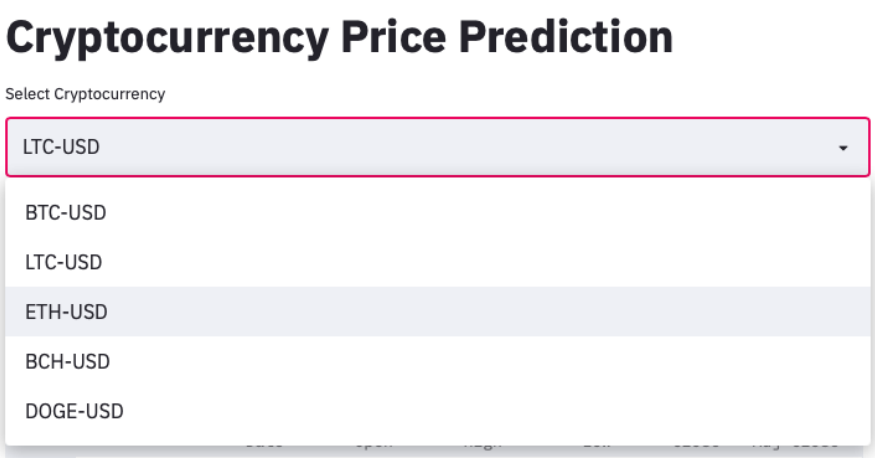
APPENDIX C: Extended User Documentation

Documentation:

Upon first opening the application, the user will find a well-designed web page for predicting cryptocurrency prices:



The user can select which cryptocurrency they would like to view the data for by clicking the drop-down menu. Selecting a different cryptocurrency will refresh the page:



Once the page has refreshed and the selected data has loaded, the page will prompt the user that the data has been successfully loaded:

```
Loading data...done!
```

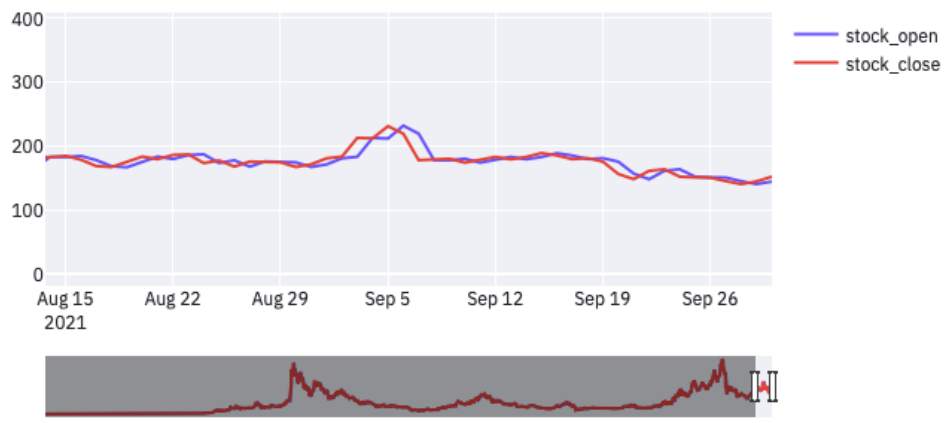
Raw data

	Date	Open	High	Low	Close	Adj Close
2091	2021-09-26T00:00:00	151.1856	153.4053	142.5475	150.6409	150.6409
2092	2021-09-27T00:00:00	150.8633	154.8250	145.2765	145.3496	145.3496
2093	2021-09-28T00:00:00	145.2203	148.4471	140.3709	140.5971	140.5971
2094	2021-09-29T00:00:00	140.6720	149.6635	140.1452	144.9186	144.9186
2095	2021-09-30T00:00:00	144.6769	152.5168	144.2387	152.3158	152.3158

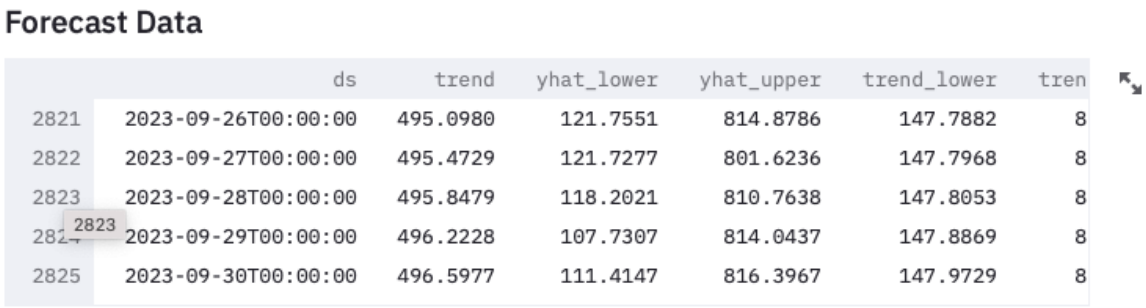
This raw data is also displayed on the 'Time Series Data' graph, and the slider underneath the graph:

Time Series Data

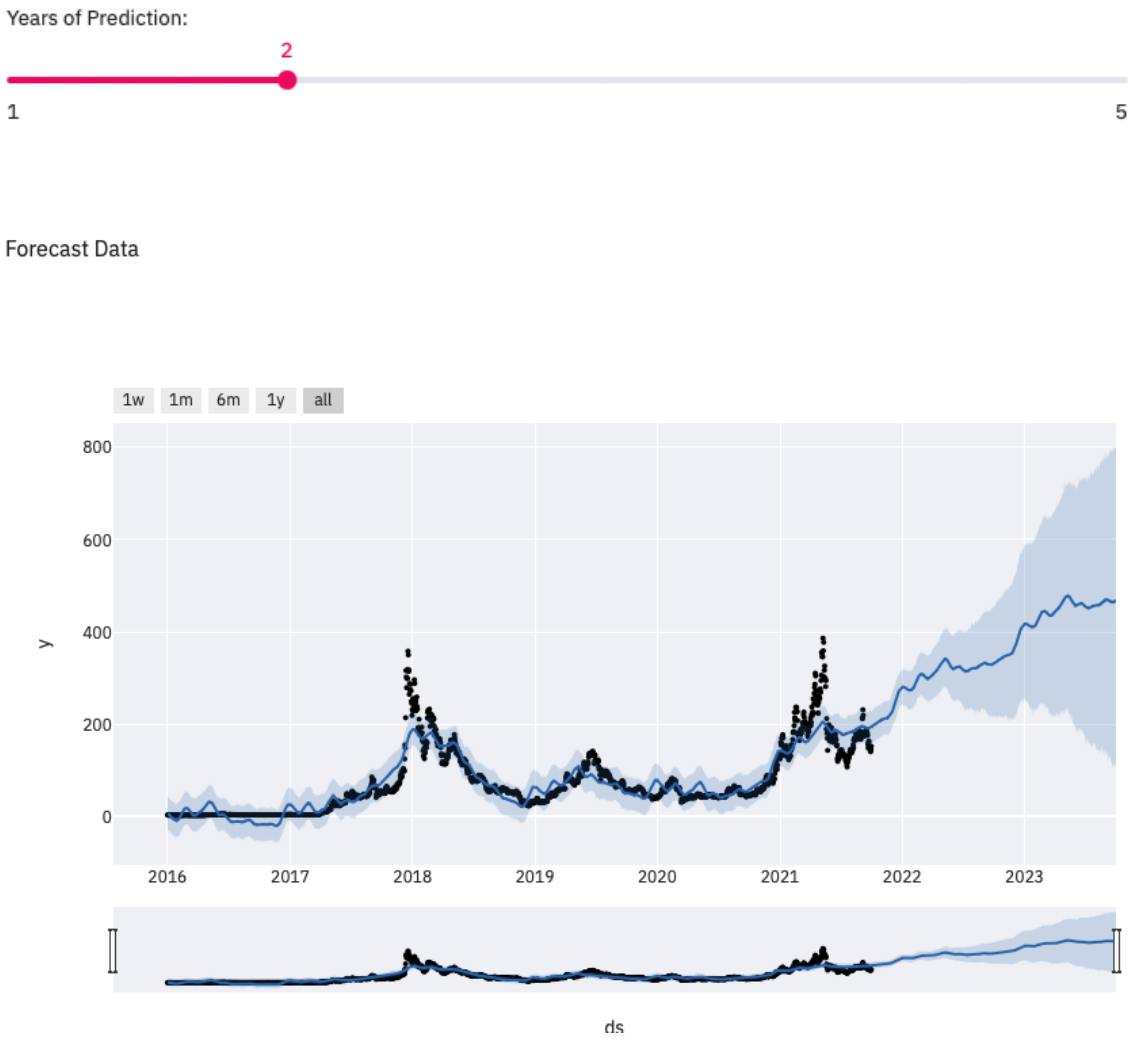
Time Series Data



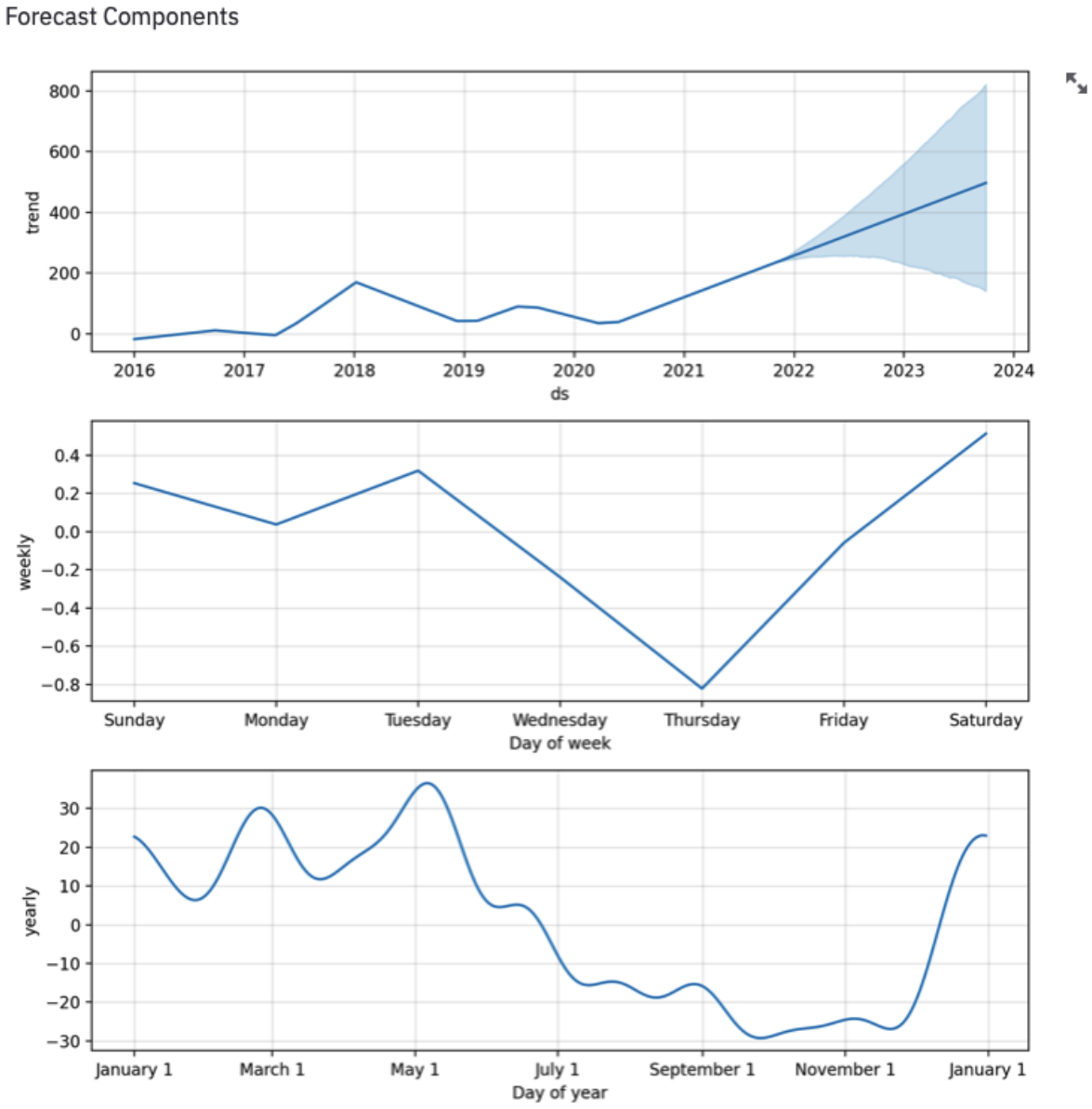
The forecast data for the selected cryptocurrency can be viewed underneath the ‘Forecast Data’ heading:



The user can select how far in the future they want to see forecast data for by using the slider labeled ‘years of prediction’:



The user can also view the forecast components for the selected cryptocurrency:



APPENDIX D: Project Proposal

A formal proposal for a MSc project that will be submitted in partial fulfillment of a University of Greenwich Master's Degree

Enter your MSc Project Title here – Forecasting Cryptocurrency Prices using Convergence Analysis of Machine Learning Algorithms

Name: Trevor Kiggundu

Student ID: 001001720

Course of Study: MSc Computing and Information Systems

Date Proposal Submitted: 07/06/2021

Submission Date: 16/09/2021

Supervisor: Solomon Ebenuwa

Topic Area: FinTech, Information Systems, Machine Learning, Algorithms, Statistics, Data Structures

Keywords associated with the project: Regression, Algorithms, Statistics, FinTech

MSc Modules studied that contribute towards this project: Project Management, Systems Modelling, Systems Design and Development, Software Quality Management

1. Overview

The project that is being proposed is the creation of a predictive modelling application that aims to accurately predict financial data, more specifically the ever-changing entity that is cryptocurrency prices. The research that will be undertaken will involve researching the fluctuation in cryptocurrency prices over the last few years, with a special concentration on the last year or so, as the COVID-19 pandemic has brought along a lot of changes that both positively and negatively impacted the merging world of technology and finance.

Mathematical Analysis is very important in this regard, as it helps to compute and provide statistics that can lead to much more accurate predictions about the future of influential entities such as the cryptocurrency market. The intended result is to create an algorithm that accurately predicts the price of various leading types of cryptocurrency, with the price of the selected type of cryptocurrency being the dependent variable, and an undetermined value being the independent variable.

The intended result is to create an algorithm that accurately predicts the price of various leading types of cryptocurrency, with the algorithm accurately considering past historical values of both variables. To do this, a set of test data will be used encapsulates the entire cryptocurrency market, and not just 'Bitcoin' for example. This will help the consumer better understand the relationship between cryptocurrency price and the chosen variable and type of currency. The findings will then be displayed using a Java application programmed using the Netbeans IDE, as this will allow for the creation of an application that boasts visually pleasing UX and UI designs, so that the user can see the price prediction on a graph as well.

The proposal discussed above encompasses many elements that have been studied this term:

- Software Tools and Techniques: Application development and Java programming
- Project Management and Software Quality Management: Project Planning, Execution, Risk Management and Quality Assurance
- Organizational Awareness: Understanding the way certain companies operate and run their business
- Systems Modelling/Design and Development: UX/UI design

This project will meet the QCF guidelines because of its originality in approach; it was not a staff project given to me, so a lot of effort went into finding something that would keep my interest. It is also a project based on an extremely relevant topic, as cryptocurrencies are increasingly becoming a trusted way to spend and invest money. The fluctuation in crypto price is also ever-changing problem, so there is no one size fits all solution, especially given the growing number of currency types. Finally, I believe that this project will show that I am ready and able to work in the field, as it not only encompasses elements of technology, but also those of economics and finance, two of the main aspects of the rapidly growing field of FinTech.

2. Objectives

There are two main objectives that are essential to the project, as well a few secondary ones that are important but not imperative to the process:

2.1 Primary: Undertake and create multiple predictive models using the selected algorithm to accurately predict the prices of leading types of cryptocurrency.

- create an algorithm that accurately predicts the price of various leading types of cryptocurrency.
  price of the selected type of cryptocurrency being the dependent variable, and an undetermined value being the independent variable.

2.2 Primary: Deduce and propose the variables and factors that affect the price fluctuation of the selected cryptocurrency.

Create an application that accurately depicts the cryptocurrency price prediction.

- Design the UX/UI so that it is visually pleasing and easy enough for the user to understand.
- Build the application.

2.? Project Plan

- To analyse the current state of the cryptocurrency market and compare it to previous years as well.
- To investigate and conclude the best resources to find test data for the given currency.
- To research the already available products on the market and compare their effectiveness to tweak the final product.
- To conduct an extensive literature review on topics such as the UX/UI designs, cryptocurrency, regression analysis, and algorithms.
- Search for journals and similar projects that might give a better insight on the project at hand.
- Initial/Interim report

- Final Product

3. Legal, Social and Ethical Issues

Can it be that some people might find the product offensive?

- No.

Will you be using copyright material? How will you get permission to use it?

- No. The data trends for Cryptocurrency prices over the last few years are easily accessible online, and the name of the regression tool created will be quite basic and ambiguous.

Does the product adhere to accepted standards and is it secure enough to hold data? In case you are dealing with confidential data have you seek advice from the Ethics Committee?

- Yes. None of the data is confidential.

Is there a chance that there might be health and safety issues with the product or with the project in general?

- No. The product does not require any physical engagement between the developer and the consumer.

Is there a chance that there might be accessibility issues with the product?

- Yes. This is always a risk when developing products that must be programmed. The current virtual environment does not help the situation either, but a large amount of quality assurance should be done to make sure there are no tedious errors especially, such as syntax errors.

Will you need permits to record video/audio/photography?

- No

How does the Disability discrimination act affect you?

- I will strive to make sure that the final product is as accessible as possible to all users. Even trivial things like the colour scheme of the application should be paid attention to.

Are there any social issues that your project may impact on?

- No

Are there any issues with the Data Protection Act (DPA), Data Protection Directive and the Human Rights Act?

- No

4. Resources

What resources are needed to do these activities?

- Access to VMware Horizon Desktop
- Access to libraries with journals for literature review and annotated bibliographies.
- Access to the relevant Java libraries to create the best version of the application possible.
- Access to reliable internet connection.

  All these risks can become issues especially given the nature of interaction caused by the COVID-19 pandemic.

5. Critical success factors

What are the critical activities, people or resources that can make or break your project?

- Loss of access to VMware desktop: This can be dangerous as communication with the supervisor would be lost, but it can be handled.
- Being unable to access the correct libraries would not be catastrophic, as the application could still be made without it.
- Lack of proper internet connection would also be catastrophic.
-

APPENDIX E: Borrowed Code

#CODE REFERENCED FROM https://github.com/python-engineer/python-fun

Imports from pip install

```
import streamlit as st
from datetime import date
import yfinance as yf
from fbprophet import Prophet #main class
from fbprophet.plot import plot_plotly
from plotly import graph_objs as go #interactive graphs via plotly, graph objects


START = "2016-01-01" #start date of crypotcurrency data
TODAY = date.today().strftime("%Y-%m-%d") #todays date converted to a string

st.title("Cryptocurrency Price Prediction") #title of the web page

stocks = ("BTC-USD", "LTC-USD", "ETH-USD", "BCH-USD", "DOGE-USD" ) #stock
names as shown on NASDAQ/YF
selected_stock = st.selectbox("Select Cryptocurrency", stocks) #creates a dropdown box to
select the stock


n_years = st.slider("Years of Prediction:", 1, 5) #creates a slider to show # of years for
prediction
period = n_years * 365 #period has to be in days, so multiply by 365
```

Loading stock data from Yahoo Finance

```
@st.cache #caches data so it does not have to be loaded again

def load_data(ticker):
        data = yf.download(ticker, START, TODAY)
        data.reset_index(inplace=True) #returns it in a pandas database
        return data

data_load_state = st.text("Load data...")
data = load_data(selected_stock)
data_load_state.text("Loading data...done!")

st.subheader('Raw data')
st.write(data.tail()) #takes raw data and displays it in table form (tail: last 10 values)
```

```
#Plotly graph object creation

def plot_raw_data():
        fig = go.Figure()

        #Plots only open and close price for selected cryptocurrency
        fig.add_trace(go.Scatter(x=data['Date'], y=data['Open'], name='stock_open'))
        fig.add_trace(go.Scatter(x=data['Date'], y=data['Close'], name='stock_close'))

        #creates a slider to show open and close prices in greater detail
        fig.layout.update(title_text="Time Series Data", xaxis_rangeslider_visible=True)

        #takes raw data and plots it using Plotly Graph
        st.plotly_chart(fig)

plot_raw_data()
```

Forecasting using Facebook Prophet

```
df_train = data[['Date', 'Close']] #Forecasting only using Date and Close Price
df_train = df_train.rename(columns={"Date": "ds", "Close": "y" }) #changes name of axis
for FBProphet to work

#df_train = data[['Date', 'Open']] #Forecasting only using Date and OPEN Price
#df_train = df_train.rename(columns={"Date": "ds", "Close": "y" }) #changes name of axis
for FBProphet to work
```

Creating FB Prophet Model

```
m = Prophet()
m.fit(df_train) #fits training data and starts training the model
future = m.make_future_dataframe(periods=period) #df that goes into the future for
prediction
forecast = m.predict(future) #returns prediction in 'forecast'

st.subheader('Forecast Data')
st.write(forecast.tail()) #takes forecast data and displays it in table form (tail: last 10 values)

#uses plotly chart to plot forecast data

st.write('Forecast Data')
fig1 = plot_plotly(m, forecast)
st.plotly_chart(fig1)

#uses regular graph to plot components (daily, weekly, monthly) data

st.write('Forecast Components')
```

```
fig2 = m.plot_components(forecast)
st.write(fig2)
```