Prediction of Forked Cryptocurrency Prices with Machine Learning Models

MSc Research Project
MSc in Fintech

Lokesh Raj Duraisamy Nagarajan
Student ID: x18142231

School of Computing
National College of Ireland

Supervisor: Theo Mendonca

Prediction of forked cryptocurrency prices with machine learning models

Lokesh Raj Duraisamy Nagarajan
X18142231

Abstract

Cryptocurrency has attained immense popularity in recent years and is considered a significant asset among investors and media. Most of the research in the past emphasises on Bitcoin; whereas, a division of the crypto coins popularly known as the Forked coins have not yet been deeply investigated. This paper primarily aims to predict the closing prices of two fork coins, namely Bitcoin Cash (BCH) and Ethereum Classic (ETC), as part of Bitcoin and Ethereum cryptocurrencies group. The prediction is a result of an hourly analysis of Open, High, Close And Volume(OHLCV) data provided by the Kaiko website and comparing modern algorithms such as 1) LSTM Deep Learning Algorithm, 2)Xtreme Gradient Boosting(XGBoost),3)Facebook's forecasting model PROPHET, and traditional time series 4)ARIMA model. Furthermore, the results were evaluated with RMSE performance metrics. Experimental results of this research indicate that the LSTM deep learning model outperforms other models for both the forked coins followed by ARIMA. The study concludes that the modern algorithm performs efficiently because of the error rate through the traditional ARIMA model.

Keywords: *Cryptocurrency, Forked coins, Close price, Machine learning, Forecasting* Bitcoin *Cash (BCH) and Ethereum Classic (ETC).*

1 Introduction

Bitcoin, as a cryptocurrency, has received much attention because of its first implementations (Pierro, 2017). We reflect the centre of modern cryptocurrency development, together with Ethereum, blockchain implementation with a focus on smart contracts. Decentralised money was excessively used as a theoretical concept in the past; however, Nakamoto's (2008) seminal piece of work, led to the sustained development and viability of the Bitcoin and the blockchain technology in the last decade, With the increasing proven record of blockchain implementations, adopters must concentrate on developing distribution applications, instead of developing new ones from scratch. The main objective of the blockchain technology and development of cryptocurrencies was to create a decentralised environment in which no third-party controls the transactions and data (Cannon, 2018). It is now a mainstream technology because it solves many problems that people were unable to do previously, thereby generating a commercial value of around $176 billion by 2025 and $3.1 billion by 2030.

The main reason for this expansion is that the blockchain for financial transactions is widely being accepted across different countries for cross border payments. The concepts of blockchain and cryptocurrencies have become financially significant, and are one of the biggest discoveries in the field of Information Technology (Cannon, 2018).

1.1 Cryptocurrencies

There are more than 1,000 cryptocurrencies that are available in the market. Different cryptocurrency coins are using various incentive structures to promote new computer nodes around the world. Cryptocurrencies provide autonomy to the users to validate the transactions and access them at the same time. Most of the cryptocurrencies are primarily created to replace the usage of fiat currency, which is achieved by a fixed market cap for the circulation of the coins; nevertheless, there are some cryptocurrencies like ripples which are limitless. Chohan (2017), in his research, examines that cryptocurrency has a high potential to become the main payment method in the future. Bitcoin was the first efficient and competent blockchain and cryptocurrency implementation. As per the survey[1] conducted by PwC, 6% of the respondents in the survey conveyed that they are familiar with cryptocurrencies; moreover, only 3% of the respondents used cryptocurrencies within that year. Various strategies to check payments have led to the invention of cryptocurrencies such as Bitcoin cash, which promotes more substantial scalability and faster transaction rate.

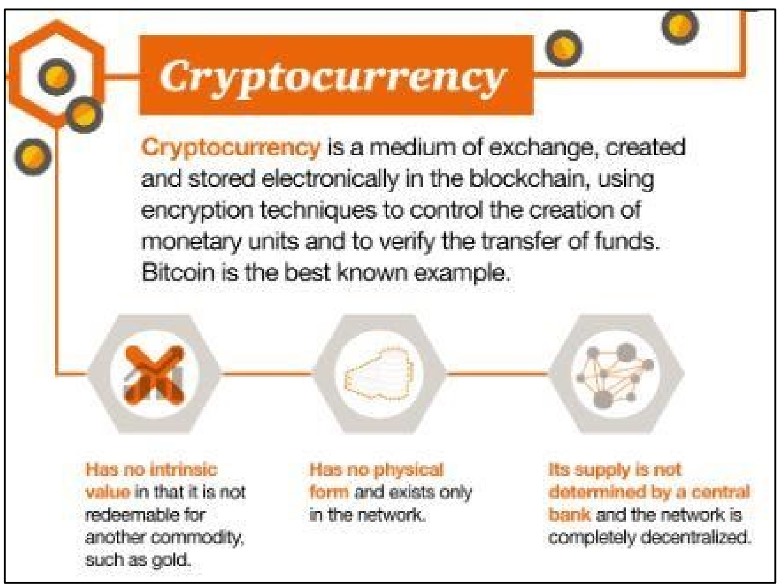

Figure 1 Cryptocurrency (PwC,2018)

[1] Pwc(2015):https://www.pwc.com/us/en/financial-services/publications/assets/pwc-cryptocurrency-evolution.pdf

1.2 Blockchain

Blockchain is defined as a decentralised public ledger which maintains a track of all the transactions which take place in the network. Nakamoto further describes accomplishing peer to peer transactions without the involvement of the third party. All these transactions are performed in a systematic process where user A intends to send data to user B, this transaction between A and B is denoted by a block which is broadcasted to every node in the network, the users of the blockchain then validate the transactions, soon after the verification, it is added to the ledger, and the new block is added to the existing blockchain. The process of validation involves miners; miners are required to solve the puzzle to authenticate the transactions; this puzzle-solving procedure is known as "Mining." A miner who finishes the puzzle at a faster rate is awarded the bitcoin. At present, the miners are awarded 12.5 BTC for every single completion of the puzzle. The new block cannot be added to the blockchain unless 51% of users approve the solution for that transaction. These blocks, once accepted, cannot be modified. As mentioned in the report by Accenture[2], blockchain has a significant potential to unlock various business values in different industries, such as aviation, healthcare, entertainment, hospitality. 53% of the respondents in one of the surveys conducted by Deloitte[3] (2019), stated that blockchain technology h become the first adoption in the organisation, and is increasing at a rate of 10 per cent as compared to 2018.

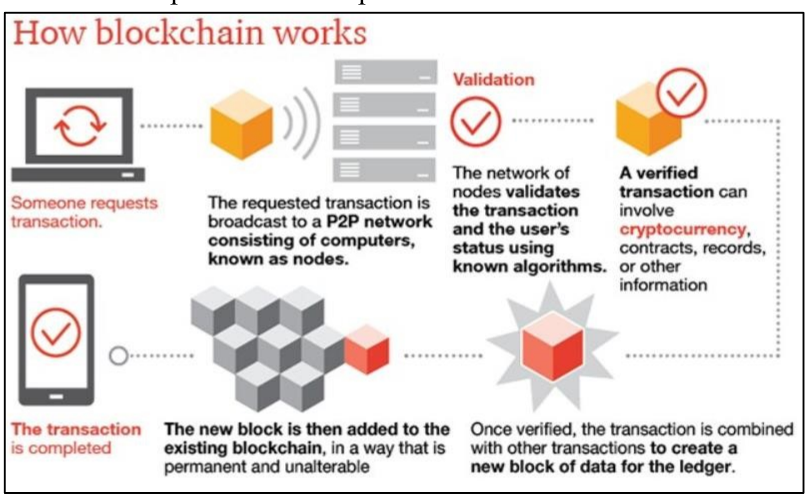

Figure 2 Blockchain Functionality (PwC,2018)[4]

1.3 Forked Coins

As per Shahsavari *et al.*,(2019), inconsistencies in the form of forks can be found in the blockchain consensus processes. *Forking* refers to a type of software upgrade that can process or defer backward compatible techniques. Forks create an alternative version of the blockchain, which means that two blockchains simultaneously run on the same network, Forks can also

[2]Accenture(2019)https://www.accenture.com/_acnmedia/PDF-88/Accenture-20180514-Blockchain-Interoperability-POV.pdf#zoom=50

[3]Deloitte(2019)https://www2.deloitte.com/content/dam/Deloitte/se/Documents/risk/DI_2019-global-blockchain-survey.pdf

[4] Pwc(2018) https://www.pwc.com/us/en/industries/financial-services/fintech/bitcoin-blockchain-cryptocurrency.html

occur in multiple cryptocurrencies, such as Ethereum, Bitcoin and Z Cash. According to Kiffer *et al*., (2017), a large-scale fork occurs in Ethereum, where a new blockchain platform allows both money transfers and smart contracts. Islam *et al*., (2019) elaborate that forks can split the blockchain either on a permanent or temporary basis. Forks fall into two categories (1)Soft fork and (2)Hard fork, which is initiated according to consensus rules to upgrade a cryptocurrency. A soft fork is a backward-compatible software upgrade, wherein the old nodes can facilitate the transactions and add new block into the blockchain and must follow the new consensus rules. For instance, block size 'A' is lowered from 4MB to 3MB; the older nodes can still involve the transactions and push new blocks. However, the old nodes push a block that is more than 3MB, which then violates the rules and is considered invalid in the cryptocurrency protocol. On the other hand, a hard fork is defined as an incompatible previous version of the blockchain, wherein the older nodes that are not upgraded will not be able to push the block in the blockchain.

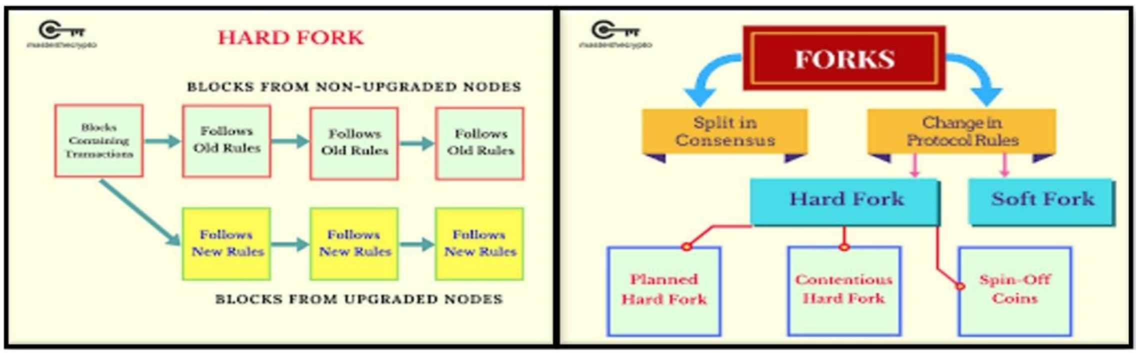

Figure 3 Formation of Fork and Hardfork[5]

1.4 Bitcoin Cash

According to Kwon *et al*., (2019), Bitcoin has become the most popular peer-to-peer network cryptocurrency. In August 2017, the original Bitcoin (BTC) and Bitcoin Cash (BCH) was subdivided. Since then, BTC and BCH mining are mostly preferred because of the compatible Proof of Work algorithms in both the coins. Bitcoin Cash is distributed in P2P digital cash. It was developed and created on 1st August 2017 and is based on the SHA-256 Proof of Work (PoW) algorithm of Bitcoin, but with certain modifications to the code. Bitcoin Cash is a 'hard fork' of the Bitcoin blockchain in the crypto-community. It is the result of two very different visions about Bitcoin and blockchain futures, which separated the Bitcoin blockchains into two routes. Some Bitcoin developers wished to increase their block size limits from 1 MB to 8 MB to reduce transaction fees and improve confirmation times. Bitcoin Cash, like Bitcoin, uses the process Proof of Work (PoW), which means that it can be mined. Nonetheless, anyone who owned the Bitcoin during the development period always kept the same sum of Bitcoin cash, which is a direct result of the hard fork.

[5] https://masterthecrypto.com/guide-to-forks-hard-fork-soft-fork/

Even though Bitcoin has the same attributes as Bitcoin Cash, one disadvantage often discussed is with regards to scalability. Bitcoin Cash is a solution to resolve the problems of high transaction costs and long duration confirmation times, wherein all other features are similar when sliding around technical upgradations; thus, the Bitcoin cash is preferred over the Bitcoin (P and Schmidsfelden, 2018). Roger ver, [6] who is an early investor of bitcoin, claims that the price value bitcoin cash would quickly rise to 90% in the future. From Figure 4, Bitcoin cash is considered to be one of the forked coins to be in the top 5 of the popular cryptocurrency list. According to one of the reports[7], Bitcoin Cash usage has exceeded Bitcoin in Venezuela, more than 360 venues are now accepting bitcoin cash in their businesses, and initiatives are being taken to diversify bitcoin cash.

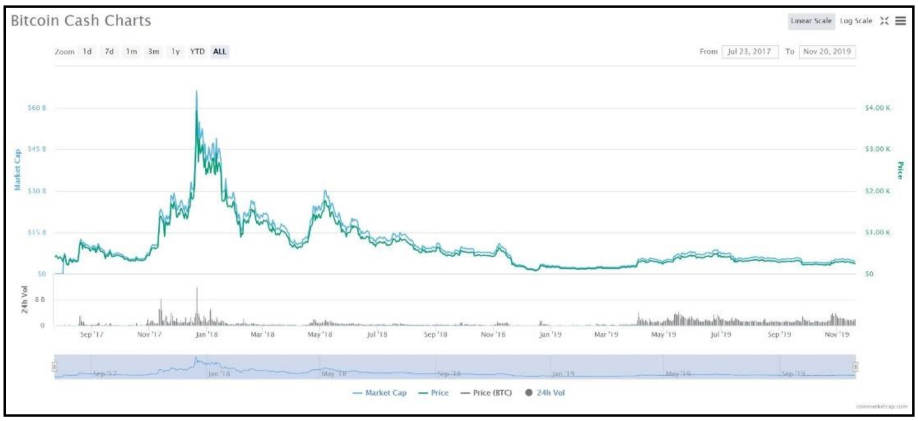

Figure 4 Market Cap of Bitcoin Cash[8]

1.5 Ethereum Classic

Ethereum is considered as one of the most distinguished cryptocurrencies across the globe. In the year 2016, the Decentralized Autonomous Organization (DAO) took the initiative to develop Ethereum with an investment of $168 million. The principal motive was to invest in profitable ventures with the benefit of smart contracts. An intriguing fact featured in the same year, when $50 million from the DAO was found transferred to other accounts without the consent of the owners (Andreas et al., 2018). This incident increased the susceptibility, and eventually manoeuvred committee members of DAO and Ethereum to discuss and avert such incidents in the future[9]. A referendum voted in July 2016 later, implemented the use of hard fork as an independent form of the Ethereum code. The researcher opines that the officials must have conceptualised the idea to develop a hard fork as a separate cryptocurrency soon after the

[6] Fadilpašić, S. (2019). Roger Ver Says Bitcoin Cash may be Worth 99,900% More 'Some Day'. [online] Cryptonews.com. Available at: https://cryptonews.com/news/roger-ver-says-bitcoin-cash-may-be-worth-99-900-more-some-da-5004.htm [Accessed 21 Nov. 2019].
[7] Hamacher(2019),https://decrypt.co/11370/bitcoin-cash-just-trumped-bitcoin-in-venezuela
[8] Coinmarketcap(2019).https://coinmarketcap.com
9 Michael del Castillo (2016). The DAO Attacked: Code Issue Leads to $60 Million Ether Theft - CoinDesk. online] CoinDesk. Available at: https://www.coindesk.com/dao attacked-code-issue-leads-60-million-ether-thef

hackers' incident. After the attack[10], It conveys that several parties had not acquainted themselves with discussing the opportunity of the Ethereum channel, which was forbidden.
Furthermore, the Ethereum Classic, as observed, grew only after the denials. It is the first kind of project that was managed entirely by a community, where they made an extraordinary effort to design the logo, website content, layout, and many other aspects. The Ethereum Classic thus came into existence after the website launch as the token summed up to the exchange of Poloniex.
It is worthy to note that Ethereum Classic came into existence soon after the members of the Ethereum community invalidated the hard fork based on immutability grounds. According to the blockchain principle, it is not possible to change the hard fork and adopt the usage of the Ethereum, which is unforked. Besides, the first Ethereum Classic was involved in the forked Ethereum chain with a block number of 1,920,000 (Andreas *et al*., 2018), which was created by Classic miners in July 2016. The Ethereum classic went through the hard fork technical to internally alter the pricing for different opcodes, especially on the Ethereum Virtual Machine. It is observed that individuals who follow Ethereum Classic assess the immutability of the blockchain against the Ethereum coins.

The background of the Ethereum Classic cannot be defeated. The details of the Ethereum Classic are gathered from various sources, which is why it seems complicated to understand for amateurs who are new in this space; this is mainly because the DAO hard fork did not respond to reverting the transactions received from the hackers. Moreover, the accounts vanished with an estimated 100 wallets balance and yielded other wallet balances. After facing a range of uncertainties, the Ethereum Classic made a vital rule that it is used only as an individual bank, particularly with the hard fork at the beginning and the entire community.
The Ethereum Classic enables users to quickly produce any number of applications on its blockchain, which the bitcoin can never do. It also facilitates blockchain creation from scratch for those who desire to create a blockchain-based application. The creators can thereby create an imaginary application by leveraging the Ethereum infrastructure. The two substantial challenges of the Ethereum Classic coins were to 1) accept the hacking incident and 2) to admit that nothing could be done because the blockchain behaves like an immutable ledger. Scholarly articles also highlight that there is a need to implement the Ethereum Classic as a tool that can be used to cease hackers and discontinue their transactions and other processes that can be detrimental (Kiffer *et al*., 2019). At present, the Ethereum Classic of the Ethereum protocol is Ether Zero.

The aim of Ether zero is to facilitate thousands of transactions per second as opposed to Ethereum's current rate that provides only 15 transactions per second at zero fees and remains

10 Kuzmina, C. (2019). The Painful History of Ethereum Classic - Bitstarz News. [online] BitStarz News. Available at: https://news.bitstarz.com/the-painful-history-of-ethereum-classic.

an ambitious goal for the company. The "Ethereum Foundation," a Swiss non-profit organisation encourages and supports the development and development of Ethereum. More recently, According to CoinDesk[11] Ethereum classic accomplished system-wide Software hard fork upgrade named “Atlantis,” To enhance the security, improve the stability, and performance of pre-compiled contracts, zk-SNARKs (Zero-Knowledge Succinct Non-Interactive Argument of Knowledge) this Upgradation was initiated by the organisation Ethereum Classic Labs where 10 Ethereum improvement proposals (EIPs) where included added to Atlantis. Ethereum classic[12] is gaining traction over the upcoming software upgrade “Agharta” hard fork, which makes it more compatible and seamless with Ethereum.

1.6 Project Specifications

Most of the research in the past is mainly focused on the price prediction of Bitcoin and Ethereum, where scholars have majorly used the time series forecasting methodologies in their studies. The researcher, through this study, attempts to explore and analyse the forked coins using machine learning models by comparing them with a deep learning algorithm.

Research Question: Can modern machine learning models outperform the benchmark set by ARIMA Model in forecasting forked cryptocurrency prices?

Aim and objectives :

This research aims to predict the hourly closing prices of Bitcoin Cash and Ethereum Classic based on hourly OHLCV.

Objectives of this study are as follows:

• Implementation of machine learning techniques to predict hourly closing prices through data analysis.
• Comparison and evaluation of the models with performance metrics.

Research Purpose:
This research intends to investigate the prediction of forked cryptocurrencies using machine learning and deep learning algorithms and aims to provide a comprehensive understanding of the forked coins to the existing body of knowledge and investors in the market to track the performance of the cryptocurrencies. The researcher analyses three modern machine learning algorithms by assessing and comparing individual performances in terms of forecasting the cryptocurrencies.

The models used in this study are 1) KERAS with LSTM (Long Short-Term Memory) model 2) PROPHET Algorithm 4) Xtreme Gradient Boosting (XGBoost) and 4) ARIMA. The data

[11] https://www.coindesk.com/ethereum-classic-successfully-forks-improving-interoperability-with-ethereum
[12] https://www.coindesk.com/ethereum-classics-next-crypto-code-upgrade-is-back-on-for-september

will consist of hourly prices of the Bitcoin Cash and Ethereum Classic, which is provided by Kaiko[13] Digital Cryptocurrency, which is renowned for providing institutional cryptocurrency market data.

The next segment of this research is a detailed analysis of the academic literature obtained from several secondary sources, which elaborate on forecasting models and fork coins; followed by the methodology section, starting from business understanding, data understanding, data preparation, including pre-processing of data, data modelling, and evaluation. Then the research presents the implementation procedures, which encompasses a brief of each model fitting. The results attained from unique algorithms have been constructively analysed and interpreted in the evaluation phase of this research.
Finally, the conclusion, wherein the researcher discusses the results of the research in detail; accompanied by the future scope for the study.

2 Related Work

According to Alessandretti *et al*., (2018), 1681 cryptocurrencies, which were retrieved from CoinMarketCap during November 2015 and April 2018, implemented machine learning algorithm and a deep learning algorithm. In their study, two XGBoost techniques were applied to discover the investment returns for all the currencies, and another method was used to analyse the behaviour of the market to predict a single currency; the other method mentioned was LSTM, used to predict the prices of currency based on previous estimates; the results show gradient boosting decision trees performed better for short duration; on the contrary, forecasts for longer period (50days) the LSTM recurrent neural networks achieved better results than the other two methods. Almasri and Arslan, (2018), in a similar manner, predicted closing prices of cryptocurrencies using neural networks in their research; they further investigated the price changes in the regular stock exchange. The neural network model enabled by *encog* framework was implemented on two data sets, with one based on daily closing prices and the second based on hourly closing prices. The results concluded that hourly based predictions were constant and accurate as compared to the former. The hourly forecasts presented 97.3 % accuracy rate as compared to 75% shown by the daily forecasts

Recently, Karakoyun *et al*., (2018) compared the ARIMA time series model with LSTM deep learning algorithm by evaluating the daily bitcoin prices from 2013 to 2017 to predict the bitcoin prices for 30 days, and the results indicate that LSTM performed superior with MAPE 1.40% and ARIMA with 11.86%. Likewise, McNally *et al*., (2018) assessed the accuracy direction suitable for the prediction of bitcoin price in USD using the traditional time series model ARIMA comparing it with two deep learning models RNN and LSTM. The historical open, high, low and close are retrieved from the coindesk bitcoin dataset; besides, the difficulty and hash rate is taken from the Blockchain. The information about difficulty and hash rate are

[13] Kaiko-Digital Asset Provider:https://www.kaiko.com/

taken. The result concluded that LSTM achieved the highest accuracy of 52.78% and RMSE of 6%.
Moreover, the recent development in forecasting algorithms (Yenidogan *et al*., 2018) compared the ARIMA time series model with the newly introduced PROPHET forecasting model developed by Facebook to predict bitcoin price by analysing the data in the minute format available from 2012 to 2018. In the pre-processing stage, they found that the minute wise data had transaction complexity, due to which the dataset was converted to an hourly transaction where data was trimmed before 2016 as the values were found constant. By evaluation methods, the findings state that the PROPHET surpasses the ARIMA model, as the PROPHET predictions were near actual costs with 94.5% and ARIMA model with a precision of 68%. Thus the researcher chose the hourly data for the analysis of forked cryptocurrencies. Chen *et al*., (2019) came up with a heuristic approach of sample dimension engineering for the bitcoin price prediction in various frequencies by implementing statistical methods such as logistic regression, linear discriminant analysis and machine learning algorithms like the Random forest, XGBoost, quadratic discriminant analysis, SVM, and LSTM. The Bitcoin data was classified into two types, where the machine learning models were implemented on high-frequency data of five minutes interval, and the daily prices were evaluated through the statistical methods. The findings show that the statistical models outperformed the machine learning models with an average accuracy of 65% for the daily data, whereas for the machine learning algorithm achieved an efficiency of only 67.2%. The average accuracy of statistical methods was 53%, as they planned to apply these similar interval price predictions to other cryptocurrencies.

Meanwhile, Mittal *et al*.,(2018) investigated the prediction and the validation of 1592 cryptocurrencies for 30 days with the ARIMA model, wherein the dataset had 13 variables ranging from 2013- 2018. This research only considers the highest rate of the currency for analysis, where the stationarity test was implemented before forecasting the prices of cryptocurrencies using the ARIMA model. The outcome of the model depicts 97.8% accuracy in ARIMA when compared to the multivariate regression model, where it predicted accuracy around 96.9%.

Anupriya and Garg (2018) analysed the daily prices of bitcoin from 2015-2018 with the ARIMA model and predicted the closing prices of bitcoin. They projected the forecasted values for 545 days, which was built according to trends and seasonality within the daily price data. The outcome of the mean percentage error is depicted as 6% of the total values, which shows that the ARIMA model shows an accuracy ranging from 60 to 70 %. Bakar and Rosbi (2017) examined the precision of bitcoin exchange rates in a high volatility environment by using the ARIMA forecasting algorithm, analysed the bitcoin exchange rates ranging from 2013 to 2017. They reported a mean absolute percentage error of 5.36% with two durations, Also concluded that the ARIMA method to be the most reliable forecasting model. However, Rebane *et al*., (2018) researched the ARIMA model and seq2seq recurrent deep multi-layer neural network (seq2seq) to predict the bitcoin over 40 days where ARIMA models were implemented only on closing prices of Bitcoin USD. The seq2seq model had three configured datasets where Bitcoin is traded among three fiat currencies; one model with popular altcoins and finally the third model has the data of all two models with social trend, The results indicate that recurrent neural network (RNN) outperformed ARIMA in more extended period; it is also identified that there is substantial volatility of price. Their study concluded that future research could be performed with additional design changes such as by use of LSTM instead of GRU.

In 2019, Li *et al*., (2019) researched on hard fork coins of Zcash termed as “Z classic” to classify the tweets into positive-negative-natural, and evaluated by XGBoost model. In accordance to the Altcoins, Ji *et al*., (2019) studied various regression and classification problems, thereby evaluating multiple deep learning algorithms such as deep neural network(DNN), LSTM, deep residual network and convolutional neural network (CNN) for the prediction of bitcoin prices using historical daily prices data from 2011 to 2018, where seven prediction models were implemented based on several criteria which include varied size of input sequences, blockchain hash rates, sequential partitioning and also the use of trading strategies to evaluate the profit among the chosen models. The evaluation was performed by MAPE (Mean absolute percentage error); the results showed that LSTM outperformed other models in regression problems, and DNN performed slightly better than other models. Interestingly, this study suggests evaluation of the price variations in coins such as Ethereum, Bitcoin Cash, and EOS as future research, which is the essence of this study.

3 Research Methodology

This research follows the Cross-Industry Standard Process of data mining (CRISP-DM), which is a traditional method to plan and execute data mining development. The CRISP-DM model is customisable and adaptable in terms of handling large data mining projects; the amount of data is reliable and rapidly transforms through several stages in the lifecycle and captures agile methodology which is implemented for data mining projects (Wirth and Hipp, 2000). Muhammad *et al*., (2018) used the CRISP-DM model in their study; where they presented *data understanding, data preparation, data modelling, results and evaluation* for the results. Figure 5 below illustrates the business understanding to depict the business requirements for investors and financial industries. The *Data understanding* stage involves the extraction and exploration of data from the related literature. In the *preparation stage*, the dataset selected is processed for the testing and training. Then in the *data modelling stage*, the cleaned data set is now implemented on the machine learning models for this research. Finally, in the evaluation stage, the results obtained from all the models are systematically assessed and evaluated to resolve the problems.

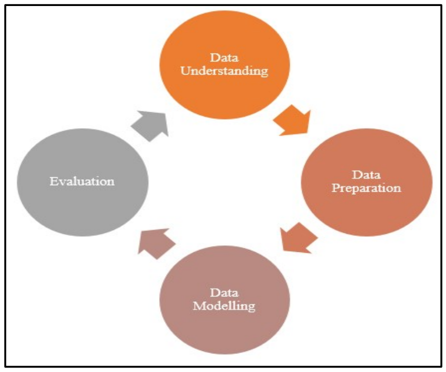

Figure 5 CRISP-DM

3.1 Data Selection:

Two cryptocurrency datasets relating to Bitcoin Cash and Ethereum Classic were obtained from Kaiko-Digital Assets. The hourly OHLCV (Open, High, Close, Volume). Yenidogan *et al*., through his research, affirms that testing the algorithm with the latest data available will ensure greater accuracy to the predictions. Likewise, the researcher in the context chose hourly datasets ranging from 1st January 2019 to 5th November 2019.

3.2 Data Preparation:

In this stage, the dataset is equipped with all the necessary tests, where the data is stationary and doesn't contain any missing values. Both the datasets are presented in the.CSV format, and the hourly data is entered from January 2019 to November 2019 for both the Bitcoin Cash and the Ethereum Classic. Both these dataset contains the features which are defined in the below (Table 1) date and the numeric values of the Open, High, Close, and Volume, where all the prices are traded and exchanged in USD. In the pre-processing stage, the data was derived in an hourly UNIX timestamp (1502380800000) format which was later converted into Year-Month- Day Hour format by **POSIXct** function from **lubridate** library, which is similar to the research methods opted by Yenidogan *et al*., (2018) and Almasri and Arslan (2018).

There are no missing values in both the datasets used in this research. In order to Implement the time series analysis, the closing value of the forked coins is chosen. A prediction data frame with the date and closing prices is created, and then the Augmented Dickey-Fuller test is initiated to evaluate the stationarity of the time series, results will conclude that both the cryptocurrencies are stationary or not. As performed by McNally *et al*., (2018) the researcher also followed the similar splitting strategy of both the datasets assigned 80% for training data and 20% test data for the prediction.

Table 1 Data Features Description

Parameter	Description	Type
Date	Hourly Trading Timestamp	Numeric
Open	Open Price of the coin	Numeric
High	High price of the day	Numeric
Low	Low price of the day	Numeric
Close	Closing price of the coin	Numeric
Volume	Total volume of the Cryptocurrency	Numeric

3.2 Modelling

ARIMA:

ARIMA abbreviates to Auto-Regressive Integrated Moving Average. It is a renowned framework used to forecast the time series applications of financial and information science (Yenidoğan *et al*., 2018). ARIMA consists of two different models, which are classified as seasonal and nonseasonal; the Non-seasonal ARIMA refers to the omission of the seasonal influences such as holidays and other period impacts, wherein only pattern lines are used. On the other hand, Seasonal ARIMA explicitly uses data collection inclusive of holidays and other term effects.

Anupriya and Garg (2018) in their research, highlight the importance of forecasting methods in the stock market and discuss issues related to the use of the forecasting method. Their results illustrate the drawbacks of using the conventional approaches to classify the samples of the best-specified ARIMA system, where prediction aims at this study. The results also establish that the ARIMA models can be useful in forecasting broad market trends. This particular model which is used in specifying the non-linear models which define the trends, errors, and seasonality independently (p, d, q).

The estimates created using alternative parameters have substantial differences. Moreover, Bakar and Sofian Rosbi’s (2017)study suggests that ARIMA is a supportive tool that can analyse significant value changes in the marketplace and that the government and the Central Bank can use ARIMA's forecasting approach to predict domestic price inflation.

PROPHET:

PROPHET is an open-source code that is available in Python, and R used to forecast the time series results. The Facebook Core Data Science Project released the algorithm for forecasting called PROPHET (Taylor and Letham, 2017). Prophet depends on a participation system in which the non-linear patterns are ideal for daily and annual cycles and holidays. The prophet is highly responsible for the missing data, records pattern trends- significant outliers and presents inbuilt parameters for tuning. Prophet uses its settings to optimise valid predictions and metrics like growth, holidays, change-points and seasonality.

They are represented by :

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \qquad (1)$$

Where g(t) is a linear growth curve, s(t) denotes weekly &Yearly seasonality, h(t) effect of holidays and Error terms indicated by εt

In fact, without manual effort, it provides a reasonable estimation of the mixed results. Purely automated forecasting methods are not versatile, as they are brittle and for creating a reasonable hypothesis. High-Quality predictions cannot be made easily and require specialised skills in data science. These are all defined by PROPHET as working inspiration because it aims to project high quality and straightforward forecasts.

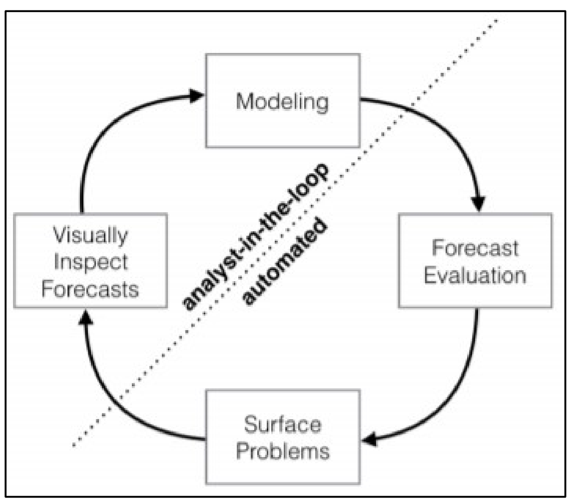

Figure 6 Functionality of Prophet (Taylor and Letham, 2017)

Nonetheless, the concept of the PROPHET works best for regular periodical data at least a year of historical data. PROPHET relies heavily on missing data, pattern shifts, and significant outliers.

LSTM (Long Short-Term Memory):

To resolve the vanishing gradient problem in Recurrent Neural Network (RNN), As stated by Ji et al. (2019), LSTM to solve the long-term dependency problem implemented the idea of gated RNN. This approach is established in various applications and also recognised in multiple cases such as machine language translations, image, and voice recognition. The KERAS model is used to organise the layers and is also useful in creating arbitrary graphs.

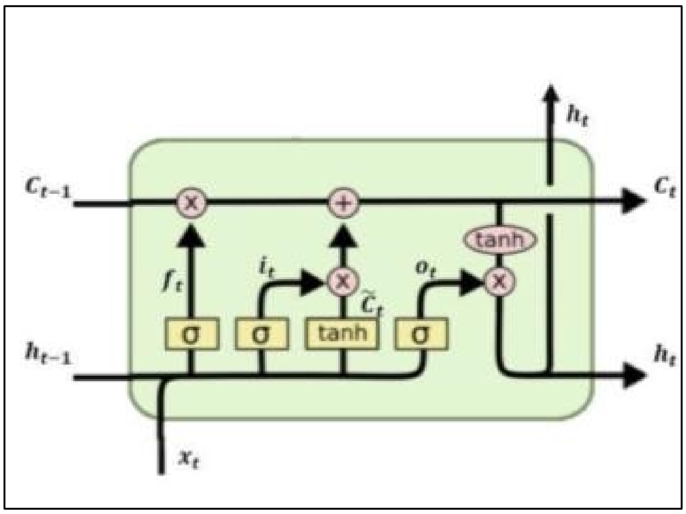

Figure 7 The structure of the Long Short-Term Memory (LSTM)

As per the diagram above, h_t is the hidden state, C_t is the cell state, and f_t is the forget gate

Scenario: LSTM takes the previous cell memory state C_{t-1} and performs multiplication function with forget gate ft

Which gives $C_t = C_{t-1} * f_t$ (2)

The previous state is forgotten if the forget gate value is 0 if it's one, it is passed to the next cell state

From the result of the equation (2), we evaluate the C layer and input state.

$$C_t = C_t + (I_t * C_t) \quad (3)$$

Here C_t is the memory state with time

The final result will be based or is on the Filtered cell state C_t; to get the hidden state h_t, we perform element-wise multiplication function with output gate O, which provides the *hidden state.*

$$h_t = \text{Tanh}\,(C_t\ ) \quad (4)$$

So, the Hidden State h_t and Cell State C_t are given as input, which follows a cycle of iterations.

Xtreme gradient boosting model (XGB Model):

Gradient boosting algorithms solve both regression and classification issues. It generates a prediction model, which is a collection of soft prediction models, including decision-making bodies. Xtreme gradient boosting was established to enhance the performance and efficiency of the Gradient boosting tree. The XGboost framework acts as a parallel processing system that regularises the tree by implementing a fitting residual evaluated from the previous tree, which eventually augments the optimisation process.

Various studies in the past used the gradient enhancement and related technologies, including intense gradient enhancement, to build predictive cryptocurrency values, estimate the number of cyber-criminal bodies within the Bitcoin ecosystem, and to identify schemes of Ponzi in the Ethereum market (Chen *et al*., 2018). Alessandretti *et al*., (2018) state that XGboosting is one of the scalable machine learning models for tree boosting, and has won various Kaggle competitions. Taieb and Borgne, (2013) proposed a Gradient boosting approach for the Kaggle Load forecasting competition and won fifth place worldwide, where the goal is to backcast and forecast the hourly electricity loads in kW for the US within 20 geographical zones.

3.3 Evaluation

Ji *et al*. (2019) stated that most of the prediction models are evaluated in root mean squared error (RMSE) metric, especially for regression problems. It's one of the commonly used parameters for price prediction and forecasting. RMSE is calculated in original units and ranges between 0 and 1. Due to the squaring of error values from the forecast, the results attained are positive and for the model effectiveness, the significant error is exhibited when compared to small errors.

$$RMSE = \sqrt{N^{-1}\sum_{i=1}^{N}(x_i - \hat{x}_i)^2}$$

Where N defines the number of observations used for testing.
x is the real value
$\hat{x}$ is the forecasted value, and t is the time script.

4 Design Specification and Implementation

The implementation of data modelling, analysis, interpretation, and evaluation was carried out in the following steps:

i. **Step One:** The datasets used for this study were provided by Kaiko- Digital Assets Data Provider. Two datasets were provided, one for Bitcoin Cash and the other for Ethereum Classic. These datasets range from January 2019 to November 2019. Both Bitcoin Cash and Ethereum Classic had 7373 observations.

ii. **Step Two:** In this step, the various packages used for data exploration, analysis, interpretation, and evaluation were installed. The installed packages include; ("zoo"), ("xts"), ("xgboost"), ("prophet"), ("forecast"), ("ggplot2"), ("timeDate"), ("timeSeries"), ("caret"), ("Car"), ("dplyr"), ("tidyverse"), ("tidymodels"),("parsnip"),("Keras") and ("tensorflow").

iii. **Step Three:** Features engineering was conducted at this stage, an entailed checking for missing values, dimensions, descriptive statistics of the datasets and investigating stationarity was performed.

iv. **Step Four:** Here, we partitioned the datasets into training and testing sets. The training datasets were the first eighty per cent of the series, while the twenty per cent constituted the testing dataset.

v. **Step Five:** The sixth step is the implementation of the data models. This stage involved the development of four predictive models; ARIMA XGBoost, PROPHET, and LSTM. Evaluation of the model required determining the best of these three models by estimating and comparing their RMSEs. The best model was thereby used for forecasting.

5 Implementation

5.1 ARIMA

ARIMA model was implemented by using forecast, tseries, lubridate and ggplot2 .
The provided dataset was pre-processed where there are no missing values and separate colum ns were removed. Then the data is converted into a time series by using ts() function. ARIM A forecasting model is processed by using auto.arima() function, that generates the best AIC or BIC value for the model and performs Auto- Correlation function to determine whether the data is noisy or not. From Figure 8 it explains Bitcoin Cash was highly volatile in March 2019 and similarly for Ethereum classic in June it attained high volatility.

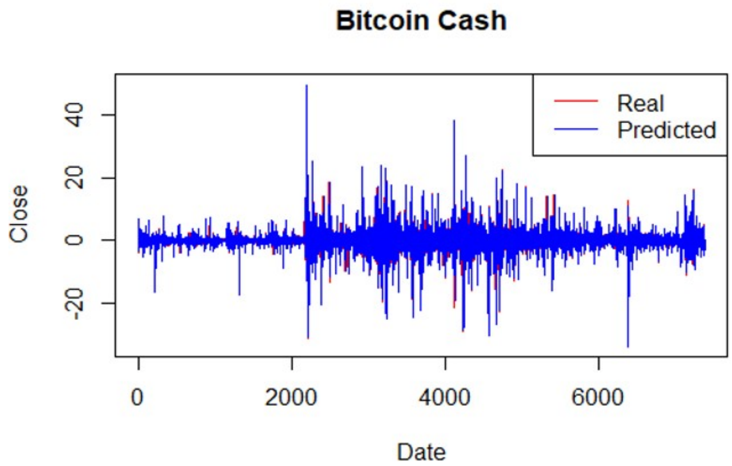

Figure 8 Bitcoin Cash

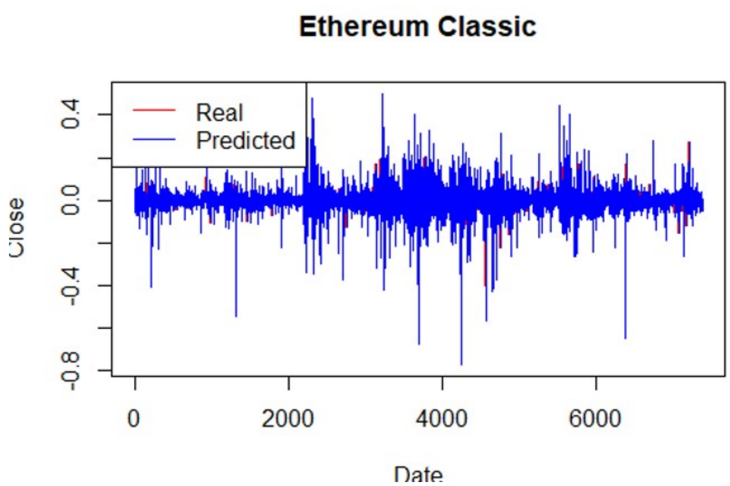

Figure 9 Ethereum Classic

5.2 XGBoost:

XGBoost is an open-source software library which provides a gradient boosting framework whereas R uses the XGBoost package, where the data is split as 80% for training and 20% for testing. After the split the mutate function is used to create a variable from the train data where we use multiple lags by the hour; organise a data frame which includes weekday and hour. Then we train and configure the model by hyper tuning boost parameters to retrieve the best fit for the model. We create a grid function, where we tune the mtry, then note a number which will project the prediction according to random sample at each split. We provide a length of 10 to 20, trees from 500 to 800 which is divided into 50 wise trees and for tree depth we provide range from the 8-10 which denotes the max no.of splits. These similar configurations are applied to the other dataset Ethereum Classic. After running multiple combinations of tries for Bitcoin Cash we select the least RMSE error which is 2.398 and For Ethereum Classic the

RMSE achieved was 0.0452, the error rate differs according to the price of the coin. Also, by using xgb.importance to understand each feature is important in building a boosted decision tree within a model, The feature importance is evaluated by using 3 scores. Gain is influenced features which contribute each tree in a model. The cover is the attempts correlated with the predictor and Frequency occurrence of a variable in ensembled trees.

Based on the score of Gain we plot a graph for the top 13 variables for bitcoin cash and Ethereum Classic which explains in Lag 5 as one the most essential feature, and in terms of Bitcoin Cash Lag 1 is the most important feature in building the model.

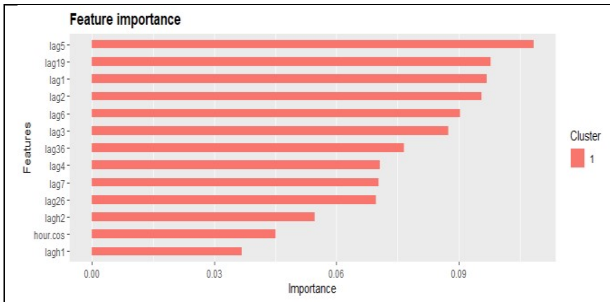

Figure 10 Ethereum Classic Feature Importance

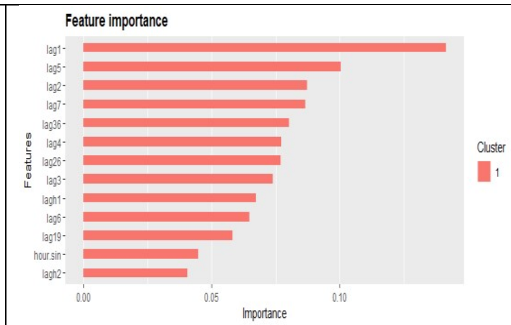

Figure 11 Bitcoin Cash Feature Importance

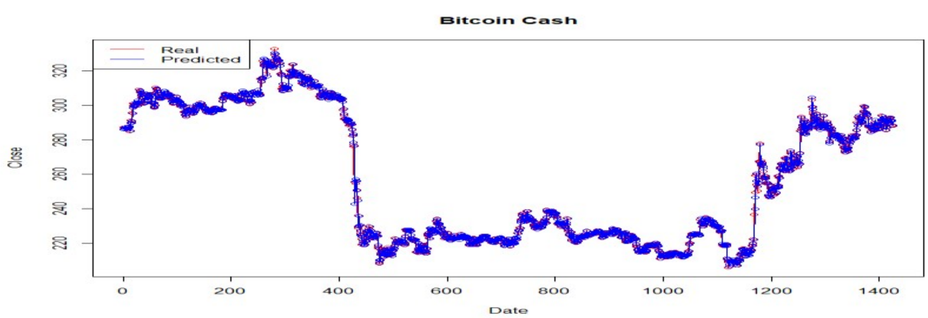

Figure 12 Bitcoin Cash

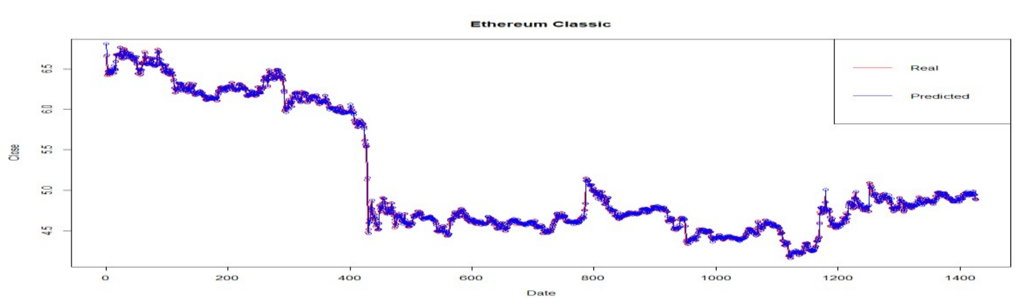

Figure 13 Ethereum Classic

The improvement of Gradient boosting with added features to make the computational speed is formed as XGboosting.

5.3 PROPHET

PROPHET is a code open source that is available in Python and R . To perform the time series analysis, we use the Bayesian curve fitting method, and it detects the seasonal trends automatically from the data with familiar parameters. After completing the 80:20 split for training and testing for both the datasets, The Prophet model is built on using **Prophet()** function followed by **Predict()** function to predict values. To forecast the future values, the **future data frame** function is used where **h** represents the no. of periods to be forecasted, and **frequency** is computed in hours (3600). The outcomes for Bitcoin Cash evaluated with the mean (yhat) value the RMSE of 35.67 and For Ethereum Classic 1.77. This figure depicts that the Dark dotted lines are the actual values and the blue line in between, is the predicted value and the shaded region at the end of the graph describes the yhat_upper and yhat_lower values.

Figure 14 Bitcoin Cash-Actual vs Predicted

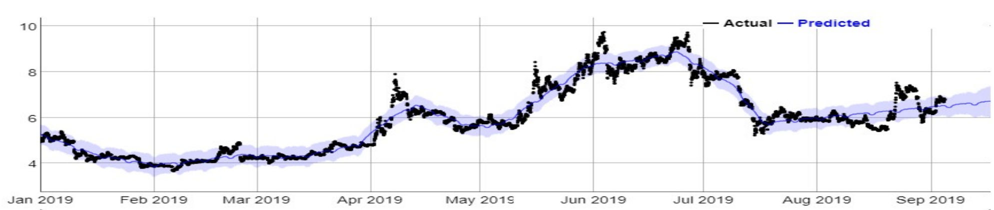

Figure 15 Ethereum Classic Actual vs Predicted

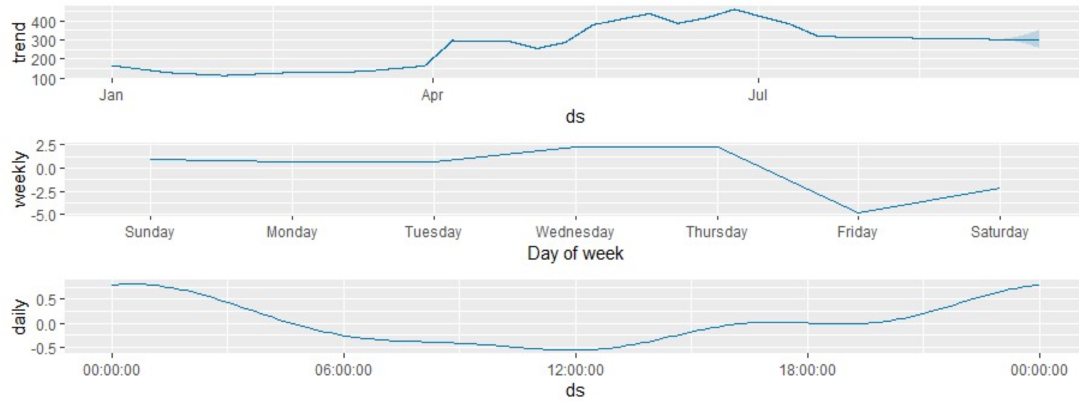

Figure 16 Bitcoin cash Component Plot

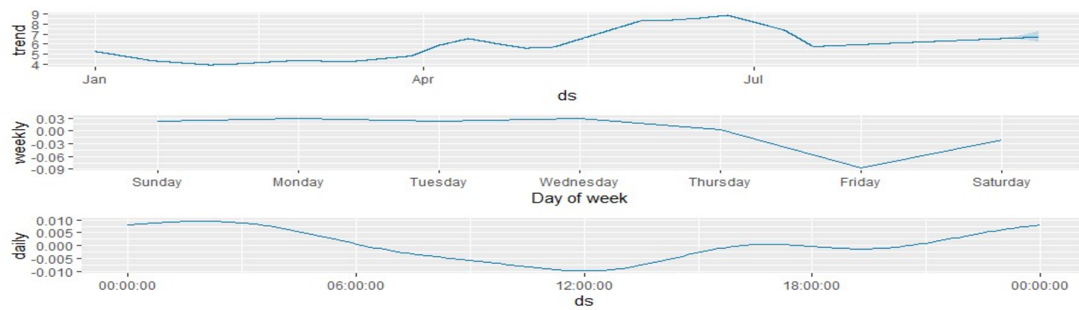

Figure 17 Ethereum Classic Component Plot

The component plot describes the trends from April to July, where the confidence interval is quite narrow. From the weekly graph, we can see that the values are high on Wednesday and Thursday and faces a sudden drop on Friday; recovers on Saturday and Sunday. From the daily hourly graph, we understand that prices are high at the start of the day and slowly reach the lowest value in the afternoon till the evening (around 6 pm) and gradually increase at around 7 PM.

5.4 LSTM

LSTM is designed to resolve the process of selecting as to which information is to be skipped and forgotten. LSTM is a special type of recurrent neural network that is capable for long term patterns. The core concept of Keras model is to arrange the layers in a sequential manner, to choose what to hold and what to remove. In order to forecast the values though LSTM we use keras,tensorflow,Mlmetrics and ggplot2 libraries. As of data, we use hourly variables Open, High, Low,Close, Volume Prices (OHLCV) prices of two fork coins collected from 01/01/2019 to 05/11/2019. To omit the non-stationarity we use difference in time series. In general LSTM random splitting is done. As these datasets involve time series we split 80% for train and 20% for test. In order to normalize the data, we use **feature range** option with default value (0,1). The **default sigmoid function** range from (-1,1) and impute the projected values to the original

scale. The main hyper parametrics is used to create a model are **Samples** which is the batch size which includes both training and testing. Timesteps and features are categorized as one dimensional with value 1. Accuracy metric is used to asses the model accuracy. We choose the optimal value as 50 for **epochs.**
A lesser value leads to underfitting and higher value leads to overfitting. Achieved RMSE of 0.35 for bitcoin cash and 0.012 for Ethereum Classic.

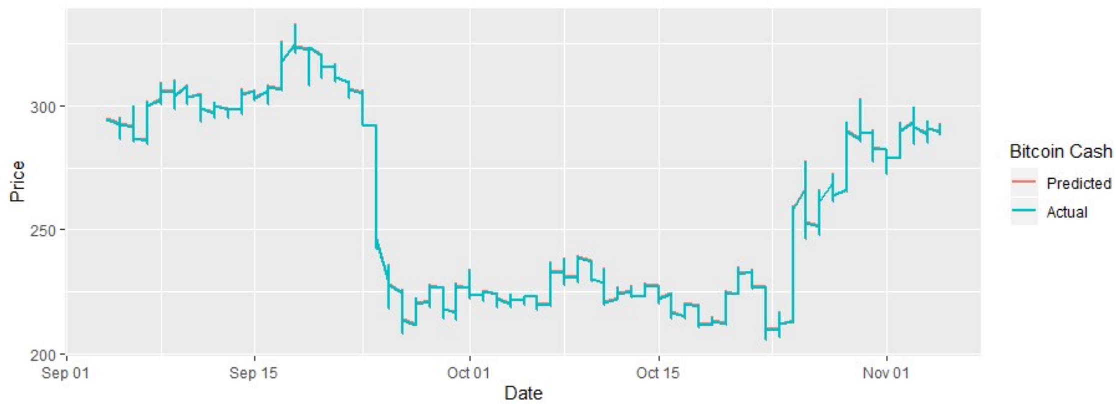

Figure 18 Bitcoin Cash Actual vs Predicted

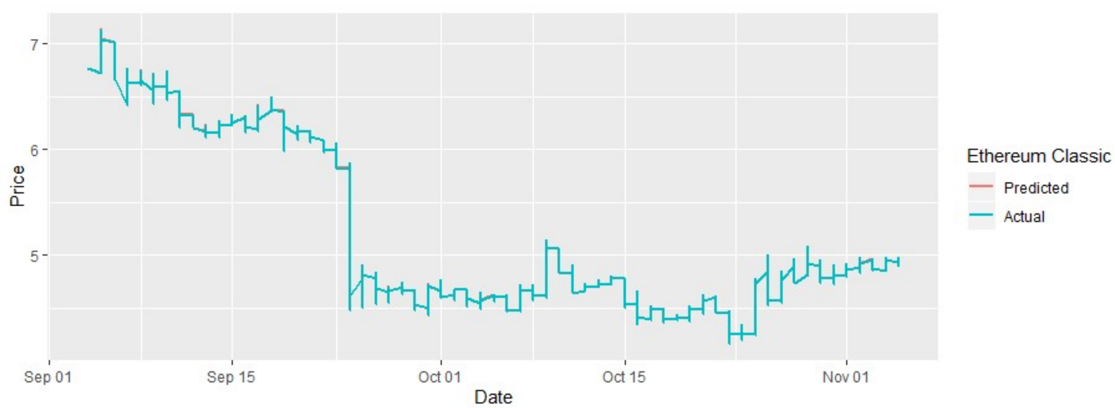

Figure 19 Ethereum Classic Actual vs Predicted

6 Evaluation and Discussion

Evaluation Metrics:

MODEL	BITCOIN CASH	ETHEREUM CLASSIC
ARIMA	2.387186	0.04501566
PROPHET	35.67221	1.778694
LSTM	0.3505858	0.001700209
XGBOOSTING	2.398	0.04532564

The table above presents different RMSEs of the four models (ARIMA, PROPHET, XGBOOST and LSTM) for each of the two forked cryptocurrencies (Bitcoin Cash and Ethereum Classic). It is believed that a model is better than others when the RMSE values generated by it is smaller when compared with other models (evaluation of performance using RMSE)

Discussion:

The above metrics indicate the following values for Bitcoin Cash, the RMSE value is 2.387186 for ARIMA, PROPHET 35.67, LSTM 0.350 and Xgboosting 2.398; and for Ethereum Classic it is 0.04501566, for ARIMA, PROPHET for 1.77, LSTM for 0.0017 and XGboosting is 0.045. Interpreting these values with RMSE, in both the fork cryptocurrencies the least error is achieved by LSTM that is 0.35 for Bitcoin Cash and 0.0017 for Ethereum Classic; followed by XGboosting, PROPHET. Finally, ARIMA presents a significant error rate. By analysing these values this research concludes that LSTM is the best performing model for forecasting of fork cryptocurrencies while PROPHET model stands last. The results obtained from LSTM analysis can be taken as a benchmark for the investors for their informed decision. However, this study used default parameters (refer 4.4 above)for LSTM model and the results can vary if any tuning in the parameter are performed. Further, the RMSE is calculated on both the coins as they have significant variations due to the volatility factors in closing price.

7 Conclusion and Future Work

A minimal amount of research available on forked cryptocurrencies which is built on the basis of upgradation of the parent cryptocurrency namely Bitcoin and Ethereum. This study analysed four machine learning methods namely: ARIMA, PROPHET, LSTM and XGBoost on fork cryptocurrency using time-stamped data to predict the closing price. The data consisted of hourly price ranging from January 2019 to November 2019 provided by Kaiko. These models are evaluated based on the RMSE performance metric. The results conclude that the LSTM model performed better while compared to other models.

Moreover, PROPHET model forecasted high error terms indicate that the other methods LSTM, XGBoost and ARIMA, perform better. Also, the results interpret that ARIMA performs better than XGBoosting, due to the addition of features variables in XGBoosting model, In future studies, ARIMA can add similar features like XGBoosting model. Further, based on this study investors can predict the value of the closing price of both the fork coins using LSTM algorithm and the investor have sound decisions. However, it is important to highlight that there are other factors which influence the performance of the sentiment analysis on both economic and social factors and should be considered by the investor before investing. This study explored only two fork cryptocurrency, however, the algorithm used in this paper can be replicated to other fork coins as well and can be part of future studies.

8 Acknowledgement

I am grateful to the almighty who has guided me in the right path of life and bestowed his grace on me. Secondly, I am immensely thankful to my parents who constantly encouraged and motivated me throughout this journey. Though miles away this wouldn't have been possible without my family's love and support. I would like to express my sincere gratitude and appreciation to all the people who invested their precious time for the completion of this report. Special thanks to my supervisor Mr Theo Mendonca whose brilliant guidance with stimulating suggestions led to the final output of my work. I would also like to thank Clara Medalie from Kaiko for providing me with the Data sets crucial for my research.

References

Andreas, M. and Wood, G. A. (2018) ʻIn mastering Ethereum: Building smart contracts and DApps', *Ethereum Timeline* O'Reilly Media; 6(1), p. 329-424. ISBN 9781491971918

Anupriya and Garg, S. (2018) 'Autoregressive Integrated Moving Average Model (ARIMA)- based prediction of Bitcoin Close Price,' *International Conference on Smart Systems and Inventive Technology* (ICSSIT). [Online] Available at: https://ieeexplore.ieee.org/document/8748423 [Accessed 26 Nov 2019].

Cannon, R. (2018) Bitcoin and Cryptocurrency: What you need to know about Bitcoin Cryptocurrencies and Blockchain Technology,1(2), p. 34- 42. ISBN:978-1-9867-1089-3

Chen, W., Zheng, Z., Cui, J., Ngai, E., Zheng, P., and Zhou, Y. (2018) 'Detecting Ponzi schemes on Ethereum', *Proceedings of the 2018 World Wide Web Conference on World Wide Web - WWW '18*. [Online] Available at: https://dl.acm.org/citation.cfm?id=3186046 [Accessed 2 Dec 2019].

Chen, Z., Li, C., and Sun, W. (2019) 'Bitcoin price prediction using machine learning: An approach to sample dimension engineering,' *Journal of Computational and Applied Mathematics*, [Online], Available at https://reader.elsevier.com/reader/sd/pii/S037704271930398X?token=B073612570059B1D4D8DFC77AB895D97C446A55668C8C8E296E867E7EADD9F9173ADF450350602037F52086CCBFB1776 [Accessed 29 Nov 2019].

Chohan, U. W. (2017) 'Cryptocurrencies: A brief thematic review, *SSRN Electronic Journal*, 7(1), pp. 2- 10.

Di Pierro, M. (2017) 'What is the blockchain?' Computing in Science and Engineering, 19(5), pp. 92-95. Available at https://aip.scitation.org/doi/abs/10.1109/MCSE.2017.3421554

ElBahrawy, A. L. Baronchelli, A. and Aiello, L. M. (2018) ‘Anticipating cryptocurrency prices using machine learning,’ [Online]. Available at: http://openaccess.city.ac.uk/id/eprint/19829/ [Accessed 21 Nov 2019].

Islam, A.K.M.N., Mäntymäki, M., and Turunen, M. (2019) ‘Why do blockchains split? An actor-network perspective on Bitcoin splits’, *Technological Forecasting, and Social Change* [Online] 148, p.119743. Available at: https://www.sciencedirect.com/science/article/pii/S0040162518319711 [Accessed 9 Dec. 2019].

Ji, S., Kim, J., and Im, H. (2019) ‘A comparative study of Bitcoin price prediction using Deep Learning mathematics’, [online] 7(10), p.898. Available at: https://www.mdpi.com/2227-7390/7/10/898 [Accessed 29 Nov 2019].

Karakoyun, E.S., and Cibikdiken, A.O. (2018) ‘Comparison of ARIMA Time Series Model and LSTM Deep Learning Algorithm for Bitcoin Price Forecasting’, *13th Multidisciplinary Academic Conference in Prague,* pp. 171-180.

Kiffer, L., Levin, D., and Mislove, A. (2017) ‘Stick a fork in it: Analyzing the Ethereum network partition,’ *16th ACM workshop on hot topics in networks*, pp. 94-100.

Kwon, Y., Kim, H., Shin, J., and Kim, Y. (2019) ‘Bitcoin vs Bitcoin Cash: Coexistence or Downfall of Bitcoin Cash?’ *IEEE Symposium on Security and Privacy*. [online] Available at: https://ieeexplore.ieee.org/document/8835333 [Accessed 9 Dec 2019].

Li, T.R., Chamrajnagar, A.S., Fong, X.R., Rizik, N.R. and Fu, F. (2019)‘Sentiment-Based prediction of alternative cryptocurrency price fluctuations using gradient boosting tree model, *Frontiers in Physics*, [online] Available at: https://www.frontiersin.org/articles/10.3389/fphy.2019.00098/full.

McNally, S., Roche, J., and Caton, S. (2018) ‘Predicting the price of Bitcoin using machine learning’, *26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP)*, [Online] Available at https://ieeexplore.ieee.org/abstract/document/8374483 [Accessed 27 Nov 2019].

Mittal, R., Gehi, R., and Bhatia, M. (2018) 'Forecasting the price of cryptocurrencies and validation using ARIMA'. [online] Ssrn.com. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3364967 [Accessed 27 Nov 2019].

Muhammad, F.A., Azah S.N., Mustapha, A., Razali, N. and Ahmad K.S. (2018) *'Regression-based analysis for Bitcoin Price Prediction*', International Journal of Engineering and Technology, 7 (4), pp. 1070-1073.

Nakamoto, S. (2008) 'Bitcoin: A Peer-to-Peer Electronic Cash System. [Online] Available at https://bitcoin.org/bitcoin.pdf. [Accessed 26 November 2019]

Bakar, N. A., and Rosbi, S (2017) 'Autoregressive Integrated Moving Average (ARIMA) Model for forecasting cryptocurrency exchange rate in high volatility environment: A new insight of bitcoin transaction, [Online] Available at https://ijaers.com/detail/autoregressive-integrated-moving-average-arima-model-for-forecasting-cryptocurrency-exchange-rate-in-high-volatility-environment-a-new-insight-of-bitcoin-transaction/ [Accessed 29 Nov 2019].

P, H., and Schmidsfelden, J, S. (2018) 'Bitcoin compared on price, liquidity, and volatility: Crypto "currencies" or an asset class of their own?', [Online] Available at: https://is.muni.cz/do/econ/sborniky/2018/Proceedings_finalni_verze_September_3.pdf#page=128 [Accessed 21 Nov 2019].

Rebane, J., Karlsson, I., Denic, S., and Papapetrou, P. (2018) 'Seq2Seq RNNs and ARIMA models for cryptocurrency prediction: A comparative study, [Online]Available at https://pdfs.semanticscholar.org/c1ec/480f005244a2cfb93f1d3ad15c2d22b864d6.pdf [Accessed 27 Nov 2019].

Shahsavari, Y., Zhang, K., and Talhi, C. (2019) 'A theoretical model for fork analysis in the Bitcoin network', *IEEE International Conference on Blockchain*

Taieb, S. B. and Hyndman, R. J(2013). *A gradient boosting approach to the Kaggle load forecasting competition*. [online] Available at: https://www.semanticscholar.org/paper/A-gradient-boosting-approach-to-the-Kaggle-load-Taieb-Hyndman/11cc72c0ade933a1ea20ea37d31829420a9bc5c3 [Accessed 28 Nov. 2019].

Taylor, S.J. and Letham, B. (2017) ‘Forecasting at scale’ [Online] Available at: https://peerj.com/preprints/3190/ [Accessed 23 Aug 2019].

Vujičić, D., Jagodić, D., and Ranđić, S. (2018) ‘Blockchain technology, Bitcoin, and Ethereum: A brief overview’, *International Symposium Infoteh- Jahorina*, pp.1-6.

Wirth, R. and Hipp, J. (2000). *CRISP-DM: Towards a standard process model for data mining*. [online] ResearchGate. Available at: https://www.researchgate.net/publication/239585378_CRISP-DM_Towards_a_standard_process_model_for_data_mining [Accessed 28 Nov 2019].

Yenidogan, I., Cayir, A., Kozan, O., Dag, T. and Arslan, C. (2018) ‘Bitcoin forecasting using ARIMA and PROPHET,’ *3rd International Conference on Computer Science and Engineering (UBMK*). [online] Available at: https://ieeexplore.ieee.org/document/8566476 [Accessed 2 Dec 2019].