Assessing the Predictability of Bitcoin Using AI and Statistical Models

Keshanth Jude Jegathees [1], Aminu Bello Usman[1], and Michael ODea[1]

[1] Department of Computer Science
York St John University, UK
keshanth.jude-jega@yorksj.ac.uk,
a.usman@yorksj.ac.uk,m.odea@yorksj.ac.uk

Abstract. This chapter analyses Bitcoin's predictability using AI and statistical models, and then identifies the model that produces the most accurate results. A multivariate time series dataset was created in order to train an AI (LSTM) and statistical model (ARIMA) to predict the price of Bitcoin over time. The LSTM model achieved the highest accuracy of 94 percent and the lowest MAPE of 5%. ARIMA had the best overall metrics, but when it came to forecasting the future, it performed poorly. The results show that it is possible to predict the BTC with a reasonable error rate; however, Bitcoin is extremely volatile, making it difficult to obtain results that can be confidently used to assert its value ahead of time. The result of the study suggested that it is possible to forecast the Bitcoin price with minimal error rates, refuting the null hypothesis. However, because the bitcoin price index is affected by a variety of external sources, forecasting time series problems is intrinsically challenging. When attempting to predict stated sort of data, the following constraints must be considered: the models (1) do not account for exogenous variable uncertainty, and (2) do not account for the fact that forecast-error variances vary with time (Fair, 1986).

Keywords: Bitcoin, Long Short-term Memory, Autoregressive integrated moving average, Cryptocurrencies, Time series analysis, ARIMA, LSTM

1. Introduction

The tradeoffs and hype for cryptocurrencies have largely grown in the past, especially it being extremely popularized in recent years, Cryptocurrency is a form of online currency that circulates without the need for a central monetary authority, such as a bank. It is a decentralized financial network, most popularly known for investment, but can also be used as an exchange for goods and services.

In 2009, a pseudonymous hacker calling himself Satoshi Nakamoto created "Bitcoin", the world's first completely virtual and decentralized currency (Bouoiyour and Selmi, 2015). It has become increasingly important as an electronic payment tool and a speculative financial asset, which makes bitcoin (BTC) a digital gold (Urquhart, 2016). Bouoiyour and Selmi (2015), suggest that Bitcoin is detached from macroeconomic fundamentals and behaves as a 'speculative

bubble', Bitcoin lacks the underlying value of real currency, meaning its value cannot be derived from neither consumption nor production (such as, gold).
Bitcoin has exponentially grown over the past decade, making it an attractive asset. The first block in the blockchain was mined in Jan. 3, 2009. During this age, bitcoin had no real monetary value, the first bitcoin transaction took place in May 2010, where 10,000 bitcoins were used to buy $25 worth of pizza. In 2011, Bitcoin passed its first threshold, achieving a worth of $1, the next couple of years were sailing smoothly, and finally in 2017 bitcoin was worth over $3000. The following years had high fluctuations, during June 2019 the price surpassed $10,000 and immediately surged to $6,635.84 by mid-December. Finally, after the pandemic caused the economic shut down, bitcoin resurfaced into activity, it started at $6,965.72. and by December it increased by over 400% in value. It reached its all-time high in mid-April 2021 over $63,000, which followed another surge which caused the prices to drop by 50%. Since then, the prices experienced extreme fluctuations (Edwards, 2022). This paper also discusses the possible factors that cause this effect in the chapters ahead.
Generally, the forecasting approaches can be categorized into two, Artificial Intelligent (AI), and Statistical models. Common examples of AI models include MLP (Multilayer Perceptron), ANN (Artificial Neural Network), RNN (Recurrent Neural Network). Some of the statistical models that kept repeating in all literatures were, Exponential Smoothing, ARIMA (Autoregressive Integrated Moving Average), GARCH (Generalized Autoregressive Conditional Heteroskedasticity). (Wang et al., 2012).

1.1 A blockchain enabled Cyber-Physical System

Cyber-Physical Systems (CPS) connect the physical world of machines and manufacturing facilities to cyberspace. Implementing CPS in the real world requires connectivity and a computational platform.

Bitcoins' underlying technology, blockchain, is a distributed ledger that can authenticate financial transactions and thwart fraud. Blockchain mitigates the risks of a centralized architecture by decentralizing the process of information validation across network peers. To put it another way, it's the most secure form of validation that also happens to be the most effective approach to speed up the provision of financial services, giving customers greater autonomy and control. Bitcoin has increased at an exponential rate over the last decade, making it an appealing asset. On January 3, 2009, the first block in the blockchain was mined. During this time, bitcoin had no monetary value; the first bitcoin transaction occurred in May 2010, when 10,000 bitcoins were used to purchase a $25 pizza. Bitcoin crossed its first boundary in 2011, reaching a value of $1; the next several years were smooth sailing, and by 2017, bitcoin was valued more than $3000. The following years saw significant changes; in June 2019, the price topped $10,000 and quickly dropped to $6,635.84 by mid-December. Finally, after the epidemic caused an economic shutdown, bitcoin emerged into activity, starting at $6,965.72 and increasing in value by nearly 400% by December. It reached an all-time high of $63,000 in mid-

April 2021, following another surge that led prices to tumble by 50%. Prices have fluctuated dramatically since then (Edwards, 2022). In the following chapters, this study will also address the probable variables that generate this impact.

This chapter focuses on investigating the predictability of Bitcoin price using AI and Statistical models and identifying the approach/model which provides the best results in predicting the closing price of bitcoin. Additionally, the impact of other external factors such as blockchain, mining, social attraction, and macroeconomics are also used in conjunction to the bitcoin historical data to try improving the accuracy of the predictions, creating a multivariate time series data. However, making accurate forecasts of this type is a challenging task due to the inherently noisy and non-stationary nature bitcoin price data.

2. Literature Review

There are several studies that revolve around achieving a similar goal to this literature. This section of the literature addresses them by the data that was used, and the types of approaches and their results.

Many literatures fall into one category, there are also papers which implement hybrid models. However, there are very few studies that compare both the two forecasting approaches. For instance, a study by authors Lee, Yoo and Jin, (2007), compare prediction performance of a NN and SARIMA model in forecasting the Korean Stock Exchange (financial time series data). Their results showed both models had robust forecasts, while the SARIMA model provided accurate forecasts for the Korea Composite Stock Price Index (KOSPI), the neural network model proved to better predict its returns. Xie, Ueda, and Sugiyama, (2021) implement a hybrid model with LSTM and MLP to forecast loads in operation management of a power system. Khashei, Bijari and Raissi Ardali, (2012) modeled a hybrid of ARIMA and PNN (probabilistic neural network) to obtain higher accuracy in time series prediction. Pai and Lin, (2005) combine the ARIMA and SVM (Support vector machine) model to exploit the unique strengths of each model in forecasting stock price

Study by Greaves and Au, (2015) uses blockchain network features to try predicting the prices of bitcoin, the researchers conducted supervised machine learning to achieve this. However, they reported an average accuracy of only 55% using an ANN model, they concluded blockchain data alone is inadequate for means of predicting bitcoin prices. Research by McNally, Roche and Caton, (2018) suggests, the most commonly used model for stock, and other time series prediction, Multilayer Perceptron (MLP) is not as effective as a Recurrent Neural Network (RNN) model. Agarwal and Sastry, (2015) prove this claim by successfully implementing predicting stock returns with a combined RNN and genetic optimization algorithm. Although, a study by Giles, (2001) supporting this claim conducted exchange rate prediction and obtained predictable results with a forty percent error rate, which is a significant gap compared to the literature by Valencia, Gómez-Espinosa, and Valdés-Aguirre, (2019), in which they compare three neural networks to predict four major cryptocurrencies price movement, out of which MLP outperformed with an accuracy rate over seventy two percent. Similar to the comparison made above, research by Raudys and Mockus, (1999) suggest the statistical model ARMA (which is comparatively less efficient than ARIMA) method is much faster than the MLP.

Some literatures that follow a statistical model approach; Cermak, (2017) applies the GARCH (1,1) model to analyze Bitcoin's volatility in respect to macroeconomic variables, results proved that Bitcoin appeared to be not only an attractive asset but also behave similarly to fiat currency in certain countries, especially China. Letra, (2016) also uses the same model to analyze the relationship between social trends and bitcoin prices. Ariyo, Adewumi, and Ayo, (2014) conducted a study in which the authors use an ARIMA model to predict stock price, and their research revealed the ARIMA model has a strong potential for short-term prediction. In their paper, they also argue that ARIMA models perform efficiently in forecasting financial time series data and especially in near future predictions than most popular ANN models. Similarly, Nochai and Nochai, (2006) apply the ARIMA model in forecasting palm oil prices of Thailand. Wang, (2011) compares ARIMA and Fuzzy time series method in forecasting Taiwan exports, in which ARIMA showed minor prediction errors and closer predictions for prolonged sample periods, whereas the fuzzy method performed well with smaller sample periods.

According to (Bitcoin.org, 2021), the price of a bitcoin is determined by supply and demand. When demand for bitcoins increases, the price increases, and when demand falls, the price falls. Research done by Kristoufek, (2015), identifies the main drivers of bitcoin price. The author conveniently categorizes each factor into areas such as, the Bitcoin Price Index (BPI), Macroeconomics & Policies, Technical Drivers, and Interest/Popularity. The author uses a first order AR model to identify correlation with wavelet coherent analysis. His literature mentions two, significant relationships between the Bitcoin price, firstly, is a negative relationship between the Bitcoin price and estimated output volume, i.e., an increase in output volume brings a drop in Bitcoin price notably in the long run. Secondly, bitcoin mining inherently leading its price, studies also indicate mining difficulty is positively correlated with the price (Kristoufek, 2015). Similarly, research done by, Ciaian, Rajcaniova and Kancs, (2016), tests the significance of the Market forces of bitcoin, it showed the number of bitcoins in circulation and transaction volume has a significantly high impact on bitcoin price. Unlike the fiat currency that is issued and supervised by the nation's central bank, the value of a Bitcoin is determined by how much the investors are willing to pay for it. (Teker, Teker and Ozyesil, 2019). The relationship between investors and their interest stimulates the price of bitcoin, during episodes of explosive prices, this interest drives prices further up, and during rapid declines, it pushes them further down (Kristoufek, 2015).

Study done by Aggarwal et al., (2019) discusses the social factors that govern the cryptocurrency market and analyzes the impact of these factors, the authors conclude there is no strong correlation between the cryptocurrency and social factors. On contrary, study by Wołk, (2020) focused on predicting the prices of cryptocurrencies like Bitcoin, Ethereum, and more, in relation to social media, more specifically twitter and google search trend data, the study made use six advanced predictive and descriptive models and sentiment analysis in conjunction to identify correlation between social influence and its effect on crypto prices. Similarly, many other studies have been conducted in support of social media data and its significance. Tandon et al., (2021) use the popular ARIMA model to predict the price of Bitcoin and Dogecoin in correlation to Tweets by popular tech idol, Elon Musk, the authors concluded by indicating no one person can control the utter volatile world of cryptocurrencies. Similarly, authors Matta, Lunesu and Marchesi researched the cross correlation between the bitcoin price to the volumes of tweets and web search data, with the use of sentiment analysis. Overall, the results indicate a significant positive relationship between the social factors and the change in price.

Google's Chief Economist Hal Varian suggested the potential of search data in speculating interests of a variety of economic activities in real time. Choi and Varian, (2009) support this claim by providing evidence that search data can predict home sales, automotive sales, and tourism, which are all time series data. Although Search Volume Index (SVI) does allow the opportunity to find patterns in attraction of a cryptocurrency, it does not link to investor interests. It is difficult to identify the motives of internet users searching for information about the Bitcoin (Kristoufek, 2015), the search term used in this, and many other literatures is “bitcoin”, therefore it is difficult to differentiate the effect, whether it is positive or negative (Kristoufek, 2013). SVI patterns analyzed by Da, Engelberg and Gao, (2011) claim that it primarily affected the decisions of individual or retail investors, meaning that more sophisticated retail investors do not show signs of correlation by these trends. Ciaian, Rajcaniova and Kancs, (2016) also support this claim, in their research, they conclude that search data (Wikipedia/Google) had a stronger impact in the starting years because when Bitcoin was little known, queries were made online about Bitcoin (views on Wikipedia).

There are only a limited literatures that take macroeconomics into consideration when forecasting bitcoin prices. A majority of the literatures analyze the significance of the affect it causes. Studies show that bitcoin appears to react to macroeconomic changes, for instance, in 2017, China suspended users from withdrawing bitcoins amid concerns cryptocurrency replacing its fiat currency, in reaction to this, bitcoin drastically dropped by over 30% within a week. Since its recovery and its peak, denial in application for the Bitcoin Trust ETF by the SEC caused yet again a 24% price crash (Cermak, 2017). Research by Bouoiyour and Selmi, (2015) describes the relationships between macroeconomic variables and bitcoin price. In his paper he finds no significant relationship between gold prices and Bitcoin prices. In support of this, Zhu, Dickinson, and Li, (2017) conducted an analysis on the macro factors that affect the Bitcoin price index (BPI) and concluded gold price has no impact in the BPI in the long run, the paper also identifies that the effect US Dollar Index has on the BPI in the long run is highly significant.

The reviewed literatures show evidence that both approaches have been implemented with success, and which exogenous variables have the most significance in the predictions. However, predicting cryptocurrency is still a widely researched topic, there are papers that approve and disapprove claims of another. Therefore, this paper will implement the strategies that were proven to be successful previously. In the aspect of data, a number of variables discussed in the literatures will be used in conjunction because cryptocurrency, bitcoin especially, has grown in different directions from when most papers were written.

The primary goal is to assess the predictability of the closing price of bitcoin and by doing so identifying which approach provides a prediction with higher accuracy and few complex variables. Finally, the paper's testable hypothesis are Bitcoin price can be predicted, and AI can be used to predict the Bitcoin price. The Null Hypothesis, or H0, in this case would be AI cannot be used to predict cryptocurrency.

3.0 Methodology and Experiment

3. Models used in the study: LSTM and ARIMA

This chapter attempts to answer the research question using an AI and a statistical approach, namely LSTM (Long Short-Term Memory) and ARIMA. Previous studies used machine learning techniques such as ANN and MLP to solve similar problems, similarly, ARIMA demonstrated enormous potential in numerous papers and has been widely used in price prediction problems.

The LSTM (Long Short-Term Memory) was first introduced by Hochreiter & Schmidhuber in 1997, LSTM is a type of deep neural network architecture that is recurrent, meaning that the output at one step is fed back into the system creating a loop (Tinawi, 2019). LSTM is a subset of RNN; however, LSTM can effectively solve the problem of long-term dependency, also known as the vanishing gradient problem in the RNN model (Guo et al., 2021).
Briefly understanding how an LSTM works; a typical LSTM network comprises of four major components, the cell state, and a series of 'gates' which control how the information in a sequence of data comes into, is stored in, and leaves the network, as shown in the image below. Namely, the forget gate, input gate, and output gate respectively. First step is deciding whether the information stored in the cell state needs to be disregarded or stored, this is achieved by implementing a sigmoid layer, which is the forget gate. Secondly, the input gate layer, decides whether to store the given information, following that the cell state is stored and the information that needed to be forgotten is replaced with the current state, and is finally output a filtered version of the cell state (Olah, 2015).

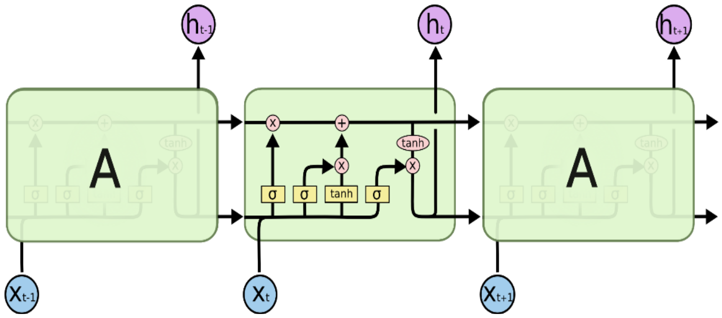

Figure 1: LSTM Network

Box and Jenkins in 1970 introduced the ARIMA model. The acronym ARIMA stands for Auto-Regressive Integrated Moving Average. The term "Autoregressive" implies the past values are used as predictors for the current values, this is also referred to as lagged values. In contrast to this, the "Moving average" component uses the lagged values prediction errors as its predictors. Finally, "Integration" simply means, differencing the stationary time series.

A general form of an ARIMA model is ARIMA(p,d,q)(P,D,Q), p refers to the order of the autoregressive process, d indicates the order of integration/differencing, and q is the order of moving average/error process. In each case the uppercase variables define the seasonal level of the model, while the lower case indicates non seasonal levels (Barão, 2008).

A general equation that defines the ARIMA model was presented in equation (1) . Where, Yt is the actual value and Et is the random error at tϕi and θj are the coefficients, p and q are integers that are often referred to as Autoregressive (AR) and Moving Average(MA), respectively. (Ariyo, Adewumi, and Ayo, 2014).

$$y_t = \varphi_0 + \varphi_1 \mathrm{Y_t} - 1 + \varphi_2 \mathrm{Y_t}-_2 + \cdots + \varphi_p \mathrm{Y_t} - \mathrm{p} + \varepsilon_t - \varphi_1 \varepsilon_t -_2 \ldots - \varphi_1 \varepsilon_t -_q \qquad (1)$$

3.1 Data and Variables

It is crucial to choose your sample points which defines your time series data correctly. The graph below shows bitcoin had begun its growth from the year 2017-2018, therefore data following that period should provide a good direction in learning the Bitcoin's price movement. The sample period spans to five years, starting from January 2017 to January 2022.

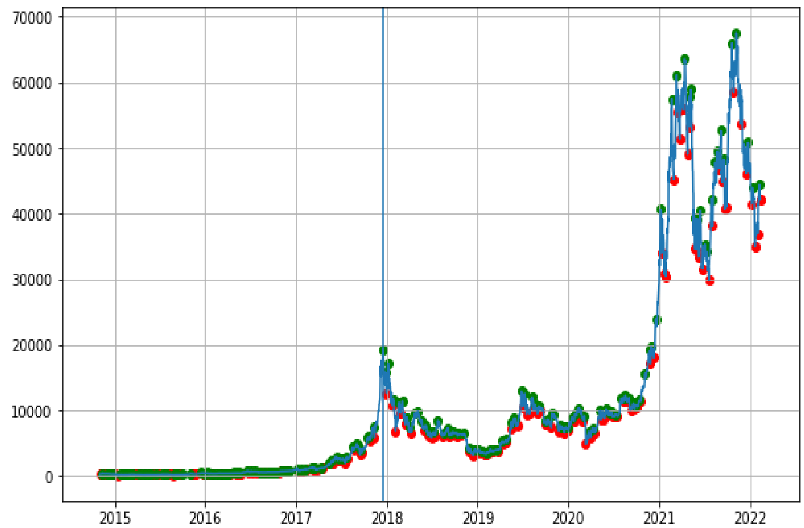

Figure 2: Data selection based on BTC peaks

Data that is used to train the models will be extracted from online resources that have been previously used by similar studies and have proved the authenticity of the sources. Primary source of data for this study is extracted from blockchain.com. More information on data sources available in Appendix A.

3.2 Dependent Variable

The dependent variable for this study is going to be the closing price of the cryptocurrency, as the goal is to predict the market value of Bitcoin at the end of the time period. The stock of money base of Bitcoins is denominated in a traditional government-controlled fiat currency, such as US Dollars, for the sake of comparability.

3.3 Independent Variables

The volatile nature of cryptocurrencies render it complicated to accurately forecast its value each day. There are several factors that need to be considered when trying to predict its market value. These variables were previously used in similar studies; Lamothe-Fernández et al (2020), Balcilar et al (2017), Da, Engelberg and Gao (2011), Bouoiyour and Selmi (2015), McNally, Roche and Caton, (2018), Greaves and Au, (2015), Kristoufek, (2015), Guo et al., (2021).

1. Demand and Supply - Although the Bitcoin is usually considered a purely speculative asset, standard fundamental factors—usage in trade, money supply and price level—play a role in Bitcoin price over the long term (Kristoufek, 2015). This also includes the number of mined bitcoins circulating on the network, the number of transactions per day, cost per transaction and miner's revenue, Exchange Trade volume, mining difficulty and hash rates following

Ciaian, Rajcaniova and Kancs (2016) and Lamothe-Fernández et al. (2020), Bouoiyour and Selmi, (2015).

2. Trends and Attractions - Search trends aggregated from the past to the present accumulatively depict the level of attraction a cryptocurrency has brought forth. A study done by Da, Engelberg and Gao, (2011) states the representation of internet search, most internet users use search engines to make these queries.

3. Macroeconomics – External factors such as gold price and the exchange rate of USD to CNY is considered to have a significant affect in the price of bitcoin. Because Gold and Bitcoin have many similarities. As for the exchange rates, historically, the cryptocurrency looks to have put in a positive performance during bouts of weakness in the Chinese currency. Table 1 lists all the selected independent variables.

Table 1: Independent variables

Variable	**Description**
(a) Demand and Supply	
Number of Bitcoins	Number of mined Bitcoins currently circulating on the network
Transaction value	Value of daily transactions
Transaction volume	Number of transactions per day
Exchange Trade volume	The total USD value of trading volume on major bitcoin exchanges.
Mining difficulty	Level of difficulty in mining a Bitcoin block
Hash rates	Times a hash function can be calculated per second
Cost Per Transaction	Miners' revenue divided by the number of transactions.
Output Value	The total value of all transaction outputs per day
(b) Trends and Attractions	
Google & Wikipedia Trends	Number of searches/views accumulated by the term "bitcoin".

(c) Macroeconomics	
Gold Prices	Gold price in US dollars per troy ounce
Exchange rate	The exchange rate between USD to CNY

3.4 Feature Selection

The selection of any model shouldn't rely only on its performance but instead, also on its complexity. Irrelevant or partially relevant features can negatively impact model performance (Brownlee, 2016). In order to select only the most significant IV and to achieve an optimal model complexity, we use measures of predictive accuracy. These measures are carried out in addition to calculating the p-values, J Hyndman, and Athanasopoulos, argue p-values could be misleading when two or more IV are correlated to one another, statistical significance does not always indicate predictive value.

There are various measures that calculate the predictive accuracy between the dependent and independent variables, such as R2, AIC (Akaike's Information Criterion), BIC (Schwarz's Bayesian Information Criterion), Cross-Validation, and there are also other techniques such as the Chi-Squared Test, and RFE (Recursive Feature Elimination) (Paul, 2020).

3.5 Unit Root Testing

Many time series include trend, cycles, and seasonality. When choosing a forecasting method, we will first need to identify the time series patterns in the data, and then choose a method that is able to capture the patterns properly (J Hyndman and Athanasopoulos, 2018).

Firstly, the data must be tested to check if it is stationary or not, not checking the stationarity degrees of the series, may result incorrect findings (Elder and Kennedy, 2001). When using models like ARIMA, it requires its dataset to be stationary. Unit root tests must be conducted on time series data to identify any non-stationarity or trends, if it is present, then it must be removed prior to analysis, as it will have no predictable patterns in the long-term (J Hyndman and Athanasopoulos, 2018). Therefore, stationarity will be tested using ADF (Augmented Dickey Fuller) test.

4. Results, Findings, Analysis, and Discussions

4.1 Time Series Analysis

Preparing a multivariate timeseries dataset is an intricate procedure. Merging multiple time series datasets without losing or overwriting data.
A shortcoming that occurred while collecting data was the frequency of data available on the google search hits (SVI) was weekly, while the rest of the time series data was collected on daily frequency. To cover for the lack of daily data for

the SVI variable, the dataset was modified with the feature's missing values interpolated with its mean values. Below is a table showing the summary statistics for the dataset. This sample is fragmented into portions which will be used for training (70%), validation (10%), and testing (20%).

Summary Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
close	1826	778.8397800	67553.94893	15470.65612	16954.38725
open	1826	778.8397830	67554.84000	15446.30799	16943.04089
high	1826	822.8602390	68990.90000	15900.04676	17426.13078
low	1826	748.6983680	66316.00000	14932.58260	16377.80915
estimatedtransactionvolumeusd	1826	112142904.0	1.46426E+10	2004008560	2112128161
ntransactions	1826	124640.0	490644.0	282361.699	55605.4601
hashrate	1826	2147763.141	198514005.7	74789091.54	54229901.87
difficulty	1826	3.17688E+11	2.50465E+13	1.03452E+13	7.52477E+12
costpertransaction	1826	5.612592289	300.3105491	72.06656832	58.82795010
Goldprice	1826	867.4000000	1888.700000	1377.316553	217.8559908
outputvolume	1826	421940.2071	24528670.35	1858919.703	1334891.637
tradevolume	1826	.000000000	4956849516	382889165.1	439863463.9
USDCNYPrice	1826	6.269000000	7.178600000	6.729909853	.2037990750
SVI	1826	4.000000000	100.0000000	18.57471264	5.753845318
Wikiviews	1826	1538	131165	8577.15	9960.120
Valid N (listwise)	1826				

Figure 3: Summary statistics for daily data

4.2 Feature Engineering

This paper conducted and compared the results from an AIC, Chi-squared, and RFE test. The model with the lowest value for AIC, and the top score of the other two tests is considered as the best model for forecasting.

A cross-validation grid search algorithm provided by Sci-Kit-Learn package was used, with the RFE model as its estimator and r2 as a score metric, it was evident the mean scores showed improvement after six features and remained constant after nine features (mean score= 0.999462). Similarly, a cross validation score with n number of parameters revealed the mean scores peaked at nine features (mean score=0.999435) and remained constant. This means, a minimum of six features should be sufficient for accurate results.

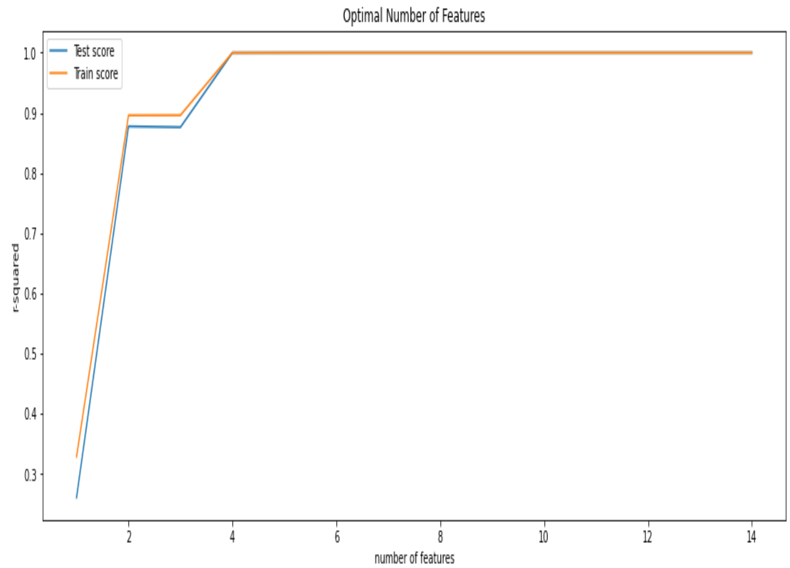

Figure 4: Feature selection: Optimal feature graph

Following this, a regression model was fit with the nine predictors that ranked first in the two tests (chi-squared and RFE), AIC score and the p-values for each predictor were calculated (AIC=2.703e+04, Adj.R2=0.99, Total features=14). The tests resulted in six features being selected (high, open, low, n-transactions, cost-per-transaction, and estimated-transaction-volume-usd), while other features scored above the significance level.

However, as discussed above, simply disregarding all variables whose p-values are greater than 0.05 is not always the right choice, a variable that is completely useless by itself can provide a significant performance improvement when taken with others (Guyon and Elisseeff, 2003). Which is why, visual, and statistical tests must be conducted before fitting the model

4.3 Feature Visualization

Study by Buchholz et al., (2012) suggested by visualizing the relationship between two variables should be a good estimate of significances. Below are a few graphs displaying the relationship between the BPI (Bitcoin price) and some of the predictor variables that proved to have a significance. The selected predictors have also been tested by previous literatures and results appeared to be the same.
Figure 6 shows subplots of three features, when visualized, show correlation to the dependent variable. These three features were chosen in regard to its high significance in previous literatures. The first plot shows the inverse correlation between the bitcoin price and the USD/CNY exchange rate. It is notable these two features are negatively correlated. In the second row, the first subplot depicts the Google Search hits (SVI) against the Bitcoin Price (BPI), although the SVI does not indicate whether a search emotion is positive or negative, it is evident that both

SVI and BPI displayed a positive correlation. The plot to its right shows the effect mining difficulty has on the bitcoin price. This was also discovered by Kristoufek, (2015), as mentioned in the author's literature, Mining difficulty is positively correlated to the bitcoin price

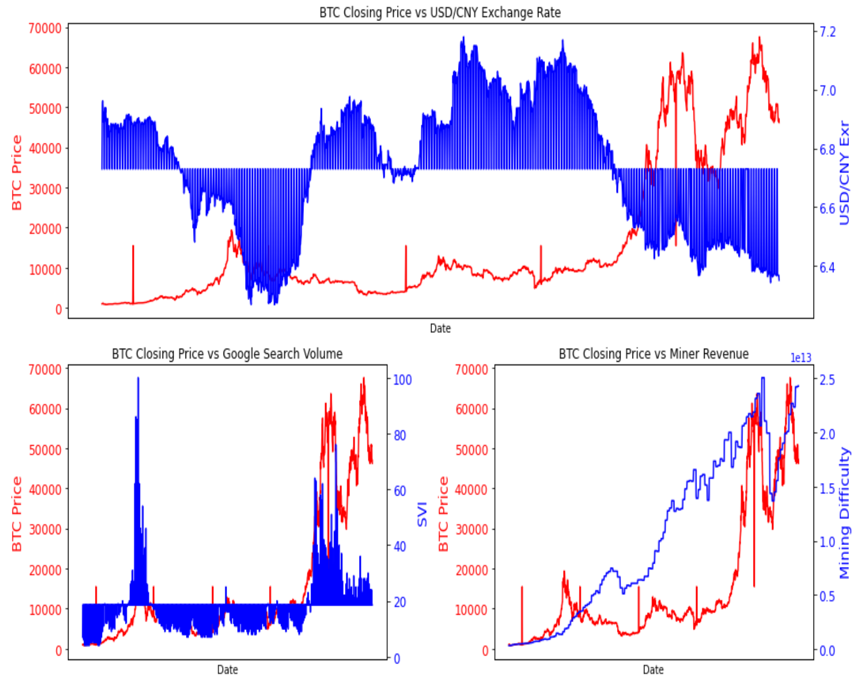

Figure 5: Correlation plots for features

It is also important to understand the difference between correlation and causation. A variable may be useful to forecast another variable, however that does not necessarily mean x is causing y. The above variables may not necessarily cause the bitcoin price to go up/down, but their change seems to be relative to the Bitcoin price

4.4 Unit Root Result

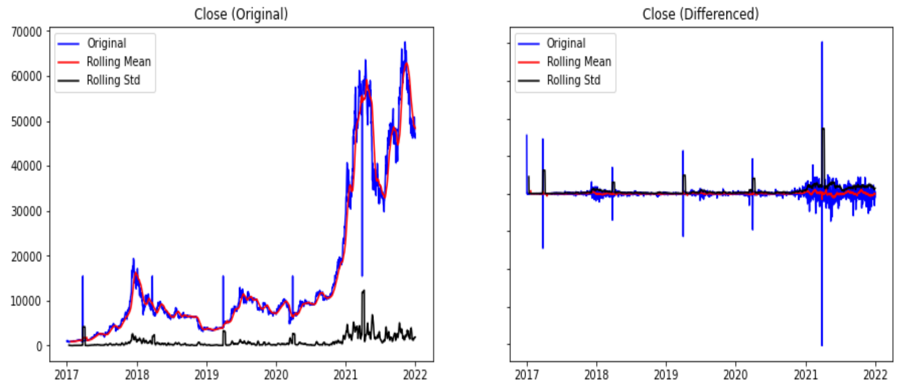

Figure 6: Original vs. Differenced Close Price

All columns were firstly visualized with ACF and PACF plots to understand the autocorrelation and to determine the number of differencing required to make the series stationary. Both plots should exhibit a decaying pattern, which is a sign the data is stationary. Unfortunately, all of the columns displayed a growing/reducing or stable pattern, which indicates the data is non-stationary.

ADF has a null hypothesis of a unit root ($d = 1$) against the alternative of no unit root ($d < 1$) (Kristoufek, 2013). All columns in the dataset were tested for stationarity, table in Appendix C shows the p-values before and after differencing the data. A majority of the features had p-values greater than the significance level before differencing. After a first order differencing, the p-values resulted to be significant (< 0.05). Therefore, a first order differencing should make the dataset stationary.
Further data preprocessing done for the models is explained in Appendix B.

4.5 Time Series Prediction

The following metrics are used in evaluating the accuracy of the model; Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE). Smaller values of MAPE and RMSE generally indicates more accurate predictions and good performance of a model. In the context of BTC price.

- MAE of 5 means that the predicted price is ± USD 5 from the actual price.
- MAPE quantifies the error in terms of percentage. For example, a MAPE of 3% can mean either USD 3 or 30 depending on whether the actual price is USD 100 or 1000, respectively
- RMSE indicates the spread of the forecast errors. A model that predicts erratic values will have a higher RMSE value, although it may have still had lower MAE or MAPE. Thus, the models should be evaluated with

respect to all the three metrics. (Mudassir et al., 2020)

4.6 LSTM Results

Four models were built and tested in search of the model with the best scores and lowest parameters. All four models were compiled with the Adam optimizer and mean squared error as the loss function. Selecting the prediction timestep (number of days to forecast) was an iterative process, the models seemed to perform poorly in windows that were too far ahead or too short, it's most effective timestep found was 30 days. For a time-series task two layers is enough to find nonlinear relationships among the data (McNally, Roche and Caton, 2018). LSTM models converged between 20 and 100 epochs, but it was found that anything over 20 epochs showed no improvements or was prone to overfitting, this may be due to the relatively small size of the dataset.
Graph below shows the results obtained by each model with validation (test) data, in terms of MAE and RMSE.

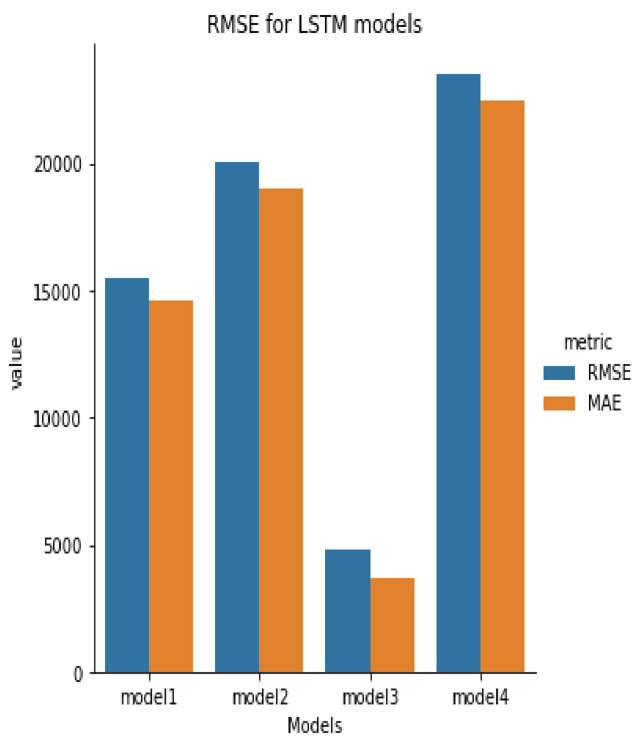

Figure 7: Accuracy Metrics: LSTM

Figure 8 shows a visualization of the predictions made by each model on the validation data. It can be observed, model3 has the closest prediction accuracy. Model3 outperformed the other models; it also had the lowest number of trainable params (20,289).

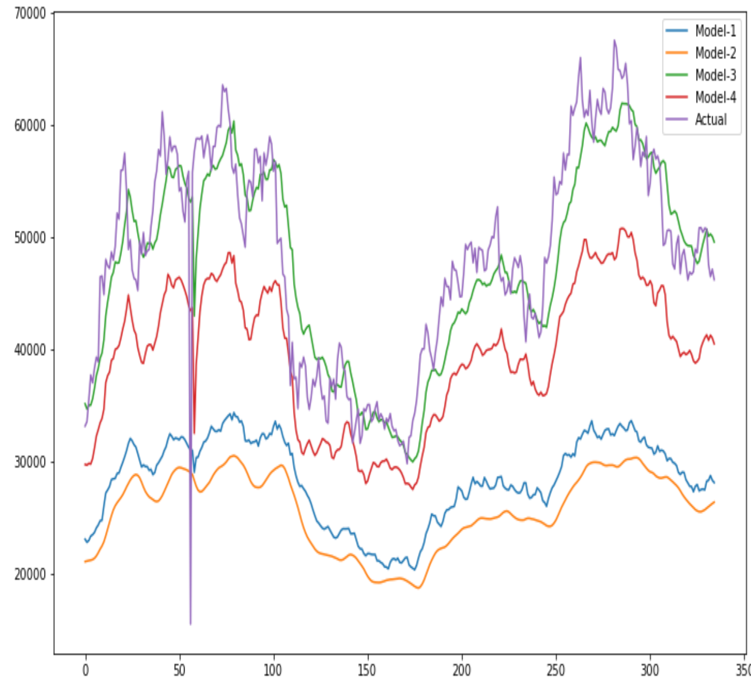

Figure 8: LSTM Model predictions comparisons

The results for the predictions made for the next 30 days showed drastic changes in accuracy metrics in comparison to validation data (MAPE=5.756%, RMSE=2696.141). The graph below shows the predictions made ahead of time compared to the actual prices during that time period, the predictions had an accuracy of 84.32%. The present data was collected using the Yahoo Finance API in python. The vertical dotted line shows the period where the training data ends, and the predicted and the present prices begin.

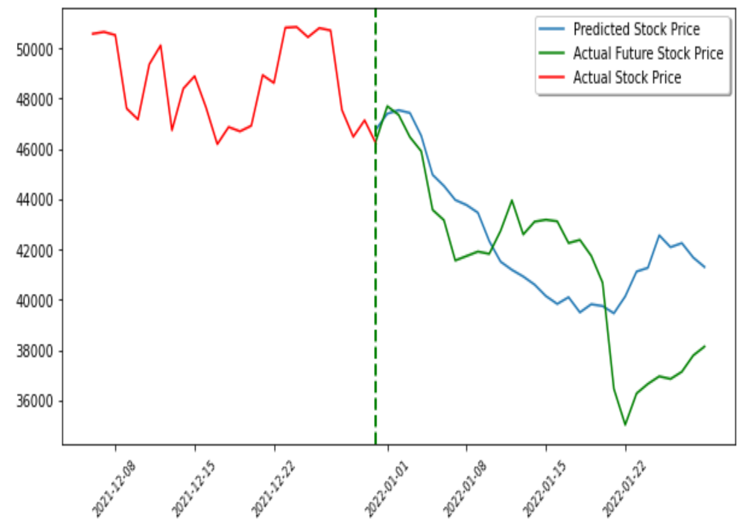

Figure 92: LSTM Future prediction graph

All models with the exception of one (Model2 with 4 hidden layers) had only 3 layers, Input, Hidden, and the Output layer. Model2 evidently took the longest time to train, only to be outperformed by models with less layers and complexity. Model4 and Model3 were built identically, however, Model4 was only trained with the features which were found as significant

4.7 ARIMA Results

The models were configured using the best params discovered from a Grid search procedure. This procedure fits a model with different combinations of orders and the model with a desirable AIC score is selected.

In addition to using ARIMA, two SARIMAX models were also fit using the same procedure. SARIMAX(Seasonal ARIMA with eXogenous factors) is an updated version of the ARIMA model where it also deals with exogenous variables. Below are the evaluation metrics on the validation data. All models produced impressive results compared to LSTM. However, the model performed poorly when trying to forecast the prices in the future.

Table 2 : ARIMA Results

Model (p,d,q)	**MAPE**	**RMSE**	**AIC**
ARIMA(1, 1, 2)	2.12%	1255.57	18872.110
SARIMAX(2, 1, 3)	1.79%	1019.90	18883.238
SARIMAX(1, 1, 2)	3.662%	1991.73	19270.158

Figure 11 shows the model's prediction on the test data against the actual values. The model has learnt the patterns from the historical data with high precision. The confidence intervals start to widen from June, which was also when the prices began to fluctuate heavily, which indicates the problem discussed in many literatures, the volatility of Bitcoin affecting its predictability

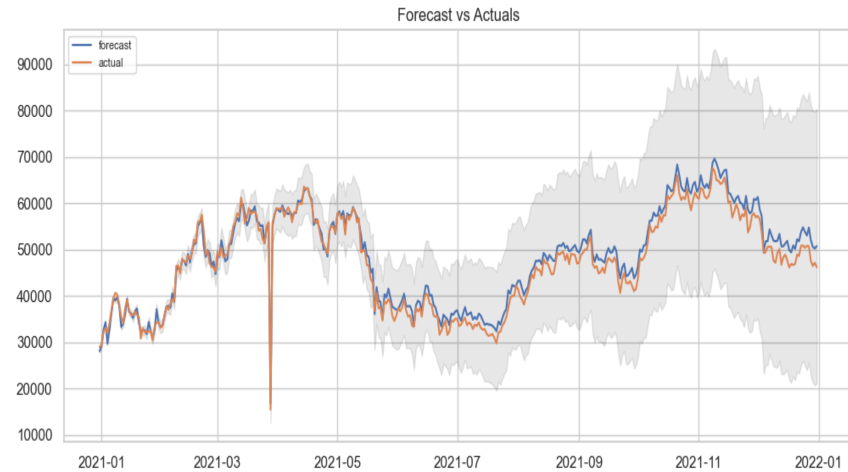

Figure 10: ARIMA Validation prediction

ARIMA is well suited for making short term predictions, the predictive accuracy worsened as the timesteps increased, the best results were found within 10-30 timesteps. All predictions start off high, and follow the actual values with decent accuracy, though there is a very high marginal gap in the prices, as displayed by the large confidence interval.

The graph below shows comparisons for n days of forecast against the actual values. Forecast 30 had a lower error (MAPE=18.141%) in comparison to forecast 60 (MAPE=48.994%). Although, forecast 5 had the overall lowest MAPE, it is too short of a time frame.

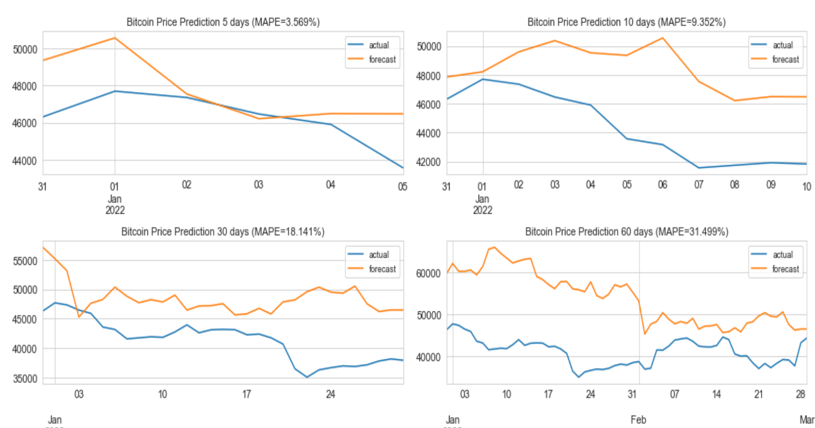

Figure 31: Prediction 5/10/30/60 Days ARIMA

30 days is an adequate time frame and also performs comparatively well; however, figure 13 shows a graph of the present prices compared with the predicted values with the historical data. The predictions appear to have identified the direction of change in the prices but was predicted marginally higher than the actual values (Approx. MAE ± USD 17794.66).

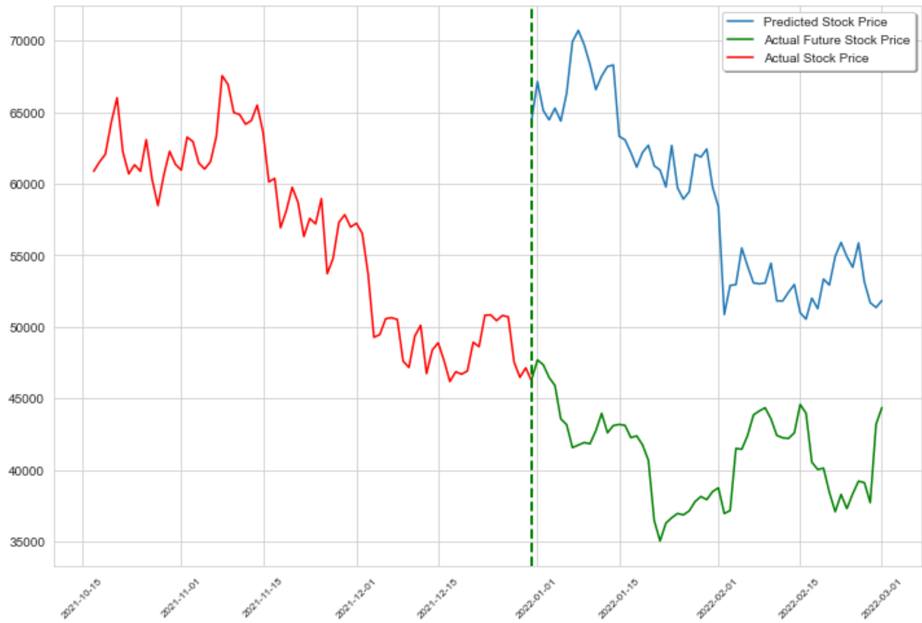

Figure 12: ARIMA Future predictions graph

Comparing predictive accuracy is not an easy task, its difficulty is why there is very few agreements as to which model produces the best predictions of bitcoin price.

4.8 Feature Significance

Models were trained/fit in two ways, using the features that were found statistically significant (sf) and all features (tf). The following models were trained using the features marked in the table below.

Table 3: Trained Models

Feature	***LSTM Models***		***ARIMA Models***
	M1, M2, M3	**M4**	
Close	X	X	X
Open	X	X	X
High	X	X	X
Low	X	X	X

Estimated Trans. Vol. USD	X	X	
N-Trans.	X	X	X
Hash Rate	X		
Cost Per Trans.	X		X
Gold Price	X		X
Output Vol.	X		X
Trade Vol.	X		
USD-CNY Exr.	X		X
SVI	X	X	X
Wiki Views	X		X

Table below shows the evaluation metrics on validation data by the features used from the best models in each approach.

Table 3: Evaluation metrics

	Sf		***tf***	
Model	***MAPE***	***RMSE***	***MAPE***	***RMSE***
LSTM : Model 3	34.833%	14601.12	5.756%	2696.14
ARIMA: ARIMA(1,1,2)	3.662%	1991.739	4.242%	2282.727

The ARIMA model's performance had no substantial changes when trained with sf or tf. However, the LSTM models performance improved drastically when trained with tf instead of sf.

5. Results Discussion

The graph below shows the Prediction-1, and Prediction-2 from the LSTM models that were trained with *sf* and *tf* respectively. It can be seen, the model that was trained with only the sf made predictions that were extremely higher than the actual values but appears to have identified the direction of change, while the other model produced significantly better predictions that fit closer to the actual price.

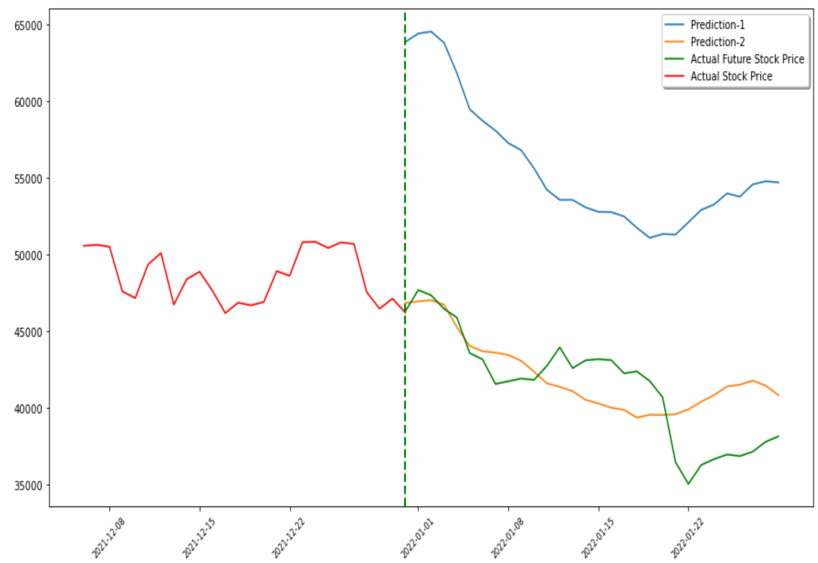

Figure 13: LSTM prediction with sf and tf

This proves that the features that were selected does help the model learn well, unfortunately it has a high error rate.
Figure 15 displays the evaluation metrics for all models, out of which the LSTM model produced the results with the highest accuracy, on the other hand ARIMA's evaluation had overall lower error than LSTM. Additional findings discussed in the Appendix D.

The results based on the training data shows both the models were able to successfully learn and predict the prices directions ahead a certain temporal frame. ARIMA performed significantly better on validation data compared to the LSTM models. On validation data the LSTM achieved MAPE of 8.08% while the ARIMA achieved 3.6%. Contrastingly, LSTM outperformed ARIMA when forecasting prices in the future.

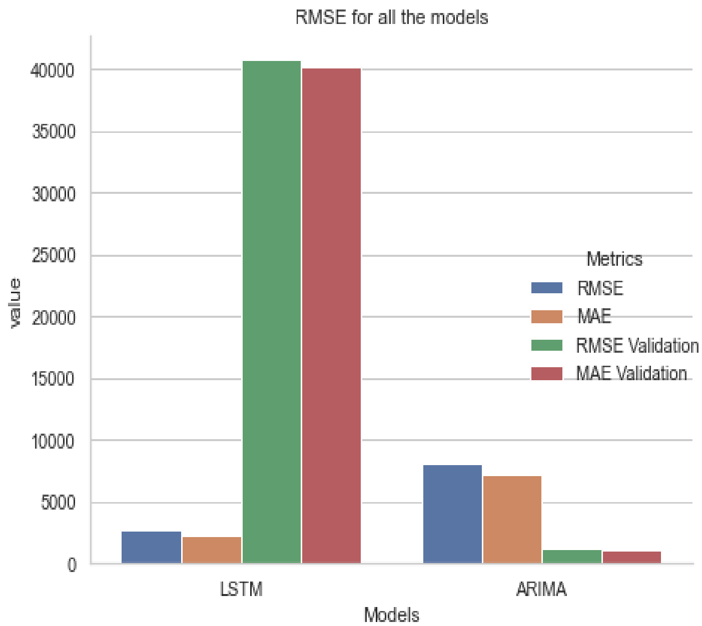

Figure 14: Evaluation metrics for all models

Although ARIMA had obtained the overall lowest metrics it does not indicate the model has good performance but has done a decent job in identifying the price's direction. Study by McNally, Roche and Caton, (2018) had also obtained excellent numbers in terms of error, however its predictive accuracy was imbalanced. Upon further analysis on the results of ARIMA, it was evident the model's predictions after the 20 days began to replicate the patterns and noise from the historical data with marginal errors. Figure 16 shows the predictions made for 90 days ahead using the past 90 days in the dataset. The green line shows the past prices relative to the given future dates, the forecast replicates the patterns and noises too closely.

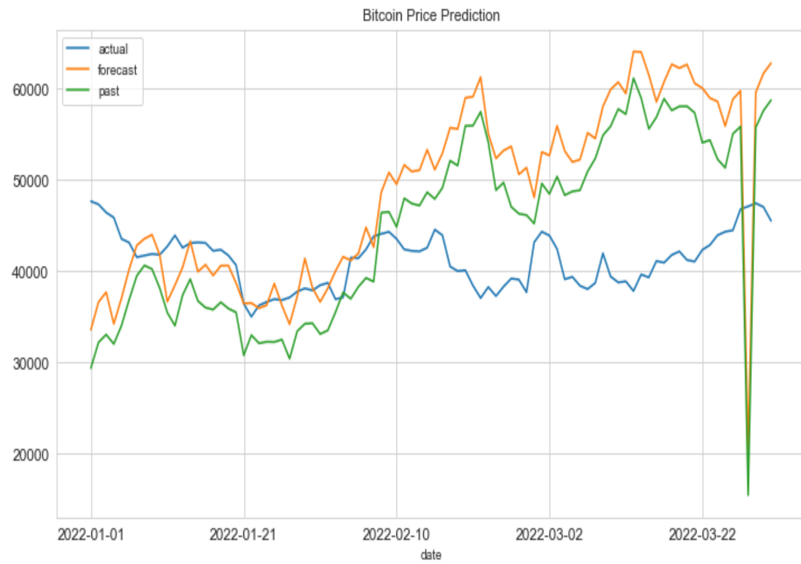

Figure 45: ARIMA Comparison chart for predicted/actual/past prices

Finally, the graph below shows both model's predictions and the present price of Bitcoin. Both models were given the past 30 days as input to forecast the next 30 days. ARIMA was unsuccessful in predicting the direction of the price ahead, instead it has made a forecast that looks a lot like the data it was trained with. LSTM, on the other hand, has done an excellent job in finding the direction of the price, though it is not precise, the line follows the actual price's direction

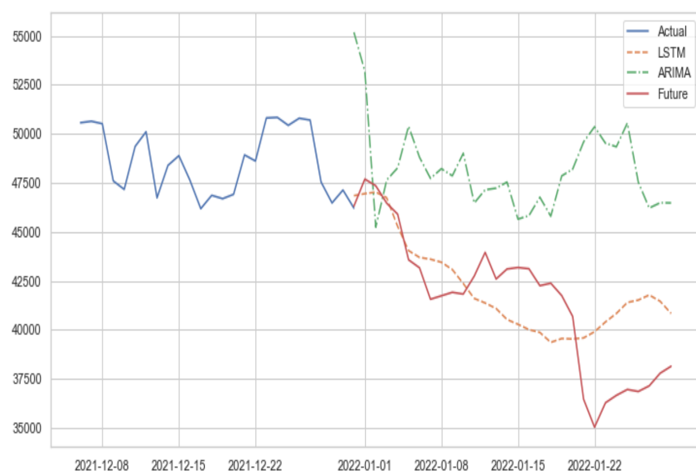

Figure 56: Model prediction comparison graph

6. Conclusion and Future Work

The results show it is possible to predict the Bitcoin price with low error rates, therefore rejecting the null hypothesis. However, prediction of time series problems are inherently difficult, especially the bitcoin price index, as it is affected by various external factors. The primary concern when trying to predict said type of data, its limitations must be taken into consideration, which are; the models (1) do not account for exogenous variable uncertainty, (2) do not account for the fact that forecast-error variances vary across time (Fair, 1986).

In this experiment, both models complement each other extremely well. LSTM is best suited to make long term forecasts, contrastingly, ARIMA is most useful when making short term forecasts. LSTM produced the best results when forecasting the future horizon (94.2% predictive accuracy, 82.7% validation accuracy) in comparison to ARIMA (82.7% predictive accuracy, 98.13% validation accuracy), ARIMA was not very good at making forecasts with past data, though it learned the historical data with high accuracy, LSTM on the other hand was able to identify the direction of change in the price successfully, however ARIMA performed much better in shorter time frames (Day 10 Accuracy=90.65%).

BTC prices are stochastic, and no given sets of features can provide a complete forecast (Mudassir et al., 2020). With that being said, many other researchers have previously used internal and external factors to classify the increase/decrease of BTC price, it was discovered the arrival of new information impacts the Bitcoin Price positively, for instance, the time series analysis revealed the queries made online about Bitcoin (SVI and Wiki views) had a stronger impact in the early stages and decreased gradually after Bitcoin became more established. There are various other social factors that could help improve the model's accuracy, for instance, Twitter sentiment data, which has already been used in previous literatures with debate to its usefulness. In addition to search trends, user activity in trading/forum sites could also be collected, this feature was also used by Lamothe-Fernández et al., (2020).

In conclusion, the results of this paper show it is possible to forecast BTC's direction with decent error rates, while it is extremely difficult to forecast its rise/fall in price with precision (Mudassir et al., 2020). The results are overall satisfactory and can be improved further by fine tuning the models and increasing the number of observations. Although, making forecasts for a highly erratic and volatile asset is prone to have errors and come with high risks when used for trading.

Future research on this study could focus on integrating tools that automate the data retrieval and updates the model constantly. It is said by the creator of Bitcoin, that Bitcoin will stop its supply after 21 million bitcoins. As of now, it has almost reached 90% of its maximum supply. This factor may affect the current model's predictions depending on how the market reacts to this. Therefore, progressively updating the model with new observations and data will lead to up-to-date results close to approximation. In addition to that, the models are built with minimal complexity, in the future more complex neural networks can be modelled to comprehensively learn the changes in the global events and combining an anomaly detection mechanism to assess the stability of Bitcoin and other exogenous factors.

References

Teker, D., Teker, S. and Ozyesil, M. (2019). Determinants of Cryptocurrency Price Movements. In: LAHSS-19, MEEIS-19 Nov. 12-14, 2019, Paris (France). 12 November 2019. Higher Education And Innovation Group. [Online]. Available at: doi:10.17758/HEAIG6.H1119510 [Accessed 18 November 2021].

Elder, J. and Kennedy, P. E. (2001). Testing for Unit Roots: What Should Students Be Taught? The Journal of Economic Education, 32 (2), pp.137–146. [Online]. Available at: doi:10.1080/00220480109595179 [Accessed 18 November 2021].

Selva, P. (2019). Augmented Dickey Fuller Test (ADF Test) – Must Read Guide. Augmented Dickey Fuller Test (ADF Test) – Must Read Guide. [Online]. Available at: https://www.machinelearningplus.com/time-series/augmented-dickey-fuller-test/ [Accessed 18 November 2021].

Lamothe-Fernández, P. et al. (2020). Deep Learning Methods for Modeling Bitcoin Price. Mathematics, 8 (8), p.1245. [Online]. Available at: doi:10.3390/math8081245 [Accessed 18 November 2021].

Balcilar, M. et al. (2017). Can volume predict Bitcoin returns and volatility? A quantiles-based approach. Economic Modelling, 64, pp.74–81. [Online]. Available at: doi:10.1016/j.econmod.2017.03.019.

Kristoufek, L. (2015). What Are the Main Drivers of the Bitcoin Price? Evidence from Wavelet Coherence Analysis. Scalas, E. (Ed). PLOS ONE, 10 (4), p.e0123923. [Online]. Available at: doi:10.1371/journal.pone.0123923.

Choi, H. and Varian, H. (2009). Predicting the present with Google trends. Google Research. Google Inc., Mountain View, CA.

Da, Z., Engelberg, J. and Gao, P. (2011). In Search of Attention. The Journal of Finance, 66 (5), pp.1461–1499. [Online]. Available at: doi:10.1111/j.1540-6261.2011.01679.x.

Bouoiyour, J. and Selmi, R. (2015). What Does Bitcoin Look Like? Annals of Economics and Finance, 16, pp.449–492.

J Hyndman, R. and Athanasopoulos, G. (2018). Forecasting: Principles and Practice (2nd ed). [Online]. Available at: https://Otexts.com/fpp2/ [Accessed 9 December 2021].

Guyon, I. and Elisseeff, A. (2003). An Introduction to Variable and Feature Selection. p.26.

Ariyo, A. A., Adewumi, A. O. and Ayo, C. K. (2014). Stock Price Prediction Using the ARIMA Model. In: 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation. March 2014. pp.106–112. [Online]. Available at: doi:10.1109/UKSim.2014.67.

McNally, S., Roche, J. and Caton, S. (2018). Predicting the Price of Bitcoin Using Machine Learning. In: 2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP). March 2018. pp.339–343. [Online]. Available at: doi:10.1109/PDP2018.2018.00060

Wołk, K. (2020). Advanced social media sentiment analysis for short-term cryptocurrency price prediction. Expert Systems, 37 (2), p.e12493. [Online]. Available at: doi:10.1111/exsy.12493.

Tandon, C. et al. (2021). How can we predict the impact of the social media messages on the value of cryptocurrency? Insights from big data analytics. International Journal of Information Management Data Insights, 1 (2), p.100035. [Online]. Available at: doi:10.1016/j.jjimei.2021.100035.

Brownlee, J. (2016). Feature Selection For Machine Learning in Python. Machine Learning Mastery. [Online]. Available at: https://machinelearningmastery.com/feature-selection-machine-learning-python/ [Accessed 13 December 2021].

Bitcoin.org. FAQ - Bitcoin. [Online]. Available at: https://bitcoin.org/en/faq [Accessed 14 December 2021].

Greaves, A. and Au, B. (2015). Using the Bitcoin Transaction Graph to Predict the Price of Bitcoin. p.8.

Kristoufek, L. (2013). BitCoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era. Scientific Reports, 3 (1), p.3415. [Online]. Available at: doi:10.1038/srep03415.

Urquhart, A. (2016). The inefficiency of bitcoin. Economics Letters, 148, 80–82.

Tinawi, I. (2019). Machine Learning for Time Series Anomaly Detection. p.55.

Rather, A. M., Agarwal, A. and Sastry, V. N. (2015). Recurrent neural network and a hybrid model for prediction of stock returns. Expert Systems with Applications, 42 (6), pp.3234–3241. [Online]. Available at: doi:10.1016/j.eswa.2014.12.003.

Giles, C. L. (2001). Noisy Time Series Prediction using Recurrent Neural Networks and Grammatical Inference. p.23.

Valencia, F., Gómez-Espinosa, A. and Valdés-Aguirre, B. (2019). Price Movement Prediction of Cryptocurrencies Using Sentiment Analysis and Machine Learning. Entropy, 21 (6), p.589. [Online]. Available at: doi:10.3390/e21060589.

Guo, Q. et al. (2021). MRC-LSTM: A Hybrid Approach of Multi-scale Residual CNN and LSTM to Predict Bitcoin Price. arXiv:2105.00707 [cs, q-fin]. [Online]. Available at: http://arxiv.org/abs/2105.00707 [Accessed 17 December 2021].

Wang, J.-J. et al. (2012). Stock index forecasting based on a hybrid model. Omega, 40 (6), pp.758–766. [Online]. Available at: doi:10.1016/j.omega.2011.07.008.

Cermak, V. (2017). Can Bitcoin Become a Viable Alternative to Fiat Currencies? An Empirical Analysis of Bitcoin's Volatility Based on a GARCH Model. SSRN Scholarly Paper, Rochester, NY: Social Science Research Network. [Online]. Available at: doi:10.2139/ssrn.2961405 [Accessed 17 December 2021].

Xie, Y., Ueda, Y. and Sugiyama, M. (2021). A Two-Stage Short-Term Load Forecasting Method Using Long Short-Term Memory and Multilayer Perceptron. Energies, 14 (18), p.5873. [Online]. Available at: doi:10.3390/en14185873.

Fair, R. C. (1986). Chapter 33 Evaluating the predictive accuracy of models. In: Handbook of Econometrics. 3. Elsevier. pp.1979–1995. [Online]. Available at: doi:10.1016/S1573-4412(86)03013-1 [Accessed 19 January 2022].

Matta, M., Lunesu, I. and Marchesi, M. (2015). Bitcoin Spread Prediction Using Social And Web Search Media. p.10.

Nochai, R. and Nochai, T. (2006). ARIMA MODEL FOR FORECASTING OIL PALM PRICE. p.7.

Khashei, M., Bijari, M. and Raissi Ardali, G. A. (2012). Hybridization of autoregressive integrated moving average (ARIMA) with probabilistic neural networks (PNNs). Computers & Industrial Engineering, 63 (1), pp.37–45. [Online]. Available at: doi:10.1016/j.cie.2012.01.017.

Pai, P.-F. and Lin, C.-S. (2005). A hybrid ARIMA and support vector machines model in stock price forecasting. Omega, 33 (6), pp.497–505. [Online]. Available at: doi:10.1016/j.omega.2004.07.024.

Zhu, Y., Dickinson, D. and Li, J. (2017). Analysis on the influence factors of Bitcoin's price based on VEC model. Financial Innovation, 3. [Online]. Available at: doi:10.1186/s40854-017-0054-0.

Olah, C. (2015). Understanding LSTM Networks -- colah's blog. [Online]. Available at: http://colah.github.io/posts/2015-08-Understanding-LSTMs/ [Accessed 17 February 2022].

Barão, S. M. M. (2008). Linear and Non-linear time series analysis: forecasting financial markets. [Online]. Available at: https://www.semanticscholar.org/paper/Linear-and-Non-linear-time-series-analysis%3A-markets-Bar%C3%A3o/c61536dba552c6a874f43c556e529eeb6e5409e3 [Accessed 17 February 2022].

Buchholz, M. et al. (2012). Bits and Bets Information, Price Volatility, and Demand for Bitcoin. p.48.

Paul, S. (2020). Feature Selection Tutorial in Python Sklearn. [Online]. Available at: https://www.datacamp.com/community/tutorials/feature-selection-python [Accessed 10 March 2022].

Edwards, J. (2022). Bitcoin's Price History. [Online]. Available at: https://www.investopedia.com/articles/forex/121815/bitcoins-price-history.asp [Accessed 17 March 2022].

Aggarwal, G. et al. (2019). Understanding the Social Factors Affecting the Cryptocurrency Market. arXiv:1901.06245 [cs]. [Online]. Available at: http://arxiv.org/abs/1901.06245 [Accessed 17 March 2022].

Mudassir, M. et al. (2020). Time-series forecasting of Bitcoin prices using high-dimensional features: a machine learning approach. Neural Computing and Applications. [Online]. Available at: doi:10.1007/s00521-020-05129-6 [Accessed 24 March 2022].

Appendix

4.1 Appendix A: Data retrieval sources

Bitcoin price index, and Blockchain data were obtained on 18 February 2021 from the Blockchain API (https://blockchain.com/). In addition to that, the other exogenous variables, such as Gold Price and Exchange Rates were extracted from Investing.com, (https://uk.investing.com/commodities/gold, https://uk.investing.com/currencies/usd-cny-historical-data, respectively). Search volume data were retrieved by accessing the Google Trends website (http://www.google.com/trends) on 22 February 2021 and the Wikipedia article traffic statistics site (https://www.wikishark.com/) on 2 February 2021.

All Bitcoin and Blockchain historical data were validated and cross checked for accuracy by comparing data in popular sources like CoinDesk (https://www.coindesk.com/price/bitcoin/), and Yahoo Finance (https://finance.yahoo.com/).

4.2 Appendix B: Model data pre-processing

LSTM: When training a network with data with large range of values, as large input values can slow down the learning and sometimes can prevent the network from learning effectively. Therefore, the dataset was scaled using the MinMax Scaler available in the sklearn library. This process scales the dataset values to fit between 0 and 1.

ARIMA: Selecting the order of the model is crucial to build a good model. Firstly, the data was explored as discussed in the Results section, unit root tests were conducted to identify the best order of differencing, and in search of the best model, a grid search with different orders was iteratively fit. Finally, the best model was determined by plotting the residuals and comparing AIC scores.

4.1 Appendix C: ADF Test scores

	Differencing Order		
Column	***no-diff***	***diff(0)***	***diff(1)***
close	0.868240		0.000000e+00
open	0.858296		0.000000e+00
high	0.873539		0.000000e+00
low	0.856106		0.000000e+00
estimated-transaction-volume-usd	0.361389		3.769493e-21
n-transactions	0.077388		1.535323e-19
hash-rate	0.899546		4.206086e-26
difficulty	0.948766		1.827273e-21
cost-per-transaction	0.577504		2.476612e-14
Gold price	0.885700		6.186569e-21
output-volume	0.023501	0.023501	
trade-volume	0.080964		6.361866e-22
USD-CNY Price	0.693700		9.939304e-21
SVI	0.006120	0.006120	
Wikiviews	0.016293	0.016293	

4.2 Appendix D: Findings

The boxplot shows the closing prices of each month throughout all the years, the price gradually grew during the first 4 months and began to decline over the next 5 months, and finally beginning to rise again. It must also be noted there are plenty of outliers following May, this could be due to the high gains for BTC during 2017 and 2019 May, which accumulated for over 50%. This could be what is causing the model to predictions with higher MAE, the smaller size the of observations was not comprehensive for the model to identify the outliers.

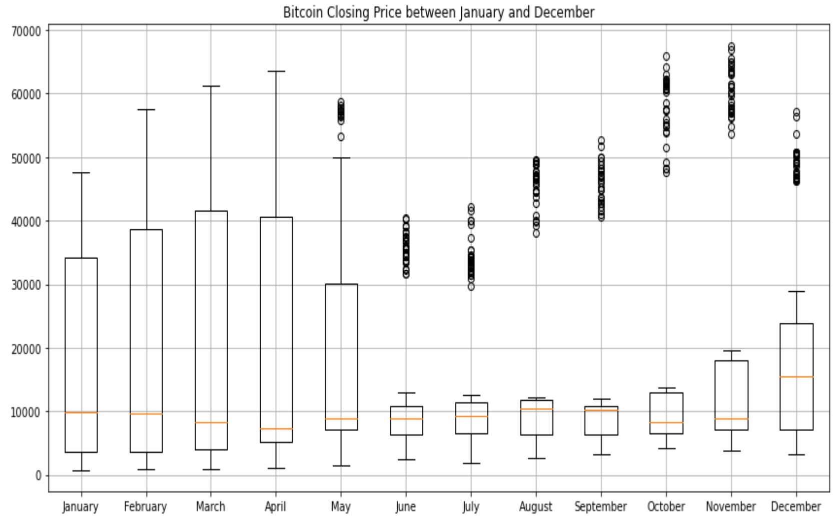

Figure 67: BTC Closing Price throughout the decade