Bitcoin Price Prediction using Time Series Forecasting Algorithm Arima

***Abstract*—We will analyze the predictability of the bitcoin market across prediction horizons. In doing so, we have to test various time series forecasting models and find have to find their outperform; such as arima and lstm. These algorithms are especially well-suited for the examined on time series data. We have to use some comprehensive feature set, including technical, blockchain-based, sentiment-/interest-based, and asset-based features. Then we will decide that which technical features are remain most relevant for most methods, followed by selected blockchain-based and sentiment-/interest-based features. Additionally, we will find that predictability increases for longer prediction horizons.**

Index Terms - bitcoin , crypto currency, block chain, arima

I. Introduction

Bitcoin is the biggest name in cryptocurrency in terms of both price index and market capitalization. Given the growing public investment into crypto, Bitcoin and other established currencies, such as Ethereum and Ripple, are facing increased competition from new, well-funded cryptocurrencies, which raised 52 million U.S. dollars in total in initial coin offering (ICO) in November 2019 [1].

Recently, virtual currency has gained more popularity as an accepted means for financial transactions. The cryptocurrencies are the most representative virtual currencies and receive more attention from the media and investors [2,3]. This is due to their attractive characteristics, such as simplicity, transparency, and increasing acceptance [4]. Bitcoin is the first and the most popular cryptocurrency on the market. It was implemented by Nakamoto [5] in January 2009 and is currently traded on over 40 exchanges worldwide and acceptable in over thirty different currencies [6].

Bitcoin allows people to sell and buy using different currencies. Bitcoins do not necessitate an institution or central bank to emit and control them. Therefore, the decentralization of Bitcoins makes possessors of Bitcoins feel safe. As Bitcoin is grounded on Blockchain as its primary database, it has some anonymity features. The username of a Bitcoin user is not disclosed during transactions; only their wallet ID is made public. Such features have made Bitcoin one of the most commonly used and valuable cryptocurrencies. Thus, Bitcoin is rising and has become an attractive investment for traders [7]. For traders or general users, the main issue is the Bitcoin exchange rate volatility. The excessive volatility of Bitcoin is a factor that prevents it being a currency; however, this volatility is thus far a motivation for traders [8]. Meanwhile, the general public are seeking solutions to cut down their risk. In fact, the Bitcoin price is remarkably volatile and changeable within a very short period of time.

II. Lirerature Review

"Bitcoin Price Prediction Using Ensembles of Neural Networks" [9]. Here they explored the relationship between the features of Bitcoin and the next day change in the price of Bitcoin using an Artificial Neural Network ensemble approach called Genetic Algorithm based Selective Neural Network Ensemble, they have constructed the neural network using Multi-Layered Perception. To better understand the practicality and its effectiveness in real world application, the entity was used to predict the next day direction of the price of Bitcoin given a set of approximately 200 features of the cryptocurrency over a span of 2 years. Over a span of 50 days, a trading strategy based on the ensemble was compared against a —previous day trend following‖ trading strategy through back-testing. The former trading strategy generated almost 85 percent returns, outperforming the —previous day trend following‖ trading strategy which produced an approximate 38 percent returns and a trading strategy that follows the single, best MLP (Multilayer Perceptron) model in the ensemble that generated approximately 53 percent in returns.

"A Research On Bitcoin Price Prediction Using Machine Learning Algorithms" [10]. In this paper they try to predict the Bitcoin price accurately taking into consideration various parameters that affect the Bitcoin value. Firstly they identify daily trends in the Bitcoin market while gaining insight into optimal features surrounding Bitcoin price. Their data set consists of various features relating to the Bitcoin price and payment network over the course of five years, recorded daily. Secondly by using the available information, they predict the sign of the daily price change with highest possible accuracy. Their methodology was to collect the data firstly from the database Quandl and CoinmarketCap they work on the time series data which was recorded daily five years at different time instances that was normalized and smoothened. To do this they implemented different normalization techniques like log transformation, z-score normalization, boxcox normalization, and so on. Interestingly they used two methods in two phases to predicting the price of the bitcoin.

"Predicting the Price of Bitcoin Using Machine Learning" [11]. In this paper their goal is to ascertain with what

accuracy can the direction of Bitcoin price in USD can be predicted. The price data is sourced from the Bitcoin Price Index . Their task is achieved with varying degrees of success through the implementation of a Bayesian optimised recurrent neural network (RNN) and Long Short Term Memory (LSTM) network. The LSTM achieves the highest classification accuracy of 52 percent and a RMSE of 8 percent. The popular ARIMA model for time series forecasting is implemented as a comparison to the deep learning models. As expected, the non-linear deep learning methods outperform the ARIMA forecast which performs poorly. Wavelets are explored as part of the time series narrative but not implemented for prediction purposes. Finally, both deep learning models are benchmarked on both a GPU and a CPU with the training time on the GPU outperforming the CPU implementation by 67.7 percent.

"Bitcoin price prediction using machine learning: An approach to sample dimension engineering" [12]. from this article we see how to predict Bitcoin price at different frequencies using machine learning techniques, they first classify Bitcoin price by daily price and high-frequency price. A set of high-dimension features including property and network, trading and market, attention and gold spot price are used for Bitcoin daily price prediction, while the basic trading features acquired from a cryptocurrency exchange are used for 5-minute interval price prediction. Statistical methods including Logistic Regression and Linear Discriminant Analysis for Bitcoin daily price prediction with high-dimensional features achieve an accuracy of 66 percent, outperforming more complicated machine learning algorithms. Compared with benchmark results for daily price prediction, they achieved a better performance, with the highest accuracies of the statistical methods and machine learning algorithms of 66 percent and 65.3 percent, respectively. Machine learning models including Random Forest, XGBoost, Quadratic Discriminant Analysis, Support Vector Machine and Long Short-term Memory for Bitcoin 5-minute interval price prediction are superior to statistical methods, with accuracy reaching 67.2 percent. To sum up the investigation of Bitcoin price prediction can be considered a pilot study of the importance of the sample dimension in machine learning techniques.

"Bitcoin Price Prediction Using Machine Learning" [13]. In this paper, they use the LSTM version of Recurrent Neural Networks, pricing for Bitcoin. To develop a better understanding of its price influence and a common view of this good invention, they first give a brief overview of Bitcoin again economics. After that, they define the database, including data from stock market indices, sentiment, blockchain and Coinmarketcap. Further in this investigation, they demonstrate the use of LSTM structures with the series of time mentioned above. In conclusion, they draw the Bitcoin pricing forecast results 30 and60 days in advance. All in all, predicting a price-related variable is difficult given the multitude of forces impacting the market. Add to that, the fact that prices are by a large extent dependent on future prospects rather than historic data. However, using deep neural net-works has provided us with a better understanding of Bitcoin, and LSTM architecture. The work in progress, includes implementing hyper parameter tuning, in order to get a more accurate network architecture. They thought it's difficult to give a mature thought on Bitcoin for the future.

"Future Price of Bit Coin Prediction Using Machine Learning Model" [14]. In this paper, they endeavor to anticipate the Bitcoin cost precisely thinking about different boundaries that influence the Bitcoin esteem. For the primary period of their examination, they comprehend and distinguish every day patterns in the Bitcoin market while acquiring understanding into ideal highlights encompassing Bitcoin cost. The informational index comprises of different highlights identifying with the Bitcoin cost and installment network throughout five years, recorded every day. For the second period of our examination, utilizing the accessible data, they anticipate the indication of the everyday value change with most noteworthy conceivable exactness. The world has in excess of 5000 advanced monetary forms, bitcoin is one of it, which has more than 5.8 million powerful customers and roughly in excess of 111 trades all through the world. Thus, the main focus on this paper is to do the close to forecast of the cost of Bitcoin in USD.

"The use of artificial neural networks in the analysis and prediction of stock prices," [15]. Instead of directly forecasting the future price of the stock, the authors predict trend of the stock. The trend can be considered as a pattern. They perform both short term predictions (day or week predictions) and also long-term predictions (months). They found that the latter produced better results with 79 percent accuracy. Another interesting approach the paper reflects is the performance evaluation criteria of the network. Based on the predicted output the performance evaluation algorithm decides to either buy, sell or hold the stock.

Also "Data mining on Romanian stock market using neural networks for price prediction" [16] , a comparison between Multi-Layer Perceptron (MLP) and Non-linear autoregressive exogenous (NARX) model is made. They conclude that MLP can also be used for stock market prediction even though it does not outperform NARX model in price prediction. The authors made use of MATLAB's neural network toolbox to build and evaluate the performance of the network.

At the very end we read a research named "Automated Bitcoin Trading via Machine Learning Algorithms" [17] they apply machine-learning algorithms to predict Bitcoin price. For the first phase of their investigation, they identify daily trends in the Bitcoin market while gaining insight into optimal features surrounding Bitcoin price. The data set consists of over 25 features relating to the Bitcoin price

and payment network over the course of five years, recorded daily. Using this information we were able to predict the sign of the daily price change with an accuracy of 98.7 percent. For the second phase of their investigation, they work on the Bitcoin price data alone and leveraged data at 10-minute and 10-second interval time points, to evaluate price predictions at varying levels of granularity and noisiness. By predicting the sign of the future change in price, they are modeling the price prediction problem as a binomial classification task, experimenting with a custom algorithm that leverages both random forests and generalized linear models. These results had 50-55 percent accuracy in predicting the sign of future price change using 10 minute time intervals.

III. Methodology

As it is a time series prediction. We have chosen ARIMA model as ARIMA model is a class of statistical models for analyzing and forecasting time series data.

Link: https://otexts.com/fpp2/arima.html

IV. Data Set

Size of the Bitcoin blockchain from January 2009 to July 12, 2021. (Statista)

Link: https://www.statista.com/statistics/647523/worldwide-bitcoin-blockchain-size

As Bitcoin is a kind of stock traded in stock market, dataset is available in plenty with all time intervals. We have collected live data from this data source during the period 2015 to till date. This provided us the most comprehensive bitcoin price in date wise data. Dataset is extracted to CSV file.

The features and attributes of the data are Timestamp, High, Low, Close, Volume(BTC), Volume(Currency), and Weighted Currency. There are 4857376 data in the dataset, but there are also many null values in the dataset. So that's why we have removed all the null values from our dataset. After removing the null values, the final dataset has a total of 3613769 data.Fig.1 shows how many null values we have in our dataset, and Fig.2 shows the average value of each of our features.

V. Result and Analysis

In this section, We will show different types of analysis on our dataset, and finally, we will show the result of our predictive model.Fig.3 shows the historical view of all the prices.

We have also found our exponentially weighted moving averages of Bitcoin price. Fig.4 shows both the original and ewma(exponentially weighted moving average) of BTC.

The main reason for finding our ewma is to find the outliers of our data. After finding ewma we have to find index-weighted moving average and then exponentially weighted

```
Timestamp              0
Open             1243608
High             1243608
Low              1243608
Close            1243608
Volume_(BTC)     1243608
Volume_(Currency)  1243608
Weighted_Price   1243608
dtype: int64
```

Fig. 1. Null Values Count

```
Timestamp          1.497898e+09
Open               6.009024e+03
High               6.013357e+03
Low                6.004488e+03
Close              6.009014e+03
Volume_(BTC)       9.323249e+00
Volume_(Currency)  4.176284e+04
Weighted_Price     6.008935e+03
dtype: float64
```

Fig. 2. Average Mean Value of the features

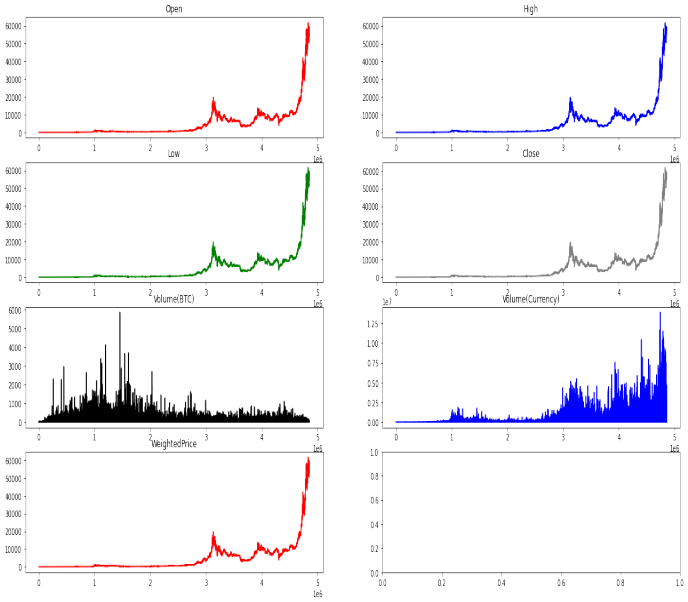

Fig. 3. Historical view of the the prices

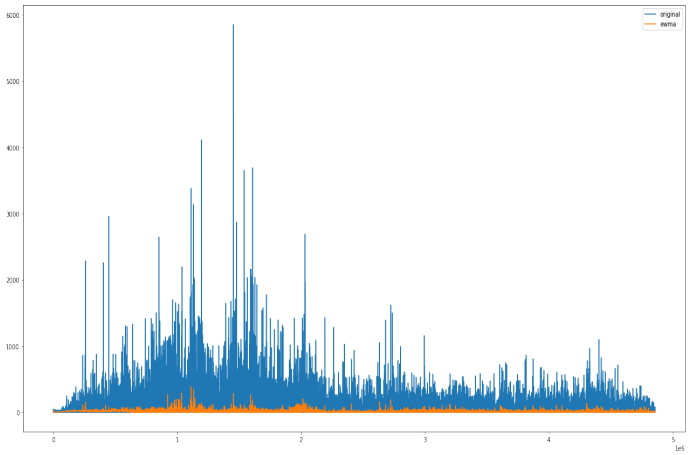

Fig. 4. Historical view of orginal and ewma BTC

moving standard deviation. Then we have Plotted data that are more than 3.0 times out of the standard deviation as outliers. Outliers are data points that are far from other data points. In other words, they're unusual values in a dataset. Outliers are problematic for many statistical analyses because they can cause tests to miss significant findings or distort real results.Fig.5 shows the outliers of the BTC features.

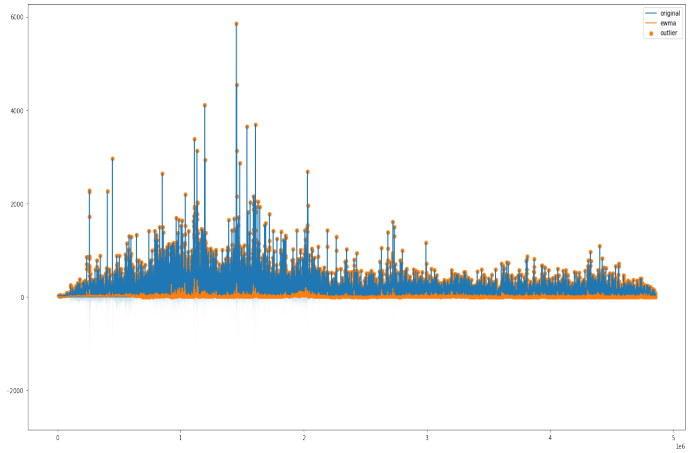

Fig. 5. Outliers of the BTC

Now that we have identified the outlier, we can use that value to remove the outlier.Fig.6 shows the Bitcoin exchange means by days, months, quarters, and years.

Finally, we have made a predictive model for BTC prediction.

We used the Arima model for our dataset. An ARIMA model is a class of statistical models for analyzing and forecasting time series data.ARIMA is an acronym that stands for Auto-Regressive Integrated Moving Average. Fig 7 shows the

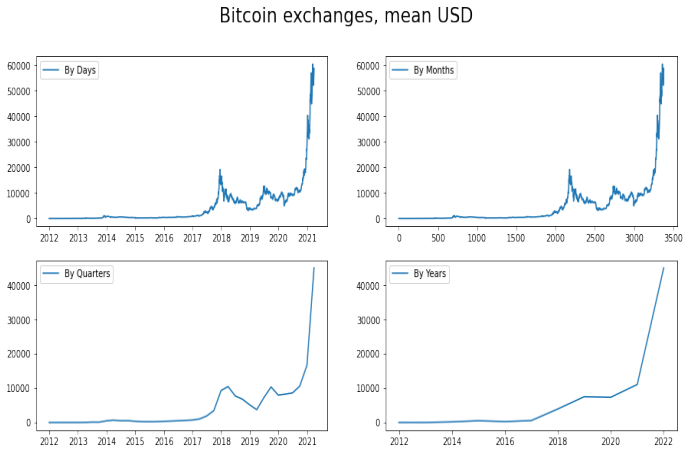

Fig. 6. Bitcoin Exchange Mean

Statespace Model Results

Dep. Variable:	Volume(BTC)	No. Observations:	3379
Model:	SARIMAX(2, 1, 3)x(0, 2, 2, 12)	Log Likelihood	-10313.644
Date:	Wed, 01 Sep 2021	AIC	20643.289
Time:	20:41:28	BIC	20692.232
Sample:	12-31-2011 - 03-31-2021	HQIC	20660.794
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	-0.5187	0.117	-4.432	0.000	-0.748	-0.289
ar.L2	0.4003	0.065	6.162	0.000	0.273	0.528
ma.L1	-0.0356	0.117	-0.304	0.761	-0.265	0.194
ma.L2	-0.8715	0.123	-7.096	0.000	-1.112	-0.631
ma.L3	0.0409	0.031	1.321	0.186	-0.020	0.101
ma.S.L12	-1.9957	0.039	-51.146	0.000	-2.072	-1.919
ma.S.L24	0.9958	0.039	25.644	0.000	0.920	1.072
sigma2	25.9579	1.041	24.924	0.000	23.917	27.999

Ljung-Box (Q):	152.71	Jarque-Bera (JB):	60401.84
Prob(Q):	0.00	Prob(JB):	0.00
Heteroskedasticity (H):	0.19	Skew:	2.70
Prob(H) (two-sided):	0.00	Kurtosis:	23.08

Fig. 7. ARIMA model summary

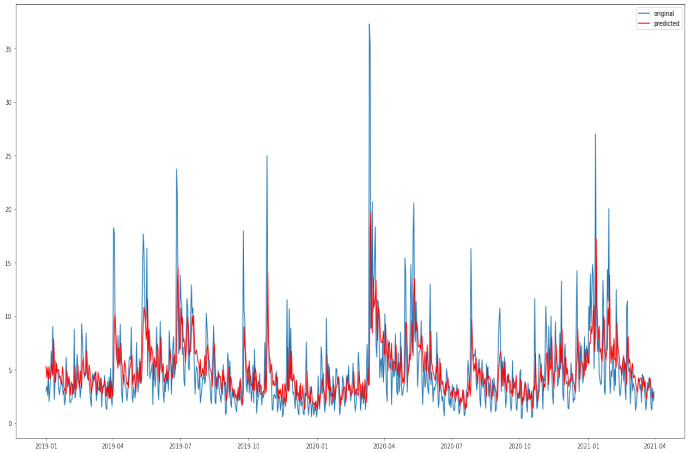

Fig. 8. Real data vs Prediction result

state-space model results of ARIMA. Finally, Fig.8 shows the graphical representation of actual data and prediction results. The forecasting model cannot reproduce the price spikes in 2019-04, 2019-07, 2019-10, 2020-04, and 2021-01, but it can reproduce the upward and downward price trends.

REFERENCES

[1] Size of the Bitcoin blockchain from January 2009 to July 12, 2021 Available: https://www.statista.com/statistics/647523/worldwide-bitcoin-blockchain-size/

[2] Y. Peng, P. H. M. Albuquerque, J. M. C. de Sá , A. J. A. Padula and M. R. Montenegro, "The best of two worlds: Forecasting high frequency volatility for cryptocurrencies and traditional currencies with support vector regression," Expert Systems with Applications, vol. 97, no. 3, pp. 177–192, 2018.

[3] A. Urquhart, "Price clustering in Bitcoin," Economics Letters, vol. 159, no. 4, pp. 145–148, 2017.

[4] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," Technical Report, 2019. [Online]. Available: https://bitcoin.org/bitcoin.pdf.

[5] S. McNally, J. Roche and S. Caton, "Predicting the price of Bitcoin using machine learning," in Euromicro Int. Conf. on Parallel, Distributed and Network-based Processing, Cambridge, UK, pp. 339– 343, 2018.

[6] R. Albariqi and E. Winarko, "Prediction of Bitcoin price change using neural networks," in Int. Conf. on Smart Technology and Applications, Surabaya, Indonesia, pp. 1–4, 2020.

[7] M. Van Alstyne, "Why Bitcoin has value," Communications of the ACM, vol. 57, no. 5, pp. 30–32, 2014.

[8] T. Shintate and L. Pichl, "Trend prediction classification for high frequency Bitcoin time series with deep learning," Journal of Risk and Financial Management, vol. 12, no. 1, pp. 17, 2019.

[9] Lipo Wang and Edwin Sin School of Physical and Mathematical Sciences Nanyang Technological University Singapore presented, "Bitcoin Price Prediction Using Ensembles of Neural Networks" in 2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD 2017)
Available: https://ieeexplore.ieee.org/document/8393351

[10] Lekkala Sreekanth Reddy, Dr.P. Sriramya, "A Research On Bitcoin Price Prediction Using Machine Learning Algorithms", in International Journal Of Scientific Technology Research Volume 9, Issue 04, April 2020 Issn22778616
Available: https://ieeexplore.ieee.org/abstract/document/8323676

[11] Sean McNally; Jason Roche; Simon Caton, "Predicting the Price of Bitcoin Using Machine Learning" in 2018 "26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP)"
Available: https://ieeexplore.ieee.org/abstract/document/8374483

[12] Zheshi Chen, Chunhong Li, Wenjun Sun, "Bitcoin price prediction using machine learning: An approach to sample dimension engineering" published in Journal of Computational and Applied Mathematics 365 (2020) 112395 (sciencedirect.com)

[13] Mrs Vaidehi, Alivia Pandit, Bhaskar Jindal, Minu Kumari and RupaliSingh, "Bitcoin Price Prediction Using Machine Learning" published on "International Journal of Engineering Technologies and Management" Research ISSN (Online): 2454-1907 May 8 2021

[14] Vigneshkumar R, Sathiskumar E, Ranjith SB, Sriram M, Dr. Ramya J, "Future Price of Bit Coin Prediction Using Machine Learning Model" in "International Journal of All Research Education and Scientific Methods" (IJARESM), ISSN: 2455-6211 Volume 9, Issue 2, February -2021

[15] F. Andrade de Oliveira, L. Enrique ZÃ¡rate and M. de Azevedo Reis; C. Neri Nobre, "The use of artificial neural networks in the analysis and prediction of stock prices," in IEEE International Conference on Systems, Man, and Cybernetics, 2011, pp. 2151-2155.

[16] M. Daniela and A. BUTOI, "Data mining on Romanian stock market using neural networks for price prediction". informatica Economica, 17,2013.

[17] Isaac Madan, Shaurya Saluja, Aojia Zhao "Automated Bitcoin Trading via Machine Learning Algorithms" (stanford.edu)