Journal of Engineering Science and Technology Review 17 (4) (2024) 136-144

Review Article

JOURNAL OF
Engineering Science and Technology Review

www.jestr.org

Using Historical Values and Social Media Sentiments to Predict Bitcoins and Altcoins Prices with Time Series Models - A Comprehensive Survey

Anamika Gupta, Shikha Gupta*, Smaran Das, Ajmera Prakash, Kartik Garg and Shreyan Sarkar

Department of Computer Science (SSCBS) University of Delhi, New Delhi, India

Received 22 January 2024; Accepted 11 July 2024

Abstract

The purpose of this study is to create an overview of the current industry practices concerning time series analysis specifically focused on cryptocurrencies. This study attempts to initially introduce the classical time series models slowly diving deeper to survey the field of altcoin price prediction to shed some light on currently available literature. Some of the models discussed below include ARIMA and its variations, LTSM, and other sentiment analysis procedures majorly targeting various alt- coins but as the first cryptocurrency, Bitcoin is observed to be the sole focus of the many studies discussed here. The study largely focuses on Time series data with social media sentiment analysis, the various factors affecting cryptocurrency prices, and finally discusses the prominent findings and results in the industry. This study is conducted with a focus to increase awareness and knowledge of the mentioned topics among fellow researchers and working professionals.

Keywords: Time Series, Cryptocurrencies, Altcoins, Sentiment Analysis, Machine learning, LTSM, ARIMA, Neural networks, Deep Learning

1. Introduction

Analysis of sequential and time-bound datasets to find hidden trends and repetitive behavior has been a major field of study for some time now. With the increase in financial awareness and an ample amount of resources within grasp, currently, a deep computer-intensive analysis of financial markets to predict future trends is heavily practiced in the industry. The stock market and other investment options have evolved with time, and so have investment techniques. Cryptocurrency is one such recent form of investment option and all existent financial techniques available for stock markets can be simply applied to it. Cryptocurrency can be sub-categorized into 2 segments: Bitcoin and Altcoins (Alternate Coins). Bitcoin was the first-ever digital cryptocurrency hence of- ten used as synonymous to all of cryptocurrency though such is not the case and other coins and tokens are often overshadowed by its popularity. Any coin or token other than Bitcoin may be called an Alt-coin, some major names being Ethereum, Dogecoin, and Ripple. Although some literature also places Ethereum out of the Alt coin category due to its own popularity and market capitalization, we will consider it as an altcoin for the sake of this study.

A plethora of the above-mentioned techniques for market prediction apply Machine Learning, Neural Networks, and concepts of Time-series analysis to tackle this problem of unpredictable and seemingly unintuitive price fluctuations of cryptocurrencies. As discussed earlier many studies have explored the relationship between Bitcoin and its predictors, however, other cryptocurrencies have not been studied upon as much. Thereupon this paper attempts to survey most of the currently available practices in the literature concerning time-series analysis of altcoins price prediction and with a general focus on studies concerning bitcoin.

2. Time Series Analysis

2.1 What is Time Series Analysis

A time series is a set of records that belong to a certain timeframe and follow a certain chronological order. The main feature that sets it apart from other datasets is its non-randomness in terms of the data points collected. Consequently, the process of finding the behavioral patterns of data that change over time or the process of making informed predictions based on previous patterns is Time Series Analysis. Time series analysis is majorly performed to extract these 3 features namely trend, seasonality, and heteroskedasticity based on which pre- dictions are made.

2.2 Various Methods in Time Series Analysis

Classically there are 11 Variations of Time Series forecasting methods [1] but different studies use a number of different approaches to tackle this time series problem. The said methods with other prominent alternatives are presented in Figure1.

2.3 Role of Social Media and Sentiment Analysis

In the current world social media is everywhere and it affects most things very greatly, be it in the form of communication or being used as a highly influential tool to guide people to form their opinion about something be it good or bad. The term Sentiment Analysis can be simply understood as the process of identifying the sentiments of the audience about a particular commodity. This process of analyzing social media sentiment and extracting the emotion out to be able to quantify it helps greatly to make estimations about the most probable future.

*E-mail address: shikhagupta@sscbsdu.ac.in
ISSN: 1791-2377
doi:10.25103/jestr.174.16

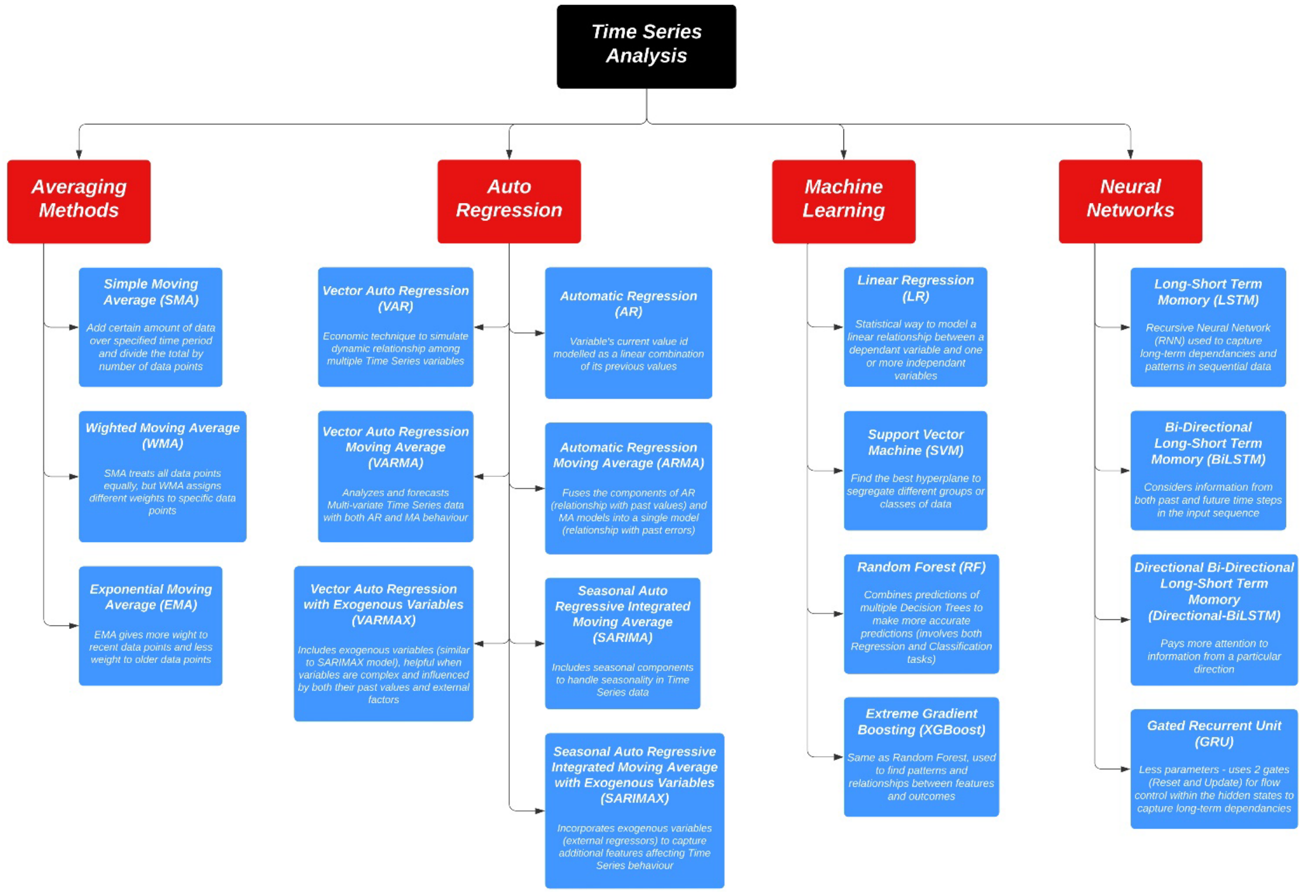

Fig. 1. Various methods

How is social media sentiment extracted

Like time series analysis, sentiment analysis is a field of study on its own and a lot of development has happened within a very short span. The process of sentiment analysis is mainly done by language processing models that always vary in terms of their complexity and are often tailor-matched for the problem at hand. One such study [2] attempted to sample a smaller time period containing 15000 tweets and found that out of the eight emotions categories, i.e. anger (2%), anticipation (18%), disgust (1%), fear (3%), joy (15%), sadness (3%), surprise (7%), trust (15%) and the two sentiments, i.e. negative (4%) and positive (33%), positive sentiments proved to be the largest category in the sample implying a positive movement in the market for the given sample. Reference [3] takes a step further by including news sentiments to the mix targeting Bitcoin, Ethereum, and Litecoin as the focus of their study. Figure 2 depicts the general procedure for sentiment analysis. To avoid deviating the survey's main focus on current time-series practices only some intensively used sentiment analysis methods are mentioned below:

1. Lexicon-based approach: All words are assigned a number to signify a positive or negative emotion based on which the sentences are calculated to be positive or negative.
2. Machine Learning based approach: Machine learning models are trained on large natural language datasets to be able to develop classifiers robust enough to bifurcate the inputs into positive or negative.
3. Hybrid: A combination of the above is also often designed to cater to the needs of respective studies to achieve promising discoveries.

Fig. 2. Sentiment Analysis

Different studies use different approaches to tackle the same problem and one such noteworthy study [4] advocates for the use of BERT language models compared to conventional Vader models as Vader fails to consider the context to be better able to extract sentiment from social media platforms. Study [3] makes use of textblob to boast of an approximately 80% and 75% success rate in successfully classifying positive tweets and negative tweets respectively. They also mention that Sarcasm being very difficult to detect may be partially responsible for the 25% false positive segment of their result. Another study [3] worth mentioning attempted determine the relationship between investors' sentiment and the volatility of cryptocurrency prices, their study forecasted the cryptocurrency prices using the Long-Term-Short-Memory (LSTM) deep learning algorithm and found very promising results during their classification of sentiments via Support Vector Machine (SVM) and Naive Bayes (NB). The multi-model study [4] for sentiment prediction compares logistic regression, linear support vector machines, and Naive Bayes. For Bitcoin, they found that logistic regression performs the best: it was able predicted 43.9% of price increases and 61.9% of price decreases correctly.

3. Prominent Findings and Results in Literature

The Stock market due to its importance and long history has attracted many studies [5] and a noteworthy observation made during the survey was that similar practices are utilized for cryptocurrencies. These studies have often served as the foundation for further research. Although Cryptocurrencies are comparatively a newer commodity, they have their fair share of market value to promote various studies to target cryptocurrency market fluctuations as the primary focus of their research. In recent years, academicians have plotted parallels among markets have managed to achieve substantial feats. Reference [6] conducted a systematic review on the relationship between cryptocurrency and the stock market, utilizing bibliometric and content analysis of 151 articles from 2008 to November 2021 showing asymmetric herd behavior and risk spillover between cryptocurrencies and stock markets of emerging economies. This section of the survey will discuss various studies concerning the said studies and shed some light on the current state of the field. Tables 1 and 2 attempt to summarize and depict the prominent features of the discussed studies in brief.

Table 1. Past researches predicting bitcoin using historical price only

Article	Dependent variable	Input set	Models	Sample period	Main finding
Vidyulatha, G.[7]	Bitcoin prices	BTC OHLC price	ARIMA and Linear Regression	July 2015 to June 2020	ARIMA outperforms LR
Changqing Luo [8]	Bitcoin prices	BTC OHLC price	Ensemble Model (VMD-LSTM- ELM)	2020/04/22 to 2020/07/20	Even when volatility is considerable, the ensemble models' prediction performance remains largely acceptable.
Muhammad J Amjad [9]	Bitcoin prices	BTC OHLC price	Arima, RF, EC, LDA and LR	1/12/14 to 13/3/15	On all metrics, classification algorithms perform better than EC and ARIMA.
Mahir Iqbal [10]	Bitcoin prices	BTC OHLC price	ARIMAX, FBPROP and XGBOOST	Jan 2012 to dec 2017	ARIMAX outperforms both FBPROP and XGBOOST
Ibrahim, A. [11]	Bitcoin prices	BTC OHLC price	vector autoregression (VAR) and Bayesian vector autoregression (BVAR)	04-01-2009 to 01/08/2020	Compared to actual values, BVAR model provided a more accurate prediction of the price of bitcoin.
Latif, N., Selvam, [12]	Bitcoin prices	BTC OHLC price (10 minutes fre- quency)	ARIMA and LSTM	12/21/2020 to 12/21/2021	The direction and the value within the determined time period were both accurately predicted by the LSTM model.
Prashant, S. [13]	Bitcoin prices	BTC OHLC price (1-minute frequency)	RNN and LSTM	Jan 2012 to March 2021	LSTM model outperforms all the models
Junwei Chen [14]	Bitcoin prices	BTC OHLC price	Random Forest Regression and LSTM	31/03/15 to 01/04/22	MAPE for RF and LSTM were 3.29% and 4.68% respectively.

Table 2. Past researches predicting bitcoin by incorporating Sentiments as feature

Article	Dependent variable	Input set	Models	Sample period	Main finding
Haritha, G.B [4]	Bitcoin prices	BTC OHLC price and Tweets	FinBERT	5 July 2021 to 5 October 2022	BERT is better than VADER as it also considers context while calculating sentiments
Krzysztof Wołk [15]	Bitcoin prices	BTC OHLC price, Google trends and Tweet sentiments (10- minutes frequency)	Hybrid model	Jan 2018 to Jan 2019	Discovered that Google Trends data and negative sentiments were the most effective predictors. Negative sentiments and carries a larger weight than positive ones.
Jacques Vella Cri-tien1 [16]	Bitcoin prices	Bitcoin OHLC prices and Tweet sentiments	Voting classifier Model (A hybrid model made from Direction-BiLSTM	30/08/2018 to 23/11/2019	Accuracy - 77% Voting classier provides higher accuracy levels by first identifying the direction of the price change and then predicting the

			and Magnitude-CNN)		actual bin.
Sai Prasanna Gontyala [17]	Bitcoin prices	Bitcoin OHLC prices and Tweet sentiments (1-hour frequency)	LSTM	11/05/,2018 to 29/05/2018	The LSTM model built using Adam optimizer outperformed Rmsprop.
Hae Sun Jung [18]	Bitcoin prices Trends	Bitcoin OHLC prices and Tweet sentiments	XGBoost	01/08/17 to 28/02/22	Accuracy of 90.57% with AUC value of 97.48%
Ifigeneia Georgoula [19]	Bitcoin prices	Bitcoin OHLC prices and Tweet sentiments	Vector Error-Correction Model and SVM	27/10/14 to 12/01/15	Wikipedia views have a positive correlation with Bitcoin price. SVM yielded an accuracy of 89.6%.

3.1 Factors Affecting Cryptocurrency

The first and foremost step in any financial study is figuring out the commodity's indicators to draw a cause-and-effect relationship to better understand the behavioral patterns of the time-series data. As this serves as the foundation for the prediction algorithms, an adequate amount of effort is put into the same. While different studies focus on different attributes, a superset of these attributes can be formed and the following subsection will attempt to bring them to light.

Where an attempt to categorize bitcoin adoption and investments on a country-based division [20], the most obvious choice would be the macro-indicators (GDP, Inflation, etc) these aren't very helpful for a higher frequency data and specific region-based studies. In such scenarios, it has been observed that the majority of studies focus on a specific set of attributes which can be fragmented into internal (supply and demand) and external (macro-financial with attractiveness levels) factors both in the case of bitcoin [21] and other altcoins [22]. Another study [23] also tries to explore this dependence between multiple cryptocurrencies and their respective price-influencing factors with the help of Bayesian networks. Reference [13] also explores this co-dependence of the mentioned factors and Bitcoin prices discovering that while some factors like Twitter sentiment, Wikipedia searches, and Hashrates have a positive effect, it is negatively associated with the Standard and Poor's 500 stock market index (which indicates the general state of the global economy). Likewise, study [24] concluded their experiments by stating that sentiment does not have an im- pact on crypto prices in the short-term perspective, but there is a long-term relationship between sentiment and crypto prices. Furthermore, a study by [25] investigated the impact of online user feedback on cryptocurrencies price volatility and trading volume. The study found that Bitcoin (BTC) showed a remarkable correlation with the amount of positive comments on social media. On the other hand, the increase in positive comments about bitcoin on social media platforms was related to a price movement and transactional activity in the cryptocurrency. Additionally, reference [13] performed a time series analysis itself to find the long-run and short-run influencers of Bitcoin prices. While the later included factors like Wikipedia views, Hashrates and Sentiment Ratio, the former comprised of the number of Bitcoins available and the S&P 500 index.

Although internal factors have a non-deniable contribution to cryptocurrency price fluctuations, there is a surplus of studies to prove a strong dependence on the cryptocurrency price with its external factors. [26] is a study that shows that although each sub-factor has some degree of relation with price, it is the combination of all these subfactors that provide the best results. While we know that a combination of these subfactors is a better choice, many studies only focus on social media sentiment to predict these market fluctuations and have managed to achieve commendable results. Some such studies [27, 28] have solely focused on Twitter sentiments to judge the price fluctuation and have a very strong relation between the two. One astonishing discovery that came to light during their study was that around 1-14% of the daily tweets that were collected were created by Twitter bots created either to manipulate or promote the overall sentiment of the respective cryptocurrency [28]. There also exists a study [29] that claims that what drives the attention of online investors is mostly the evolution of prices and not the evolution of technology. They also point out that often emojis that play a crucial rule for sentiment calculation are often removed for simpler calculations.

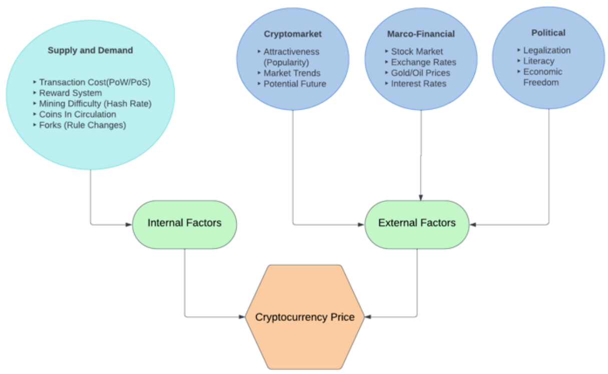

Fig. 3. Factors affecting Cryptocurrency

3.2 Accuracy of Price Fluctuations

Although the statistics of the respective evaluation metric score vary abruptly across different studies due to their respective approach to the problem and the features of the datasets the studies base their work on, these help us better visualize and grasp the essence and magnitude of their results. Broadly we can categorize these studies into 2 categories based on their dataset frequencies as a metric:

1. Low/Mid Frequency: Studies that focus on a dataset with a monthly frequency, i.e. there is only one data point representing the said timeframe or with a focus on daily frequency.
2. High Frequency: Studies with an hourly or minute-wise dataset frequency. One thing to mention before we start discussing the following studies is that it would seem that the studies with a higher frequency will often have higher statistics compared to their lower-frequency counterparts though with their exceptions, one simple reason for this may be the availability of a greater

number of data points for the algorithms to work on and this does not mean that the latter may be better than the former. Each study serves to better our understanding of the topic.

3.2.1 Low/Mid Frequency dataset-based studies

This subsection will discuss studies that have focused on a lower frequency dataset that compared to their higher frequency counterparts have lower statis- tics in terms of their accuracy scores and error scores.

One observation that was made by [29] claims that the ARIMA model is superior to the Linear Regression machine learning model in terms of predicting financial fluctuation for Bitcoin in the short run. The study focused on the time frame of 5 and a half years and managed to get on-par results with its machine-learning counterparts. Likewise, the study [10] also observed that ARIMAX is the best algorithm to forecast the change in the bitcoin price in the market with RMSE of 322.4 compared to FBProp and XGBoost algorithms with RMSE scores of 229.5 and 369 respectively. In addition, reference [30] compared ARIMA with classical machine learning models. The study [30] experimented for both cases, univariate and multivariate models in which the ARIMA, ANN (Artificial Neural Network), Kriging, and Bayesian models are used as univariate while ANN, SVM, RF, and Bayesian models were proposed for multivariate case. It was observed ARIMA and Bayesian provides better results compared to other univariate models since they have smaller RMSE and MAPE values compared to other models. However, the SVM outperforms all the univariate and multivariate models and is selected as the best model where its performance measures of RMSE and MAPE are much smaller than the values of all other models. It was also indicated by Reference [31] that a time-series approach leads to better results than a conventional one-day timeframe for bit- coin price prediction. Consequently, Reference [32] also observed that for the short-term forecast, the error of the BART (Bayesian Additive Regression Tree) algorithm was half the size of the error of the conventional ARIMA model on average. It was also observed that it was almost 15-20% lower than the error of the ARFIMA model for slowly changing periods.

Table 3. Past researches predicting altcoins using historical price only

Article	Dependent variable	Input set	Models	Sample period	Main finding
Adedokun, A.,[34]	Cryptocurrencies prices (BTC, ETH, DASH, DOGE, etc.)	OHLC price of cryptocurrencies	Recursive Residual Test and VEC Granger/Block Exogeneity Test	3 years (2015, 2017 and 2018)	Some currencies show strong bidirectional causality
Srđan Jelinek [35]	Cryptocurrencies prices (BTC, ETH, LC)	OHLC price of cryptocurrencies	Fourier Transform and F-Transform with FIS	7/8/2015 to 24/12/2018	Fourier Transform and F-Transform had accuracy of 66% and 61% respectively.
Jacques Phillipe Fleischer [36]	Cryptocurrencies prices (BTC, ETH, DOGE, EOS)	OHLC price of cryptocurrencies	LSTM	9/11/2017 to 30/06/2022	LSTM outperforms ARIMA by great margin for all the coins.
David Meijer [37]	Ethereum and Ripple prices	OHLC price of cryptocurrencies	LSTM	01/01/2017 to 30/04/2020	LSTM is not able to achieve reliable results as the con- structed models show signs of over- fitting
Sridhar, S [38]	Dogecoin prices	DOGE OHLC price	multi-head attention-based transformer encoder-decoder model	05/07/2019 to 28/04/2021	accuracy of 98.46% and R-squared value of 0.8616
Kurniawan, K [33]	Cryptocurrencies prices (BTC, XRP, DOGE)	OHLC price of cryptocurrencies	ARIMA, GARCH and Holt's Winter	07/08/15 to 30/06/22	Holt – Wintertime series model out- performs ARIMA and GARCH
Derbentsev, V. [32]	Cryptocurrencies prices (BTC, XRP and ETH)	OHLC price of cryptocurrencies	ARIMA, ARFIMA and BART	01/01/2017 to 01/03/2019	for the short-term forecast, BART is much better choice than ARIMA
Persson, E [39]	Cryptocurrencies prices (BTC, SOL and ETH)	OHLC price of cryptocurrencies	ARIMA, GARCH, LSTM, Transformer, Prophet and Naive walk	2022-01-01 to 2022-05-20	LSTM outperformed the other models in times of higher volatility
Kwon, D-H [40]	Cryptocurrencies prices (BTC, XRP, ETC, ETH, etc.)	OHLC price of cryptocurrencies	LSTM and GB	09/06/17 to 08/05/2018	LSTM model is always superior to the GB model in all metrics.
Bouteska, A. [41]	Cryptocurrencies prices (BTC, XRP, LTC, ETH)	OHLC price of cryptocurrencies	Arima, MLP, LSTM, AdaBoost, Light GBM, Simple RNN, GRU	01/04/16 to 31/08/23	Trading strategies based on deep learning (for Ripple) or ensemble learning (for Bitcoin, Ethereum, and Litecoin) gives better results.

Another study [34] that attempted to analyze the differences in trends with the help of methods similar to VAR models found that while there was a synchronization of price boom in cryptocurrency in 2017 the investors were more aware of the individual cryptocurrency projects in the upcoming year resulting in de-synchronized market fluctuation of cryptocurrency in 2018 respectively. A study [42] utilizing Random Forest Regression and LSTM with the implementation of Lags (a concept stating that there exist a delay for the effect to actually materialize) in their study targeting Bitcoin observed that whether it was random forest regression or the LSTM algorithm, as the number of past periods of the substituted explanatory variables increases, the prediction accuracy of the model decreases. Likewise reference [16] used voting classifier model (Direction-BiLSTM and Magnitude-CNN models merged together) to attain an accuracy of 77% reaffirming the above claim of decreasing accuracy in case of more lags being added after a 3 day lag. A noteworthy Reference [43] that makes use of available technical indicators with the help of machine learning based neural net- work to predict bitcoin prices had managed to attain a accuracy of 94.89% under all circumstances of technical trade indication increasing trader confidence by graphs depicting a real BTC value 5 to 10 times in 300-days of respective fiscal year. Another study [36] that relied on LTSM as their model focused on EOS- USD as their cryptocurrency. The study made use of EPOCHS and constantly fed the data points to algorithm in the different sized fragments.

Table 4. Past researches predicting altcoins by incorporating Sentiments as feature

Article	Dependent variable	Input set	Models	Sample period	Main finding
Jethin Abraham [44]	Cryptocurrencies prices of 181 currencies	Cryptocurrencies OHLC prices, Google trends and Tweet sentiments.	Pearson Correlation Coefficient	04/03/18 to 03/06/18	The volume of tweets and Google Trends were both highly correlated to pricing.
Raj Parekh [45]	Cryptocurrencies prices (Dash and Bitcoin cash)	Dash and Bitcoin-cash OHLC prices and Tweet sentiments.	DL-GuesS (hybrid GRU and LSTM-based model)	03/03/21 to 01/04/21	DL-GuesS out- performs the traditional systems with MAPE being 4.7928 for Dash and 4.4089 for Bitcoincash
Franco Valencia [46]	Cryptocurrencies prices (BTC, ETH, XRP and LTC)	Tweet sentiments (1-hour frequency)	SVM, RF and MLP	16/02/18 to 21/04/18	MLP performs better than random forest and SVM for single feature vector classification designs. Accuracy for Bitcoin was 0.72 with 0.74 precision.
Tianyu Ray Li [47]	Tweets Sentiment Score	Tweets (1-hour frequency)	Textblob	26/01/19 to 19/02/19	Positive tweets are identified with a success rate of over 80% and negative tweets with a success rate of 75%.
Cathy Yi-Hsuan Chen [29]	Cryptocurrencies prices (EW and CRIX)	Cryptocurrencies OHLC prices, and Tweet sentiments.	Autoregression	01/08/14 to 27/12/18	The precise sentiment of messages posted on social media is also extremely well captured by emojis.
M. Kabir Hassan [2]	Tweets Sentiment Score	Tweets	NRC emotion lexicon	-	Trust (15%) and positiveness (33%) found to be biggest sentiments.
S. Oikonomopoulos [3]	Cryptocurrencies prices (BTC, ETH and LC)	Cryptocurrencies OHLC prices, and Tweet sentiments.	Logistic Regression and Bernoulli Naive Bayes	24/09/17 to 30/11/17	Naive Bayes model struggles to reliably anticipate daily price fluctuations that deviate from the overall trend but works rather well for identifying broad trends in coin prices.
O. Kraaijeveld [48]	Cryptocurrencies prices (BTC, ETH and XRP etc)	Cryptocurrencies OHLC prices, and Tweet sentiments. (1-hour frequency)	Granger Causality testing	04/06/18 to 04/08/18	A heuristic method is created to determine that Twitter bot accounts posted at least 1 to 14 per- cent of the Tweets that were acquired.
Frank van Engelen [24]	Cryptocurrencies prices (BTC, ETH and ADA)	Cryptocurrencies OHLC prices, and Tweet sentiments.	Pearson Correlation and Granger Causality	01/01/19 to 31/12/22	Bitcoin can be used to predict future sentiment in the market. sentiment does not have an impact on crypto prices in the short-term perspective, but there is a long-term relationship between sentiment and crypto prices

Additionally, [35] attempted a Fourier-based approach wherein they worked with Fourier Transform with FIS and F-Transform with FIS. They concluded their study with a somewhat expected result that the F-transform is slightly better than the Fourier transform, which is to be expected since conventionally the Fourier transform is used for predicting derivatives prices. Reference [36] moves a step further to suggest LTSM (Long short-term memory) methods over the conventional ARIMA-based studies due to better RMSE values resembling the results of many studies that will be discussed in the next subsection while we should also note that some studies like [9] also exist that indicate that we should use classification algorithms in settings where the underlying time series is stationary and mixing. Furthermore, Reference [44] applied Pearson Correlation Coefficient as their experiment model and reported that the Sentiment of tweets was not a reliable indicator when cryptocurrency prices were dropping. Although both Google Trends and tweet volume were observed to be highly correlated with price also maintaining itself during periods of increasing and decreasing prices suggesting that the relationship is robust to periods of high variance and non-linearity.

A unique approach of combining different models to create a hybrid model to cater to the needs of the researchers is also a trend that was observed during our survey. Reference [45] uses DL-GuesS (hybrid GRU and LSTM-based model) claiming a better performance compared to the traditional approach with MAPE acquired for Dash and Bitcoincash as 4.7928 and 4.4089 respectively. Another prominent model used by [18] was XGBoost which outperforms the conventional machine learning models like Logistic Regression, SVM and Random Forest. While the study only predicted the trends and not the magnitude of these trends, they managed to achieve an accuracy of 90.57% and an AUC value of 97.48%.

Consequentially, reference [9] attempted to identify an effective ML algorithm for long-term Bitcoin price predictions, using technical indicators as model inputs on the historical price data from May 2017 to May 2023. As per their study LSTM was the found to be the most accurate ML algorithm among those tested, with technical indicators EMA and SMA having a significant impact on model performance. Reference [41] provide a comprehensive comparative analysis of ensemble learning and deep learning forecasting models, on various cryptocurrencies (Bitcoin, Ethereum, Ripple, and Litecoin). The results of this study reveal that gated recurrent unit, simple recurrent neural network, and LightGBM methods outperform other machine learning methods, as well as the naive buy-and-hold and random walk strategies. Reference [49] examined hybrid LSTM machine learning models that can be used for the prediction of cryptocurrency prices especially Bitcoin. The numerical outcomes attested the fruitfulness of hybrid LSTM model with impressive results like 150.96 RMSE reduction for ETH and reduced normalized one-RMSE of only 0.05. LSTM and CNN-LSTM have shown promising results in capturing now-linear relationships, long-term dependencies, and complex patterns. Moreover, consistent trends were observed in multiple cryptocurrencies during the same period, which is required to be further explored [50][51].

3.2.2 High-Frequency dataset-based studies:

As discussed earlier, studies with a frequency greater than a daily timeframe reported higher statistics compared to their other counterparts. One such study is [12] which managed to achieve accuracy figures of 98.21% and 99.73% with ARIMA and RNN-LTSM respectively. The study focuses on a 10-minute interval of bitcoin price from 21st Dec 2020 to 21st Dec 2021. Another crucial feature of the study that should be mentioned is that only 5% of data was taken as the test set while the other 95% served to train the above-mentioned models. An observation recorded during their study implies that while ARIMA could only track the trend of Bitcoin prices, the LSTM model was able to predict both the direction and the value during the specified time period. One study that uses a different model than the ones mentioned above to achieve great results is [38]. The Multihead attention-based transformer encoder-decoder model simply put is a combination of multiple attention modules where each module repeats its computations multiple times in parallel. This is quite a different model from the conventional time-series models often used by studies. The study specifically focussed on the period of 05 July 2019 to 28 April 2021 to predict dogecoin price fluctuation to achieve an accuracy of 98.46% and an R-squared value of 0.8616 which in itself is a humongous achievement and overheads many studies discussed till now by a comfortable gap.

Other noteworthy studies that need mentioning are [11] and [33]. While [11] used vector autoregression (VAR) and Bayesian vector autoregression (BVAR) and observed that where the VAR model proved to be superior in terms of predicting a great pattern for the fluctuations BVAR was much more accurate to predict the bitcoin values.

On the other hand, Reference [26] applied and compared 3 models namely: ARIMA, GARCH (Generalized AutoRegressive Conditional Heteroskedasticity), and Holt's Winter, they observed that Holt - Wintertime series model is a better Bitcoin, Ripple, and Litecoin forecasting model compared to ARIMA and GARCH. Reference [39] ups it another notch by comparing five different models: ARIMA, GARCH, LSTM, Transformer, Prophet, and Naive Walk. Their study targeted Solana (SOL) Bitcoin (BTC) Ethereum (ETH) for the period of 1st Jan 2022 to 20th May 2022. The study concluded that LSTM and ARIMA-GARCH performed best in a scenario of low volatility, while the LSTM outperformed the other models in times of higher volatility in their experiments. Another study that advocates for LSTM [19] states that the LSTM model outperforms as compared to all the models with the minimum MSE score. The study focused on a minute-by-minute frequency of bitcoin prices for the time period of January 2012 to March 2021. The experiments were performed with an 80-20 split of training to test data. Furthermore, reference [40] found a similar trend that the LSTM model is always better than the GB model for all cryptocurrencies (BTC (Bitcoin), ETH (Ethereum), XRP(Ripple), BCH (Bitcoin Cash), LTC (Litecoin), DASH (Dash), and ETC (Ethereum Classic)), they performed a time series classification upon. It should also be mentioned that [32] promotes non-linear deep learning models over ARIMA which performed poorly in their study. Although a study [37] also states that in some scenarios, the LTSM model was prone to overfitting. Another multi- currency study [46] that targets Bitcoin, Ethereum, Ripple, and Litecoin found that in the case of single feature vector classification design, MLP outperforms other classification models like random forest and SVM. An interesting feature in study [52] was observed during our survey, it focused only on 135 influential Twitter accounts rather than the full sample and were able to achieve commendable results indicating a smarter and more efficient approach for future research.

Additionally, a different approach was put into use by [14]. The study con-siders the multiple factors that affect the cryptocurrency price and generates a new Ensemble Model (VMD-LSTM-ELM) to achieve a prediction accuracy of 95.12%. The model puts into consideration the multiscale attributes of cryptocurrency price and matches different machine learning models to overcome the same. It was also observed that even when the volatility was high, the prediction performance of the ensemble models was at a relatively satisfactory level. Reference [15] also uses a hybrid model comprising of least square linear regression and Bayesian Ridge Regression Model that achieved consistently good results even with blind testing data. They concluded their study by stating that the most powerful predictors were found to be Google Trends data together with general negative sentiment (including weighted sentiment). Additionally, observing that Negative news carries a larger weight, as shown by the correlation values during their data exploration phase

According to the literature check, most of the papers only read using literal data and don't operate with real-time data. The majority of them use literal day-to-day ending prices rather than current prices and don't deal with short time intervals similar to five minutes or fifteen minutes. Cryptocurrency price is largely unpredictable and volatile, making it difficult to manage and much more complex to study. A majority of the studies used univariate time series models, which don't take advantage of the other indicators and other influential features to ameliorate the complexity. Deep learning models are effective in predicting crypto prices but have limitations like complex model training and a long training time, which makes it grueling to train the model in real time. The forecasting accuracy varies greatly between models and cryptocurrencies, and there is no obvious trend that would allow us to determine which model is best or which coin is the most predictable throughout the validation or test periods. However, when compared to other similar studies, the forecasting accuracy of the individual models generally seems low. This is not unexpected given that the top model in its class is based on maximizing the average of returns one step forward rather than on minimizing predicting error. Furthermore Reference [17] found out in their experiments that the LSTM model built using Adam optimizer outperformed Rmsprop.

4. Conclusion

This paper has discussed various time series models and different approaches utilized by different studies available in current literature to attempt to create a general overview of the current state of the respective domain. This study explored diverse time series models and various methodologies to predict cryptocurrency trends with high accuracy. It also covers various factors which affect the price of the cryptocurrencies, Internal factors which include transaction cost, amount of coin in circulation, and mining difficulties can have an effect on the potential growth and demand of the coin whereas some factors such as market trends, exchange rates, the popularity of the coin are some external factors which can have a huge effect on the price of the coin.

In low-frequency datasets, time series models like ARIMA and ARIMAX have preformed significantly well outperforming the linear Regression machine learning model in predicting short-term financial fluctuations for Bitcoin. Another model which is BART(Bayesian Additive Regression Tree) outperforms the ARIMA algorithm for short-term forecasts and the error was half the size of the error of the conventional ARIMA model on average. Furthermore, it was observed that some studies advocated LTSM (Long short-term memory) methods over the conventional ARIMA-based studies due to better RMSE values.

Consequently, for the case of high-frequency datasets, a variety of models have performed well, where models like ARIMA and RNN-LSTM have managed to predict the price of bitcoin with commendable accuracy of 98.21% and 99.73% respectively. The key observation was that ARIMA was only able to track the trend of the Bitcoin price whereas LTSM model was able to predict both the direction and the value during the specified time period. The VAR and BVAR models also performed significantly well, with BVAR being superior in predicting the actual values. It was also observed that Holt - the Wintertime series model is better than ARIMA and GARCH models in forecasting various cryptocurrencies. As a precautionary note, while this survey attempts to create a comprehensive overview of the said domain, it has its limitations and would like to inform the reader of the same and suggest further reading to achieve an even deeper understanding of the domain.

References

[1] A. S. Bharatpur, "A literature review on time series forecasting methods,." Jan. 2022. [Online]. Available: https://www.researchgate.net/publication/357786404_A_LITERATURE_REVIEW_ON_TIME_SERIES_FORECASTING_METHODS

[2] M. K. Hassan, F. A. Hudaefi, and R. E. Caraka, "Mining netizen's opinion on cryptocurrency: sentiment analysis of Twitter data," *Stud. Econom. Finan.*, vol. 39, no. 3, pp. 365–385, Apr. 2022, doi: 10.1108/SEF-06-2021-0237.

[3] S. Oikonomopoulos, K. Tzafilkou, D. Karapiperis, and V. Verykios, "Cryptocurrency Price Prediction using Social Media Sentiment Analysis," in *2022 13th International Conference on Information, Intelligence, Systems & Applications (IISA)*, Corfu, Greece: IEEE, Jul. 2022, pp. 1–8. doi: 10.1109/IISA56318.2022.9904351.

[4] G. B. Haritha and N. B. Sahana,, "Cryptocurrency Price Prediction using Twitter Sentiment Analysis," in *Natur. Lang. Process., Informat. Retriev. AI*, Academy and Industry Research Collaboration Center (AIRCC), Feb. 2023, pp. 13–22. doi: 10.5121/csit.2023.130302.

[5] H. N. Gudavalli and K. V. R. Kancherla, "Predicting Cryptocurrency Prices with Machine Learning Algorithms: A Comparative Analysis." , Department of Computer Science, p. 64, 2023.

[6] S. S. Jeris, A. S. M. N. Ur Rahman Chowdhury, Mst. T. Akter, S. Frances, and M. H. Roy, "Cryptocurrency and stock market: bibliometric and content analysis," *Heliyon*, vol. 8, no. 9, Art. no. e10514, Sep. 2022, doi: 10.1016/j.heliyon.2022.e10514.

[7] G. Vidyulatha, M. Mounika, and N. Arpitha, "Crypto currency prediction model using arima," *Turkish J. Comput. Mathem. Educ. (TURCOMAT)*, vol. 11, no. 3, pp. 1654–1660, Mar. 2020.

[8] C. Luo, L. Pan, B. Chen, and H. Xu, "Bitcoin Price Forecasting: An Integrated Approach Using Hybrid LSTM-ELM Models," *Mathemat. Probl. Engin.*, vol. 2022, pp. 1–17, Nov. 2022, doi: 10.1155/2022/2126518.

[9] M. J. Amjad and D. Shah, "Trading Bitcoin and Online Time Series Prediction," in *NIPS Time Series Workshop*, Dec. 2016. [Online]. Available: https://api.semanticscholar.org/CorpusID:14853974

[10] M. Iqbal, M. Iqbal, F. Jaskani, K. Iqbal, and A. Hassan, "Time-Series Prediction of Cryptocurrency Market using Machine Learning Techniques," *EAI Endors. Transact. Creat. Technol.*, vol. 8, no. 28, Art. no. 170286, Aug. 2021, doi: 10.4108/eai.7-7-2021.170286.

[11] A. Ibrahim, R. Kashef, M. Li, E. Valencia, and E. Huang, "Bitcoin Network Mechanics: Forecasting the BTC Closing Price Using Vector Auto-Regression Models Based on Endogenous and Exogenous Feature Variables," *J. Risk Financ. Manag.*, vol. 13, no. 9, Art. no. 189, Aug. 2020, doi: 10.3390/jrfm13090189.

[12] N. Latif et al., "Comparative performance of lstm and arima for the short-term prediction of bitcoin prices," *Australasian Account,, Busin. Fin. J.*, vol. 17, no. 1, pp. 256–276, Jan. 2023.

[13] S. Prashant, "Bitcoin Price Prediction Using Time-series Analysis and Sentiment Analysis on Twitter Data in Cloud Environment," Master's Thesis, Dublin, National College of Ireland, 2022. [Online]. Available: https://norma.ncirl.ie/5959/

[14] J. Chen, "Analysis of Bitcoin Price Prediction Using Machine Learning," *J. Risk Financ. Manag*, vol. 16, no. 1, Art. no 51, Jan. 2023, doi: 10.3390/jrfm16010051.

[15] K. Wołk, "Advanced social media sentiment analysis for short-term cryptocurrency price prediction," *Exp. Sys.*, vol. 37, no. 2, Art. no. e12493, Apr. 2020, doi: 10.1111/exsy.12493.

[16] J. V. Critien, A. Gatt, and J. Ellul, "Bitcoin price change and trend prediction through twitter sentiment and data volume," *Financ Innov*, vol. 8, no. 1, p. 45, Dec. 2022, doi: 10.1186/s40854-022-00352-7.

[17] S. P. Gontyala, "Prediction of Cryptocurrency Price based on Sentiment Analysis and Machine Learning Approach," National College of Ireland, Dublin, 2021. [Online]. Available:

https://norma.ncirl.ie/5163/1/saiprasannagontyala.pdf

[18] H. Sun Jung, S. Hong Lee, H. Lee, and J. Hyun Kim, "Predicting Bitcoin Trends Through Machine Learning Using Sentiment Analysis with Technical Indicators," *Comp. Sys. Sci. Engin.*, vol. 46, no. 2, pp. 2231–2246, Jan. 2023, doi: 10.32604/csse.2023.034466.

[19] I. Georgoula, D. Pournarakis, C. Bilanakos, D. N. Sotiropoulos, and G. M. Giaglis, "Using Time-Series and Sentiment Analysis to Detect the Determinants of Bitcoin Prices," *SSRN J.,* May 2015, doi: 10.2139/ssrn.2607167.

[20] F. Parino, M. G. Beiró, and L. Gauvin, "Analysis of the Bitcoin blockchain: socio-economic factors behind the adoption," *EPJ Data Sci.*, vol. 7, no. 1, Art. no. 38, Dec. 2018, doi: 10.1140/epjds/s13688-018-0170-8.

[21] O. Poyser, "Exploring the determinants of Bitcoin's price: an application of Bayesian Structural Time Series," 2017, *arXiv*. doi: 10.48550/ARXIV.1706.01437.

[22] Y. Sovbetov, "Factors Influencing Cryptocurrency Prices: Evidence from Bitcoin, Ethereum, Dash, Litcoin, and Monero," *J. Econom. Financ. Anal.*, vol. 2, no. 2, pp. 1–27, Feb. 2018.

[23] R. Amirzadeh, A. Nazari, D. Thiruvady, and M. S. Ee, "Modelling Determinants of Cryptocurrency Prices: A Bayesian Network Approach," 2023, *arXiv*. doi: 10.48550/ARXIV.2303.16148.

[24] L. Kulcsar and F. Engelen, "Twitter Sentiment Analysis on the Cryptocurrency Market," Jönköping University, Sweden, 2023. [Online]. Available: https://www.diva-portal.org/smash/get/diva2:1762598/FULLTEXT01.pdf

[25] Y. B. Kim *et al.*, "Predicting Fluctuations in Cryptocurrency Transactions Based on User Comments and Replies," *PLoS ONE*, vol. 11, no. 8, Art. no. e0161197, Aug. 2016, doi: 10.1371/journal.pone.0161197.

[26] M. Ortu, N. Uras, C. Conversano, S. Bartolucci, and G. Destefanis, "On technical trading and social media indicators for cryptocurrency price classification through deep learning," *Exp. Sys. Applic.*, vol. 198, Art. no. 116804, Jul. 2022, doi: 10.1016/j.eswa.2022.116804.

[27] T. Pano and R. Kashef, "A Complete VADER-Based Sentiment Analysis of Bitcoin (BTC) Tweets during the Era of COVID-19," *Big Data Cognit. Comput*, vol. 4, no. 4, Art. no. 33, Nov. 2020, doi: 10.3390/bdcc4040033.

[28] O. Kraaijeveld and J. De Smedt, "The predictive power of public Twitter sentiment for forecasting cryptocurrency prices," *J. Int. Finan. Mark., Instit. Money*, vol. 65, Art. no. 101188, Mar. 2020, doi: 10.1016/j.intfin.2020.101188.

[29] C. Y. Chen, R. Despres, L. Guo, and T. Renault, "What Makes Cryptocurrencies Special? Investor Sentiment and Return Predictability During the Bubble," *SSRN J.*, Jun. 2019, doi: 10.2139/ssrn.3398423.

[30] M. Khedmati, F. Seifi and M. J. Azizi, M. Mudassir, S. Bennbaia, D. Unal, and M. Hammoudeh, "Time-series forecasting of Bitcoin prices using high-dimensional features: a machine learning approach," *Neural Comput & Applic*, Jul. 2020, doi: 10.1007/s00521-020-05129-6.

[31] M. Mudassir, S. Bennbaia, D. Unal, and M. Hammoudeh, "Time-series forecasting of Bitcoin prices using high-dimensional features: a machine learning approach," *Neural Comput & Applic*, Jul. 2020, doi: 10.1007/s00521-020-05129-6.

[32] V. Derbentsev, N. Datsenko, O. Stepanenko, and V. Bezkorovainyi, "Forecasting cryptocurrency prices time series using machine learning approach," *SHS Web Conf.*, vol. 65, Art. no. 02001, Jan. 2019, doi: 10.1051/shsconf/20196502001.

[33] Khusrul Kurniawan and Sugiyono Madelan, "Forecasting Using Time Series Analysis Method in Crypto Currency Period 2015 – 2022," *Int. J. Innov. Sci. Res. Techn.*, vol. 7, no. 9, pp. 1454–1459, Oct. 2022, doi: 10.5281/ZENODO.7201582.

[34] A. Adedokun, "Bitcoin-Altcoin Price Synchronization Hypothesis: Evidence from Recent Data," *J. Fin. Econom.*, vol. 7, no. 4, pp. 137–147, Dec. 2019.

[35] S. Jelinek, A. Poledica, B. Petrović, and P. Milošević, "Forecasting Cryptocurrency Time Series Using Fuzzy Transform, Fourier Transform and Fuzzy Inference System," in *Proceed. 2019 Conf. Int. Fuzzy Sys. Assoc. European Soc. Fuzzy Logic Techn. (EUSFLAT 2019)*, Prague, Czech Republic: Atlantis Press, 2019. doi: 10.2991/eusflat-19.2019.88.

[36] J. P. Fleischer, G. Von Laszewski, C. Theran, and Y. J. Parra Bautista, "Time Series Analysis of Cryptocurrency Prices Using Long Short-Term Memory," *Algorithms*, vol. 15, no. 7, Art. no. 230, Jul. 2022, doi: 10.3390/a15070230.

[37] D. Meijer, "Predicting cryptocurrency price trends with long short-term memory," Tilburg University, The Netherlands, 2020. [Online]. Available: http://arno.uvt.nl/show.cgi?fid=156536

[38] S. Sridhar and S. Sanagavarapu, "Multi-Head Self-Attention Transformer for Dogecoin Price Prediction," in *2021 14th Int. Conf. Human Sys. Inter. (HSI)*, Gdańsk, Poland: IEEE, Jul. 2021, pp. 1–6. doi: 10.1109/HSI52170.2021.9538640.

[39] E. Persson, "Forecasting Efficiency in Cryptocurrency Markets : A machine learning case study," Master's Thesis, KTH, School of Electrical Engineering and Computer Science (EECS), 2022.

[40] D.-H. Kwon, J.-B. Kim, J.-S. Heo, C.-M. Kim, and Y.-H. Han, "Time Series Classification of Cryptocurrency Price Trend Based on a Recurrent LSTM Neural Network," *J. Inform. Process. Sys.*, vol. 15, no. 3, pp. 694–706, Jun. 2019, doi: 10.3745/JIPS.03.0120.

[41] A. Bouteska, M. Z. Abedin, P. Hajek, and K. Yuan, "Cryptocurrency price forecasting – A comparative analysis of ensemble learning and deep learning methods," *Int. Rev. Financ. Anal.*, vol. 92, Art. no. 103055, Mar. 2024, doi: 10.1016/j.irfa.2023.103055.

[42] J. Chen, "Analysis of Bitcoin Price Prediction Using Machine Learning," *J. Risk Financ. Manag*, vol. 16, no. 1, Art. no 51, Jan. 2023, doi: 10.3390/jrfm16010051.

[43] M. Khalid Salman and A. Abdu Ibrahim, "Price prediction of different cryptocurrencies using technical trade indicators and machine learning," *IOP Conf. Ser.: Mater. Sci. Eng.*, vol. 928, no. 3, Art. no. 032007, Nov. 2020, doi: 10.1088/1757-899X/928/3/032007.

[44] J. Abraham, D. Higdon, J. Nelson, and J. Ibarra, "Cryptocurrency price prediction using tweet volumes and sentiment analysis," *SMU Data Sci. Rev.*, vol. 1, no. 3, Art. no. 1, Jan. 2018.

[45] R. Parekh *et al.*, "*DL-GuesS*: Deep Learning and Sentiment Analysis-Based Cryptocurrency Price Prediction," *IEEE Access*, vol. 10, pp. 35398–35409, 2022, doi: 10.1109/ACCESS.2022.3163305.

[46] F. Valencia, A. Gómez-Espinosa, and B. Valdés-Aguirre, "Price Movement Prediction of Cryptocurrencies Using Sentiment Analysis and Machine Learning," *Entropy*, vol. 21, no. 6, Art. no. 589, Jun. 2019, doi: 10.3390/e21060589.

[47] T. R. Li, A. S. Chamrajnagar, X. R. Fong, N. R. Rizik, and F. Fu, "Sentiment-Based Prediction of Alternative Cryptocurrency Price Fluctuations Using Gradient Boosting Tree Model," *Front. Phys.*, vol. 7, Art. no. 98, Jul. 2019, doi: 10.3389/fphy.2019.00098.

[48] O. Kraaijeveld and J. De Smedt, "The predictive power of public Twitter sentiment for forecasting cryptocurrency prices," *J. Int. Finan. Mark., Instit. Money*, vol. 65, Art. no. 101188, Mar. 2020, doi: 10.1016/j.intfin.2020.101188.

[49] H. M. Fadhil and N. Q. Makhool, "Forecasting Cryptocurrency Market Trends with Machine Learning and Deep Learning," *BIO Web Conf.*, vol. 97, Art. no. 00053, Apr. 2024, doi: 10.1051/bioconf/20249700053.

[50] S. Yu, "Comparative analysis of machine learning techniques for cryptocurrency price prediction," *ACE*, vol. 32, no. 1, pp. 1–12, Jan. 2024, doi: 10.54254/2755-2721/32/20230175.

[51] Jayanta Aich, "Performance Scrutiny of Price Prediction on Blockchain Technology Using Machine Learning," *J. Electr. Sys.*, vol. 20, no. 7s, pp. 446–452, May 2024, doi: 10.52783/jes.3339

[52] G. Cheuque Cerda and J. L. Reutter, "Bitcoin Price Prediction Through Opinion Mining," in *Compan. Proceed. 2019 World Wide Web Conf.*, San Francisco USA: ACM, May 2019, pp. 755–762. doi: 10.1145/3308560.3316454.