JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

A SURVEY ON CRYPTOCURRENCY BIG-DATA ARCHITECTURE, PRICE PREDICTION MACHINE LEARNING ALGORITHMS AND SOCIAL MEDIA SENTIMENTS

[1]Pranav Vichur, [2]Jayashree M Kudari, [3]Abhijith Nair, [4]Raj Abhishek Pandey,
[5]Rishi Chowdary, [6]T Kanak Raj
[1,3,4,5,6] UG Scholar, [2]Associate Professor
[1,2,3,4,5,6]School of CS and IT
[1,2,3,4,5,6]Jain (Deemed-To-Be University), Bengaluru, India

Abstract – Cryptocurrencies are currently trending all over the world. Blockchain technology has made the cryptocurrency technology and its network more advanced. Cryptocurrencies are used for safe, anonymous online transactions. Despite being relatively simple to use, traditional statistical methods are flawed by seasonal effects. The best method for predicting cryptocurrency prices is machine learning. In this paper, it is illustrated how cryptocurrency's big data architecture can be applied to provide real-time analysis of cryptocurrency transactions and blocks. Understanding the various machine learning algorithms and techniques is essential when analyzing price dynamics of cryptocurrencies. Depending on the amount of historical data provided, most algorithms achieve similar accuracy. Moreover, researchers demonstrated how social media information and trends influence the reputation of certain cryptocurrencies, demonstrating the predictability of market prices.

Keywords: Lambda Architecture; ANN (Artificial Neural Network); LSTM (Long Short-Term Memory); RF (Random Forest); GRB (Gradient Boosting Machine); Tweepy; Decentralization;

I. INTRODUCTION

The rise of cryptocurrency began with Satoshi Nakamoto's Bitcoin white paper in 2008. Most cryptocurrencies are supported by blockchain networks, which process thousands of transactions every second. This process produces large amounts of data that is used to anticipate price fluctuations. [1]

Big data is described as information that is too vast for traditional technologies to handle. The term "big data" denotes a wide range of concepts, including the use of data, storing, collecting, and interpreting it, as well as the ubiquitous cultural shift that has affected businesses and society as a whole. Thousands of transactions are sent every second to multiple blockchain networks that facilitate the operation of various cryptocurrencies, such as Bitcoin and Ethereum. This process generates large amounts of data, which are processed to predict price fluctuations of cryptocurrencies. [1]
As social networking and online transactions have increased, more and more data is collected. Several clustering techniques are used to analyze this data, based on K-means, an unsupervised learning algorithm that is viral in the machine learning world. [3]

To predict crypto costs with reasonable precision, the paper employs Long Short-Term Memory (LSTM) and the Recurrent Neural Network (RNN) on freely available data. CoinMarketCap was used to compile the data for this study. Ripple, Ethereum, Litecoin, Bitcoin Cash, and Ripple were among the digital currencies for which time-series data was used. The dataset utilized six key highlights. The depictions are as per the following:
- Open Price: It is the opening price of a currency.
- Close Price: It is the everyday closing price of the market.

- Excessive cost: The daily most exorbitant cost for a cash transaction.
- Low Price: The everyday least cost for money.
- Volume: It is the amount of money that is exchanged on a day-to-day basis.
- Market Cap: The total market capitalization of the money on a daily basis. This varies depending on the fluctuation in costs on any particular day. [4]

II. OBJECTIVES

The aim is to improve the understanding of the current cryptocurrency's big-data architecture, its technological advancements and implementation as well as determine the most suitable machine learning algorithm with differing methodologies for price prediction and social media sentiment analysis.

III. LITERATURE REVIEW

1. *ARCHITECTURE*

A. Approach on Lambda Architecture

Designing an architecture to process and analyze cryptocurrency data can be extremely challenging. Lambda architecture is proposed as a solution to this problem. The system should maintain the capability of processing both historical data and real-time data that can be combined with previously processed data to create a complete picture.

The approach was developed by Nathan Marz and the Lambda architecture is considered one of the best solutions in the market for processing cryptocurrency data. The process of extracting useful information from huge amount of data can be time-consuming. Processes that take minutes to hours have very little efficiency. Analysts prefer to get real-time results to use with prior batch analysis results. In response to the problem, Lambda architecture creates two distinct data flows: hot path and cold path. [1]

All incoming data is stored as raw data in a batch layer (cold path), and it is processed in batches. Data is processed in real-time in a speed layer (hot path), reducing latency but sacrificing accuracy. The speed layer also provides data in increments based on the most recent data. [1]

B. Proposed Big-Data Architecture

The proposed solution, which is based on the Lambda architectural approach, is made up of several components. Data sources consist of multiple data streams, a message passing for inter-component communication, having three processing layers namely: batch layer, speed layer, and a serving layer, and a client application with an appropriate user interface are all shown in Figure 1. [1]

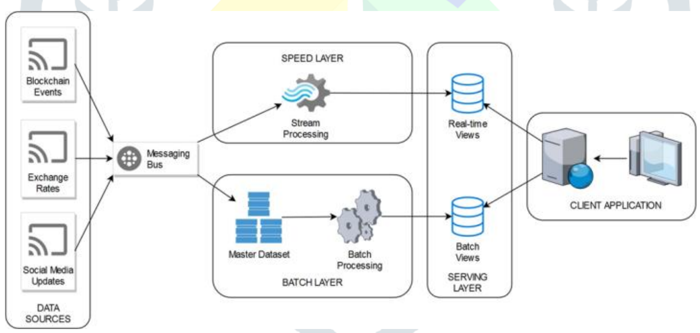

Figure 1. The proposed big data architecture as a graphical representation. [1]

1. Data Sources

The most straightforward and fastest way would utilize public APIs to recover new blockchain information. The information gathered from news websites, well-known websites, and social networking and expert advice that is relevant and connected to specific cryptographic money or the digital currency sector is considered web-based media information. [1]

2. Batch Layer

The Master Dataset is created by appending fresh data entries from a variety of Kafka topics. Each entity type's data is organized into separate directories, each of which contains many JSON-formatted text files. The data is aggregated to produce an "average" of cryptocurrency popularity over time. Figure 2 depicts a conceptual entity-relationship diagram. [1]

Data from many directories, such as the price of any cryptocurrency is read first, followed by data from social media, resulting in several data frames, each for a separate social network. The data is aggregated to produce an "average" of cryptocurrency popularity over time. [1]

3. Speed Layer

Using Spark Streaming or Kafka Streams API, the Speed layer of Apache Spark can execute real-time data processing on streaming data. A correlation calculation is conducted by consuming data streams of interest, which are created by filtering, mapping, aggregating, and joining operations. [1]

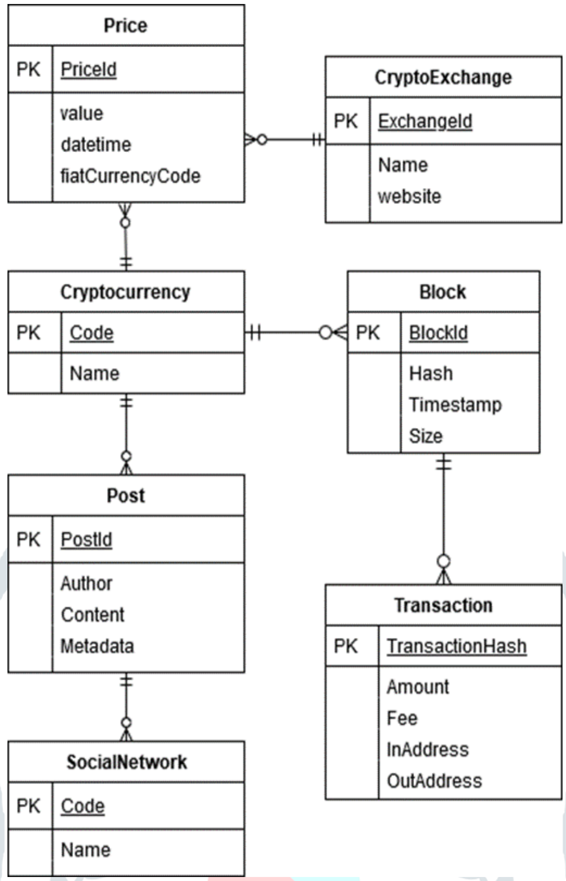

Figure 2. The master dataset's conceptual entity-relationship diagram [1]

4. Serving Layer

All data is stored on the same database for both real-time and for batch ingestion. Apache Druid serves as an excellent candidate for the serving layer. It supports both relational and time-series data structures, and its primary timestamp is ideal for blockchain data. [1]

5. The Client Application

Client applications consist of back-end web applications integrated with REST APIs and front-end JavaScript code. Time-series charts, pricing comparison diagrams, block inspectors, histograms are all included in the front-end applications. Different aggregations and sorting choices should be straightforward to express with interactive query tools. [1]

2. MACHINE LEARNING ALGORITHMS

I. Cryptocurrency Price Analysis Using Artificial Intelligence

A. Data Collection and Analysis

The study examined the values of three of the most popular cryptocurrencies: Bitcoin, Ripple, and Ethereum. For historical pricing data, 1030 trading days were taken from https://www.blockchain.com/markets. To avoid overfitting during model training, the dataset was divided into 80/20 training and testing sets. [2]

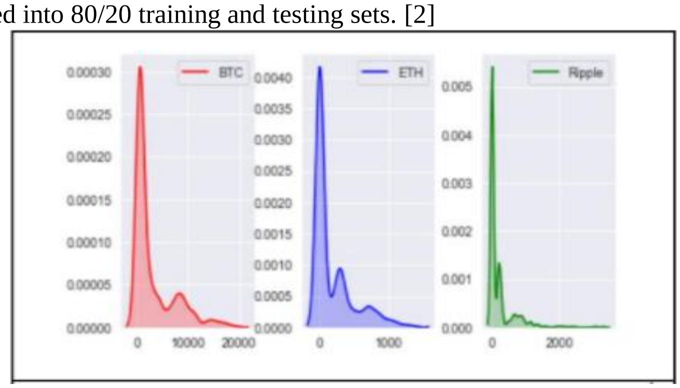

Figure 3. The density distribution of price history of three cryptocurrencies. [2]

Figure 3 illustrates the density distribution of price history for Bitcoin on the left, Ethereum being in the middle, and Ripple on the right respectively from 7th August 2015 to 2nd June 2018. [2]

B. Models

Financial markets have been predicted using deep neural networks. The two most commonly utilized deep learning models are Long-Short-Term Memory Recurrent Neural Networks (LSTM RNN) and fully connected Artificial Neural Networks (ANN). [2]

C. Results

A. ANN Estimate of the Time-Series Memory

ANN model is utilized in the first phase of the ANN model in order to predict the latest price of Bitcoin one day in the future using five different memory lengths: 7, 14, 21, 30, and 60 days. The mean square error and Pearson correlation are used to calculate the error between data and the model. Figure 4 shows the final results. [2]

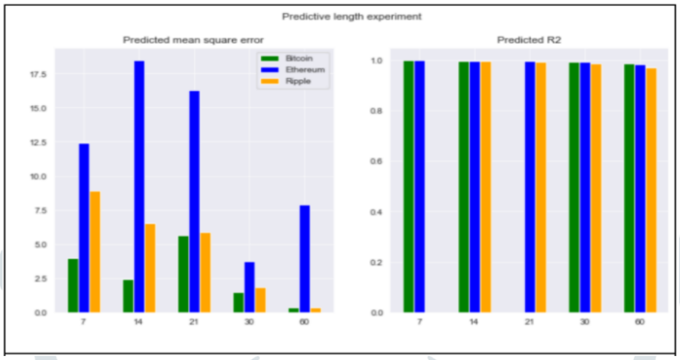

Figure 4. The model data mean square error and Pearson correlation of an ANN model according to price history. [2]

Figure 4 illustrates an ANN model with 7, 12, 21, 30, and 60 days of price history as inputs, the left and right panels depict, respectively, the model data mean square error and the Pearson correlation. [2]

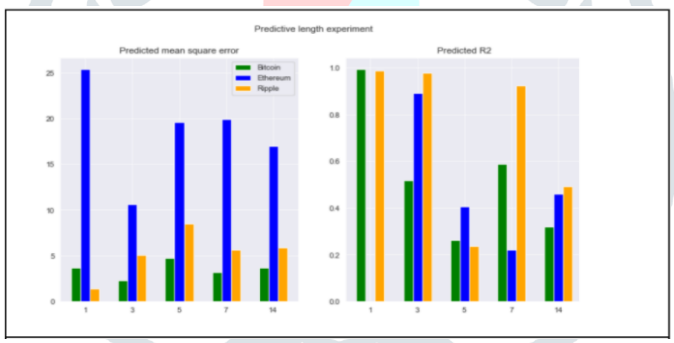

Figure 5. The model data mean square error and Pearson correlation of an ANN model according to performance. [2]

In the Figure 5, the left and the right panels provide the ANN model's performance over 1, 3, 5, 7, and 14 days. The left panel shows the model data mean square error and the right displays the Pearson correlation. [2]

B. LSTM Estimate of the Time-Series Memory

In predicting the one day future price of multiple cryptocurrencies, the LSTM model, which is based on mean square error, performs similarly to the ANN model. It shows that, despite its inherent limitations, ANN is capable of extracting and utilizing essential information concealed in previous price dynamics. [2]

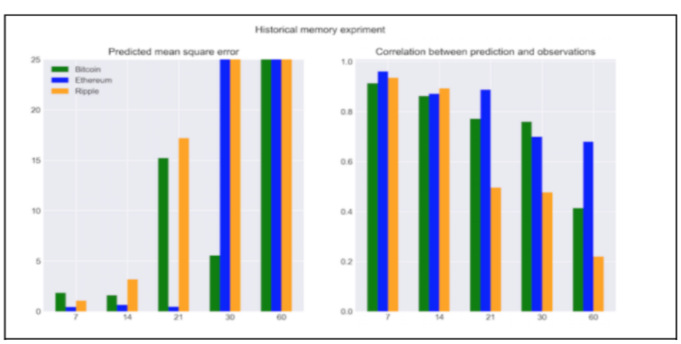

Figure 6. The model data mean square error and Pearson correlation of an LSTM model. [2]

Figure 6 illustrates the model data mean square error on the left panel and Pearson correlation on the right panel, respectively, are depicted for the LSTM model with inputs of 7, 14, 21, 30, and 60 days. [2]

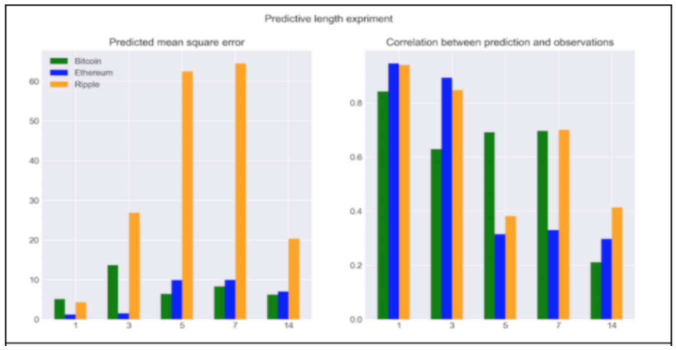

Figure 7. The LSTM model's performance on prediction period. [2]

Figure 7, the graph shows the performance of the LSTM model with prediction period of 1, 3, 5, 7, or 14 days. The right and left panels, respectively, show the model data mean square error and the Pearson correlation. [2]

C. Limitation and Future Work

The experiment is limited to the study of price trends of three digital currencies only. A drawback is the optimization procedure, and this model's parameters aren't well evaluated. [2]

III Traditional Time-Series Models Using LSTM Algorithm

1. LSTM (Long Short-Term Memory)

LSTM is the state-of-the-art neural organization that addresses a few unsolvable errands of the time series utilizing fixed size time window and is a persevering design of a neural organization. LSTM scaffolds delay more than 1000 discrete strides to take care of issues with steady blunder carousels. It can connect delays without losing the capacities of brief time frame slacks. LSTM can figure out how to connect time stretches moving toward 1000 stages without losing the capacities of brief time frame slacks. Neither detonating nor vanishing mistake course through the interior condition of extraordinary units, in this way accomplishing an efficient, gradient-based calculation for consistent design consistency. [4]

2. Proposed LSTM (Long Short-Term Memory)

RNN with a period space definition utilizes self-circle associations to recollect the past and has a straight memory block for the secret worth, and three doors to direct the progression of data between the memory block and the result. Utilizing a LSTM organization to foresee digital money value esteem, the hourly worth of cryptographic money is anticipated by utilizing the given element vectors and the LSTM organization. [4]

3. Results and Discussion

The RMSE (Root Mean Square Error) and MSE (Mean Square Error) were employed as output comparison metrics. Mistakes are used to quantify the price prediction performance of cryptocurrencies and are formally stated as:

Models	RSME	MSE	Accuracy (%)
[10]	0.0499	0.3115	50.35
[11]	0.0510	0.3342	49.08
[12]	0.0543	0.3516	50.67
Proposed Model	0.0634	0.4321	67.43

Table 1. Comparing the performance of existing models [4]

$$RMSE = \sqrt{\frac{1}{N_{test}} \sum_{n=1}^{N_{test}} \left(\overline{y}_n - y_n\right)^2}; \qquad [4]$$

$$MSE = \frac{1}{N_{test}} \sum_{n=1}^{N_{test}} \left(\overline{y}_n - y_n\right)^2, \qquad [4]$$

Here y_n defines the genuine worth of the n^{th} expectation point and $–y_n$ is the anticipated worth and N test is the quantity of the test set samples. Table 1 evaluates the presentation examination between the model that is proposed and other comparative models for the cryptographic money value forecast. It shows the created model to perform better with an exactness of 67.43%, [12] trailing by 50.67%, [10] with 50.35% and [11] with 49.08% precision, respectively. [4]

IV. PRICE PREDICTION ALGORITHMS

1. *Random Forest (RF):*

Proposed by Leo Breiman, it is a ML approach. Preparing informational collection is partitioned into numerous arbitrary subsets with substitution (bootstrap tests) and every one of the base classifiers is prepared in its own sub-set. The last classifier (x, q)$_N$ is worked as normal of the fundamental calculations (x) i$_h$ (for regression):

$$(x_i, y_i)_{i=1,2,3,\dots,N} \quad [5]$$

The number of observations (samples) in the training set is denoted by N. The base models aren't very accurate on their own, but their ensemble considerably improves prediction quality. [5]

$$a_N(x,\theta) = \frac{1}{N}\sum_{i=1}^{N} h_i(x,\theta) \quad [5]$$

2. *Gradient Boosting Machine (GBM):*

The basis of GBM is covered by following Jerome H Friedman's comparison with RF to produce a weighted summation of the N basic algorithms. It concludes with:

$$a_N(x,\theta) = a_{N-1}(x,\theta) + \lambda\big(\gamma_N h_N(x)\big) \quad [5]$$

The principal benefit of supporting is that it decreases both fluctuation and inclination in estimating, in any case the decrease of the conjecture mistake is yet completed predominantly because of the decrease in predisposition. [5]

3. *Hyper-Parameters Tuning:*

Stochastic GBM (SGBM), which is used in the paper, is a GBM modification that makes use of such a segment. The models are fitted using the preparation test by adding simple trees to their outfits. [5]

Parameters	**RF**	**GBM**
Loss of Function	quadratic	quadratic
Proposition of training or test subsamples, % (in term of percentage)	70/30	70/30
Subsample Rates of random samples	0.7	0.7
Ensemble size in trees	500	250
The maximum number of features per split	15	4
The maximum number of terminal nodes in trees	12	-
The minimum samples in a child node	150	15
The learning rate or the shrinkage		0.1

Table 2. The accounted for the last upsides of hyper-boundaries setting. [5]

V. Improvising Algorithms Using Reinforcement Learning

1. *Price Prediction with Machine Learning:*

Artificial Intelligence based machine learning algorithms such as Deep Learning make accurate predictions by tackling complex and nonlinear problems using a large amount of data because of the large number of qualities, value forecasting with exact outcomes is difficult, and the deep learning approach overcomes such problems. [6]

Summary	**Technique**	**Cryptocurrency**	**Dataset**
Using the bitcoin transaction network for the price prediction	Single layer network	Bitcoin	Coindesk
Using the ARIMA to predict the price of bitcoin	ARIMA	Bitcoin	-
Using the largest Lyapunov Exponent and Detrended Fluctuation Analysis	LSTM	Bitcoin, Ripple	-
For ensemble regression trees, the XGBoost and the LSTM are used.	RNN, LSTM	1681 cryptocurrencies	Coinmarket.com
Using four classes to engineer features	LDA, Logistics Regression, SVM, LSTM	Bitcoin	Coinmarket.com
Prediction of the telegram and trends data	LSTM	Bitcoin, Ethereum	Google trend, telegram, market data aggregate
Twitter mood and Google trends were used to forecast bitcoin prices, with the prediction being improved based on Wikipedia and Facebook posts.	RNN, LSTM, Regression	Bitcoin	Coiindesk.com, google trends, twitter API

Using correlation analysis to investigate the crypto currency market	Multivariate regression	Bitcoin, Ethereum	Etherscan.io, blockchain.info

Table 3. It represents the summary of late related examinations regarding cryptocurrency value prediction. [6]

VI. Price Prediction Time-Based Analysis

1. Price Prediction Daily Analysis:

According to Litecoin's day-to-day examination. The value records are in United States Dollars. The forecast is based on daily data on digital currency prices. The forecast is based on the size of the progression and the forecast window. [6]

2. Price Prediction Weekly Analysis:

The daily investigation calculates the average price of Litecoin over multiple weeks and 7 days in US Dollars. The prediction is based on the progress size and the RSME. [6]

3. Price Prediction Monthly Analysis:

The monthly value prediction after-effects of Litecoin are introduced. It tends to be the real worth and the anticipated worth are close. [6]

VII. Traditional Statistical And Machine Learning Predictions

Cryptocurrencies are used as online payment methods, and their popularity is growing. Many governments have begun to investigate cryptocurrency taxation and regulation. Accountants and auditors are now accepting cryptocurrency payments in exchange for business advice. According to a new estimate, the price of bitcoin will reach a peak of $18,000 per bitcoin in late 2017. [7]

The ARIMA time-series model was used to estimate the bitcoin price. Certain chain lets were found to have a strong predictive value for bitcoin values. Bitcoin, Ether, Monero, Litecoin, and Dash values are determined by their processing power and network. [7]

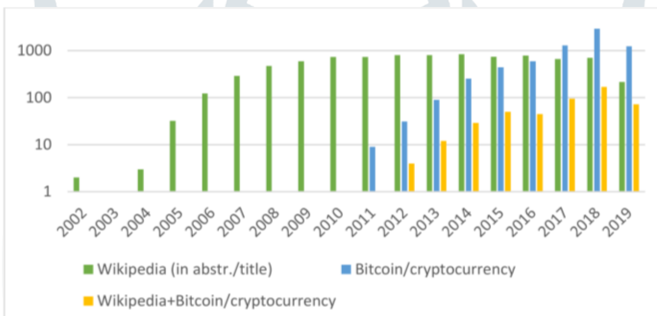

Figure 8. Bitcoin price trends from 2002 to 2019. [7]

Poyser (2019) examined the determinants of bitcoin's price level using Bayesian structural time-series and found that neutral investor sentiment, the gold price, and the Yuan to USD exchange rate all had negative effects on bitcoin's price level. Jang and Lee (2018) used a Bayesian NN (BNN) model to predict Bitcoin price variations and found that it was more accurate than SVR and linear regression benchmarked models. [7]

The tweets were analyzed using a multivariate linear regression model. Chen et al. (2020) used the machine learning technique to predict the bitcoin price, with an average accuracy of 62.2 percent versus 53.05 percent using linear discriminant analysis and logistic regression. [7]

Madan, Saluja, and Zhao (2015) used bitcoin data to test a variety of machine learning algorithms and discovered that binomial logistic regression, SVM, and RF worked well and had good accuracy. Sun, Zhou, and Lin (2019) created a prediction model for predicting bitcoin price patterns using a one-layer NN and a combination of generative and discriminative classifiers. [7]

Lahmiri and Bekiros used deep learning (DL), which included extracting hidden patterns from an underlying dynamical system and employing LSTM to identify domain-specific patterns. The performance of RNN, LSTM, and ARIMA for bitcoin price prediction was compared using a multi core and a graphics processing unit environment. Based on an RNN and LSTM, Yaoetal (2018) established a framework for predicting cryptocurrency values using deep learning based on the volume, the market cap, the circulating supply, and the maximum supply. [7]

Jiang and Liang proposed that RL should be used to solve a portfolio management issue. To get the best outcome, the researchers employed CNN without a model. Forecasting the price of cryptocurrencies, Sun et al. (2020) used the light GB machine (LightGBM) decision tree algorithm. For crypto-currency price prediction, Snihovyi, Ivanov, and Kobets (2018) and Altan, Karasu, and Bekiros (2019) developed DL and RL approaches. [7]

A genetic algorithm and neural nets-based time-series forecasting method was proposed. DL-ANN multiagent frameworks can be utilized to construct decision-support systems for time-series forecasting, according to financial trading. Loud's work (2020) demonstrates that the market has predictable short-term pricing movements. [7]

3. SOCIAL MEDIA MESSAGES AND SENTIMENTAL ANALYSIS

A. Data Pre-processing and Feature Selection

Tweepy is a free Python library used for accessing the Twitter API and gathering the Twitter data. Tweepy considers sifting hashtags or words. There are various manners by which the cryptographic forms of money of premium might be alluded into the tweets. The immediate way is by utilizing a hashtag "#" followed by "bitcoin" or "Ethereum". Because these hashtags were chosen as the only ones to be used with Tweepy to gather tweets, tweets with only "#bitcoin" and "#Ethereum" hashtags were the first to offer a large informational index. The client ID, a unique identifier that can't be changed, time stamp, and the number of times a tweet was "retweeted" (someone posted exactly the same tweet so that their followers could see it) and how frequently it was "favorited" were all collected for each post. Because tweets can be multilingual, the tweets are separated by language for this investigation. The tweets were collected using a content that ran every 15 minutes and collected 1,500 tweets per occurrence. [9]

B. Sentiment Analysis of Tweets

Around 23 million accounts of clients and tweets are from bots. Assuming the bots were sending tweets that contained polarized opinions about the digital currencies then those tweets might in any case impact individuals' interest to possess cryptographic forms of money and its costs. In any case, a significant number of the tweets don't contain any feeling and on second thought give just realities or are serving the capacity of promoting. The current USD cost of a solitary bitcoin is a reality and doesn't convey any feeling. Consequently, opinion examination of the tweets gives restricted data to the model. After pre-handling the gathered tweets, research says the data acquired from the tweets through opinion examination is of restricted worth. The truth is that "people are what they read," therefore it is obvious how their viewpoint can be greatly affected by tweets or friendly social media posts., created by bots, or notices. [9]

C. Twitter volume and cryptocurrency prices

The volume of tweets is the final model input to consider. Regardless of price changes, tweet sentiment tended to remain positive. This could be because people who continue to tweet about cryptocurrencies when prices are falling are interested in them for reasons other than their value, such as privacy.That is, however, a fact of technology, not something that ebbs and flows like price. This suggests that tweet volume, rather than sentiment, is a better metric to use because the number of people talking about cryptocurrencies on Twitter fluctuates with market prices. [9]

D. The Impact of the Tweets on a Common Man's Point of View

The fact remains: "People are what they read", thus; it is clear how the perspective of everyday person can without much of a stretch be affected by Tweets or any friendly media impact. Individuals may get roused to contribute significantly more than their abilities without understanding the unpredictability of the market. The utilization of watchwords like Tesla, rocket, Mars, starship, send-off, bitcoin, digital money, hold was represented isolating the tweets considering digital currency. The followings graph below indicates the spike in the price of bitcoin and spike in trading volume when Elon Musk changed his Twitter bio to #bitcoin [9]

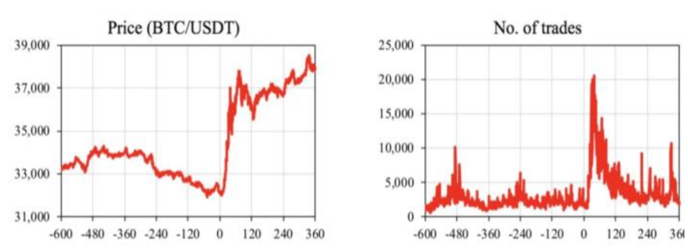

Figure 9. Twitter trend that affected Bitcoin price when Elon Musk changed Twitter bio to #bitcoin. [9]

IV. CONCLUSION

A wide array of new innovations can be applied to the field of cryptocurrencies and its applications in areas such as security, privacy, and data analysis. Yet, there are still many challenges and issues associated with accurately predicting cryptocurrency prices. After reviewing the best suited big-data architectures and machine learning algorithms proposed by various authors utilizing the historical data of cryptocurrencies like Bitcoin, Ethereum, Litecoin and Ripple. All machine learning models perform exceptionally well at predicting real-time prices, despite their fundamental structures and how social media trends expand and diminish the cryptocurrency's reputation and prices.

V. REFERENCE

[1] Horvat, Nebojša & Ivković, Vladimir & Todorović, Nikola & Ivančević, Vladimir & Gajic, Dusan & Luković, Ivan. (2020). Big Data Architecture for Cryptocurrency Real-time Data Processing.

[2] Yiying, Wang & Yeze, Zang. (2019). Cryptocurrency Price Analysis with Artificial Intelligence. 97-101. 10.1109/INFOMAN.2019.87147

[3] Mittal, Ruchi & Arora, Shefali & Bhatia, Mohinder: Pal Singh. (2018). AUTOMATED CRYPTOCURRENCIES PRICES PREDICTION USING MACHINE LEARNING. International Journal of Soft Computing. 8. 1758-1761. 10.21917/ijsc.2018.0245.

[4] Bamidele, Awotunde & Ogundokun, Roseline & Jimoh, Rasheed & Misra, Sanjay & Aro, Taye. (2021). Machine Learning Algorithm for Cryptocurrencies Price Prediction. 10.1007/978-3-030-72236-4_17.

[5] Derbentsev, V., Babenko, V., Khrustalev, K., Obruch, H., & Khrustalova, S. (2021). Comparative Performance of Machine Learning Ensemble Algorithms for Forecasting Cryptocurrency Prices. *International Journal of Engineering, 34*, 140-148.

[6] Shahbazi, Zeinab & Byun, Yongcheol. (2021). Improving the Cryptocurrency Price Prediction Performance Based on Reinforcement Learning. IEEE Access. PP. 1-1. 10.1109/ACCESS.2021.3133937.

[7] Khedr, Ahmed & Arif, Ifra & P V, Pravija & El-Bannany, Magdi & Alhashmi, Majed & Sreedharan, Meenu. (2021). Cryptocurrency price prediction using traditional statistical and machine-learning techniques: A survey. Intelligent Systems in Accounting, Finance and Management. 28. 10.1002/isaf.1488.

[8] Tandon, Chahat & Sanjana, R. & Palivela, Hemant & Parihar, Sidharth. (2021). How to 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. 100035. 10.1016/j.jjimei.2021.100035.

[9] Mohapatra, Shubhankar & Ahmed, Nauman & Alencar, Paulo. (2020). KryptoOracle: A Real-Time Cryptocurrency Price Prediction Platform Using Twitter Sentiment.

[10] Lahmiri, S., & Bekiros, S. (2019). Cryptocurrency forecasting with deep learning chaotic neural networks. Chaos, Solitons & Fractals, 118, 35–40.

[11] Wu, C. H., Lu, C. C., Ma, Y. F., & Lu, R. S. (2018, November). A new forecasting framework for bitcoin price with LSTM. In 2018 IEEE International Conference on Data Mining Workshops (ICDMW) (pp. 168–175). IEEE.

[12] Bonatti, A., & Cisternas, G. (2020). Consumer scores and price discrimination. The Review of Economic Studies, 87(2), 750–791.