Cryptocurrency: Analysis, Visualization and Prediction

*Project Report for *Indian Institute of Technology, Bombay* - DS 203: Programming for Data Science (2022)

K. Solanki
Department of Physics
Indian Institute of Technology, Bombay
Mumbai, Maharashtra
210260026@iitb.ac.in

P. Aggarwal
Department of Physics
Indian Institute of Technology, Bombay
Mumbai, Maharashtra
210260038@iitb.ac.in

A. Swamy
Department of Physics
Indian Institute of Technology, Bombay
Mumbai, Maharashtra
210260008@iitb.ac.in

***Abstract*—A cryptocurrency is a digital currency, which is an alternative form of payment created using encryption algorithms. The use of encryption technologies means that cryptocurrencies function both as a currency and as a virtual accounting system. To use cryptocurrencies, you need a cryptocurrency wallet. These wallets can be software that is a cloud-based service or is stored on your computer or on your mobile device. The wallets are the tool through which you store your encryption keys that confirm your identity and link to your cryptocurrency.**

Improving our understanding of the cryptocurrency markets and performing successful cryptocurrency investments takes a considerable amount of time and skill. With the recent emergence of many artificial intelligence (AI) and machine learning (ML) tools, investors and traders are increasingly looking to develop more efficient mechanisms of investing. Both AI and ML are proving to be efficient in crypto markets as crypto investing is adopted by further institutional and individual investors.

In this project, our primary objective is to analyze data on historical prices of various cryptocurrencies, depict the trends graphically and try to predict future prices using ML models.

I. Introduction

A. What is cryptocurrency?

Cryptocurrency, sometimes called crypto-currency or crypto, is any form of currency that exists digitally or virtually and uses cryptography to secure transactions. Cryptocurrencies don't have a central issuing or regulating authority, instead using a decentralized system to record transactions and issue new units.

Cryptocurrency is a digital payment system that doesn't rely on banks to verify transactions. It's a peer-to-peer system that can enable anyone anywhere to send and receive payments. Instead of being physical money carried around and exchanged in the real world, cryptocurrency payments exist purely as digital entries to an online database describing specific transactions. When you transfer cryptocurrency funds, the transactions are recorded in a public ledger. Cryptocurrency is stored in digital wallets.

Cryptocurrency received its name because it uses encryption to verify transactions. This means advanced coding is involved in storing and transmitting cryptocurrency data between wallets and to public ledgers. The aim of encryption is to provide security and safety.

The first cryptocurrency was **Bitcoin**, which was founded in 2009 and remains the best known today. Much of the interest in cryptocurrencies is to trade for profit, with speculators at times driving prices skyward.

B. How does cryptocurrency work?

Cryptocurrencies run on a distributed public ledger called blockchain, a record of all transactions updated and held by currency holders.

Units of cryptocurrency are created through a process called mining, which involves using computer power to solve complicated mathematical problems that generate coins. Users can also buy the currencies from brokers, then store and spend them using cryptographic wallets.

If you own cryptocurrency, you don't own anything tangible. What you own is a key that allows you to move a record or a unit of measure from one person to another without a trusted third party.

Although Bitcoin has been around since 2009, cryptocurrencies and applications of blockchain technology are still emerging in financial terms, and more uses are expected in the future. Transactions including bonds, stocks, and other financial assets could eventually be traded using the technology.

C. Cryptocurrency Examples

There are thousands of cryptocurrencies. Some of the best known include:

- **Bitcoin** : Founded in 2009, Bitcoin was the first cryptocurrency and is still the most commonly traded. The currency was developed by Satoshi Nakamoto – widely believed to be a pseudonym for an individual or group of people whose precise identity remains unknown.
- **Ethereum** : Developed in 2015, Ethereum is a blockchain platform with its own cryptocurrency, called Ether (ETH) or Ethereum. It is the most popular cryptocurrency after Bitcoin.
- **Litecoin** : This currency is most similar to bitcoin but has moved more quickly to develop new innovations,

including faster payments and processes to allow more transactions.

- **Ripple** : Ripple is a distributed ledger system that was founded in 2012. Ripple can be used to track different kinds of transactions, not just cryptocurrency. The company behind it has worked with various banks and financial institutions.
- Non-Bitcoin cryptocurrencies are collectively known as "altcoins" to distinguish them from the original.

II. DATASET AND METHODOLOGY

The dataset has one csv file for each cryptocurrency. Price history is available on a daily basis from April 28, 2013. This dataset has the historical price information of some of the top crypto currencies by market capitalization.

- Date : date of observation
- Open : Opening price on the given day
- High : Highest price on the given day
- Low : Lowest price on the given day
- Close : Closing price on the given day
- Volume : Volume of transactions on the given day
- Market Cap : Market capitalization in USD

This data has been taken from coinmarketcap.com and it is free to use data.

III. DATA VISUALIZATION

We have data on prices of different cryptocurrencies. To understand the data better, we present time-series plots showing cryptocurrency price analysis over the years.

A. Bitcoin

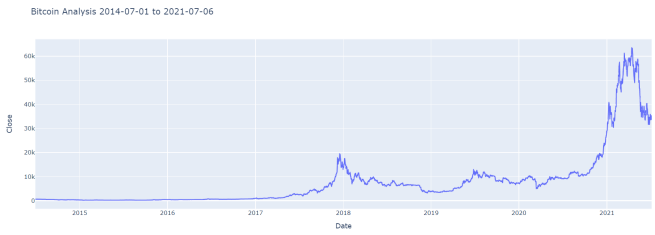

Fig. 1. Bitcoin analysis

Notice (Fig. 1) how the price of bitcoin was steadily increasing till 2018. Then there was a spike in its price in 2018 followed by another spike in 2021.

B. Litecoin

Fig. 2 shows that the fluctuations in litecoin prices is more than that of bitcoin. It showed two huge spikes in 2018 and 2021 and another spike in 2019-20.

Fig. 2. Litecoin analysis

Fig. 3. Ethereum analysis

C. Ethereum

Ethereum prices (Fig. 3) showed a a spike in 2018 and a huge peak in 2021, similar to Bitcoin.
In general, we observe that cryptocurrency prices shot up in 2018 and 2021.

This leads us to an obvious test: to check the correlation amongst cryptocurrencies.

Fig. 4 shows the heatmap, a correlation matrix amongst the multitude of cryoptocurrencies.
We see that most cryptos have a positive correlation coefficient with the price of bitcoin. This could be because the great positive shock in the price of bitcoin (as it eventually

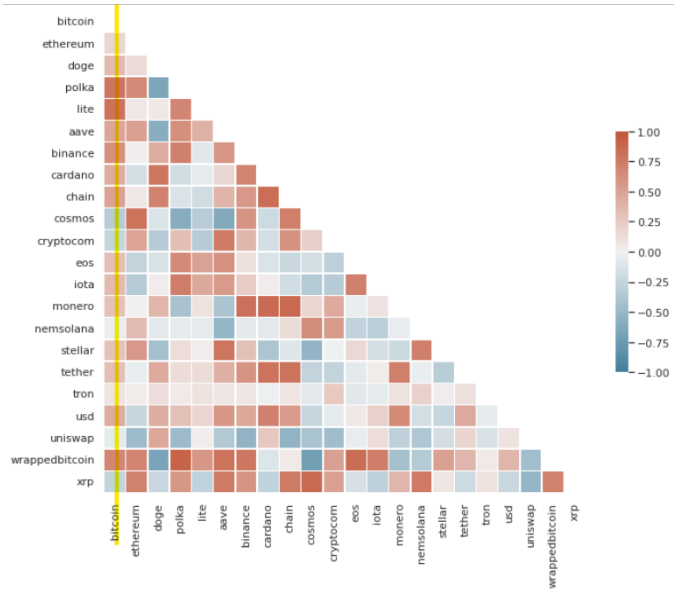

Fig. 4. Crypto Heatmap

reached 60000 USD) led to ”hype” for cryptocurrencies in the market. We now test this theory by limiting the time over the correlation coefficient is calculated to the region where bitcoin had its spike, namely 2017 to 2021.

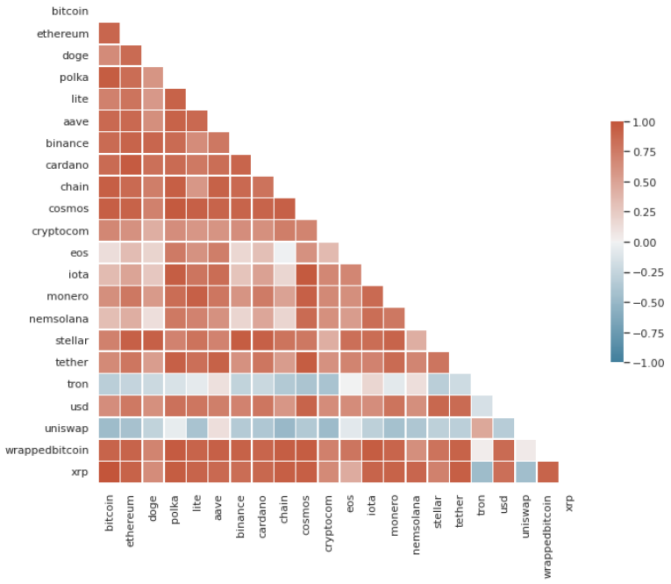

Fig. 5. Crypto Heatmap 2017-2021

(Fig. 5) We see that our hypothesis was indeed correct. Intuitively, we deduce that the sharp rise in bitcoin led to an increase in the investment of all cryptocurrencies.

IV. Random Correlations

The premise of this section is to check if certain random ’variables’ are correlated i.e, affect and/or are affect by the values of cryptocurrencies. It is important to know while navigating this section: Correlation is NOT Causation.

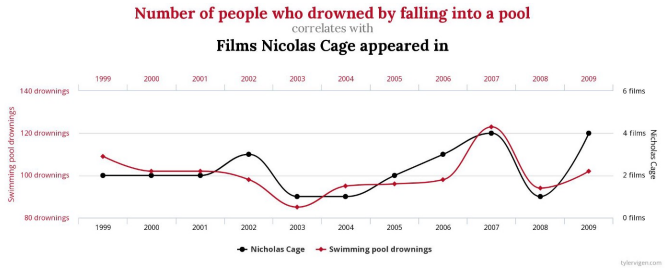

Fig. 6.

Meet Nicholas Cage, a famous American actor who has acted in movies such as National Treasure, Ghost Rider and National Treasure: Book of Secrets (Variety, I know right). In the above graph, we clearly see the correlation between the number of Nicholas Cage movies released versus the number of swimming pool drownings in a year. Obviously, these are independent events and have no causative relation.

Similarly, we look at how crypto prices correlate with seemingly random data, and if correlation coefficients predict some sort of relation, we investigate potential causes.

First, we look at some of the top Market Price Indexes, namely Nifty50, S&P500 and Dow Jones.

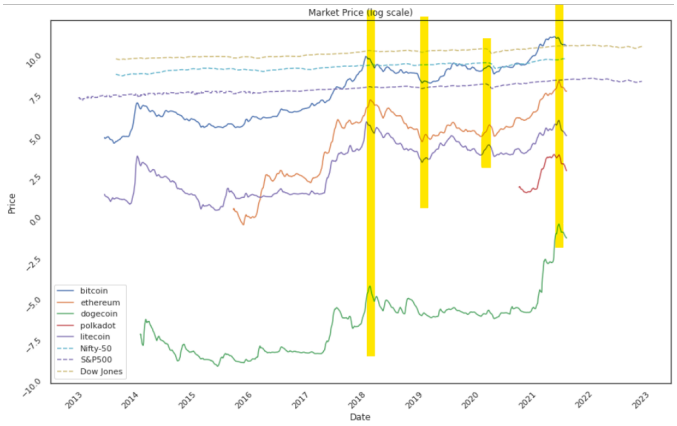

Fig. 7. Market Indexes (log scale)

Clearly, we can visually deduce that there is none, or negligible correlation between these market indexes and cryptocurrencies. We do see a small dip in the market indexes around 2020, but that was due to the COVID-19 recession. Fig. 7 also has highlighted matching peaks of different cryptocurrencies specifically in the 2018-2021 period.

The above brings to light a new idea: did the COVID-19 pandemic have any effect on the market value of cryptocurrencies? Let us analyse this. In this report, we will choose to look at the countries which were hit the hardest by COVID-19.

Consider the following graphs: (Fig. 8)

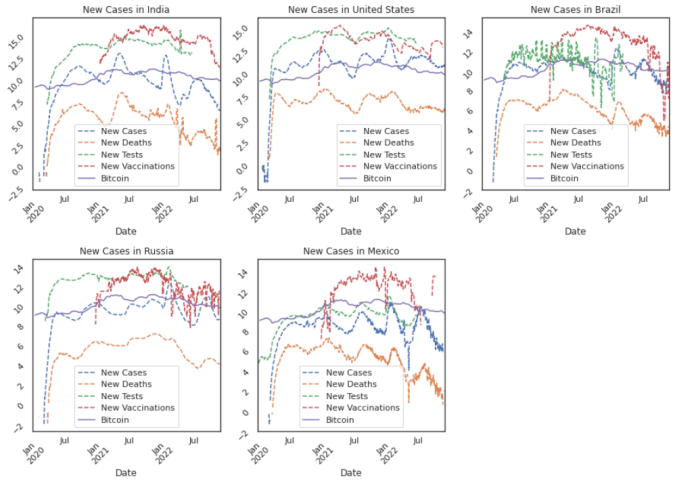

Fig. 8. COVID-19

The COVID-19 data at first looks very erratic and uncorrelated. However, when we calculate the correlation coeffi-

cients... *Drumroll...*

```
'The correlation coefficient between new cases in India and the price of bitcoin: 0.15879494365880897'
'The correlation coefficient between new deaths in India and the price of bitcoin: 0.05326204866001644'
'The correlation coefficient between new tests in India and the price of bitcoin: 0.4798398603296269'
'The correlation coefficient between new vaccinations in India and the price of bitcoin: 0.07219424430171377'
''
'The correlation coefficient between new cases in United States and the price of bitcoin: 0.16484843075475203'
'The correlation coefficient between new deaths in United States and the price of bitcoin: 0.05351935055893808'
'The correlation coefficient between new tests in United States and the price of bitcoin: 0.33587949248030996'
'The correlation coefficient between new vaccinations in United States and the price of bitcoin: 0.13751784051712884'
''
'The correlation coefficient between new cases in Brazil and the price of bitcoin: 0.14897455816513042'
'The correlation coefficient between new deaths in Brazil and the price of bitcoin: 0.029877836290027014'
'The correlation coefficient between new tests in Brazil and the price of bitcoin: 0.06368864929896587'
'The correlation coefficient between new vaccinations in Brazil and the price of bitcoin: 0.13227224630076495'
''
'The correlation coefficient between new cases in Russia and the price of bitcoin: 0.2481149029451739'
'The correlation coefficient between new deaths in Russia and the price of bitcoin: 0.4509134761239863 6'
'The correlation coefficient between new tests in Russia and the price of bitcoin: 0.23082978446375146'
'The correlation coefficient between new vaccinations in Russia and the price of bitcoin: 0.090561408815094 2'
''
'The correlation coefficient between new cases in Mexico and the price of bitcoin: 0.11059783054759639'
'The correlation coefficient between new deaths in Mexico and the price of bitcoin: 0.021539800340270688'
'The correlation coefficient between new tests in Mexico and the price of bitcoin: 0.38110039698521897'
'The correlation coefficient between new vaccinations in Mexico and the price of bitcoin: 0.09251502444019574'
''
'The correlation coefficient between new cases in France and the price of bitcoin: 0.2853036009724069'
'The correlation coefficient between new deaths in France and the price of bitcoin: 0.0669741155599272'
'The correlation coefficient between new tests in France and the price of bitcoin: 0.5000618496319337'
'The correlation coefficient between new vaccinations in France and the price of bitcoin: 0.1081066026687659'
''
'The correlation coefficient between new cases in Germany and the price of bitcoin: 0.3080389754243058'
'The correlation coefficient between new deaths in Germany and the price of bitcoin: 0.2543690152216802 6'
'The correlation coefficient between new tests in Germany and the price of bitcoin: 0.3695911306553574'
'The correlation coefficient between new vaccinations in Germany and the price of bitcoin: 0.14166671512619292'
```

Fig. 9. Correlation Coefficients

(Fig. 9) We see that there is indeed little to no correlation and deem that pursuing some causative theory is not worth it.

This section takes you through how we used data analysis to predict the trends in prices of majorly three cryptocurrencies: Bitcoin, Ethereum & Litecoin

We begin with analysing the closing prices histogram distribution and try to figure out where do majority of the closing prices lie so we roughly get an estimate is to how dispersed the closing price data is

V. Histograms (Frequency v/s Closing Price)

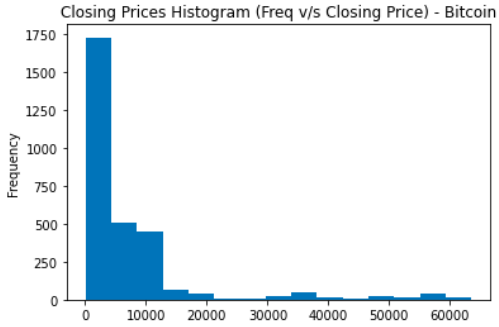

Fig. 10.

We observe (Fig.10) that most of the Bitcoin close prices fall in bins with values $\leq$ 20000.

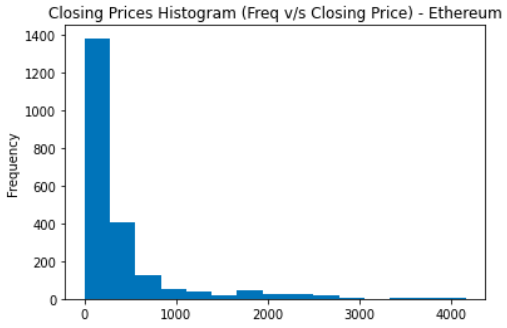

Fig. 11.

We observe (Fig.11) that most of the Ethereum close prices fall in bins with values $\leq$ 1000.

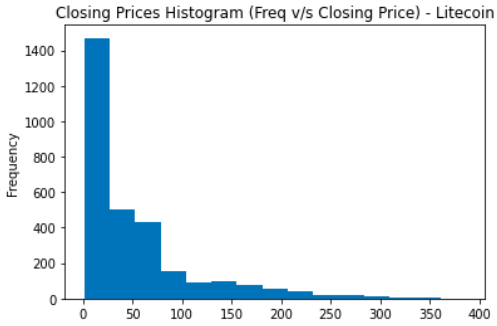

Fig. 12.

We observe (Fig.12) that most of the Litecoin close prices fall in bins with values $\leq$ 200.

VI. 30-day Rolling Prices

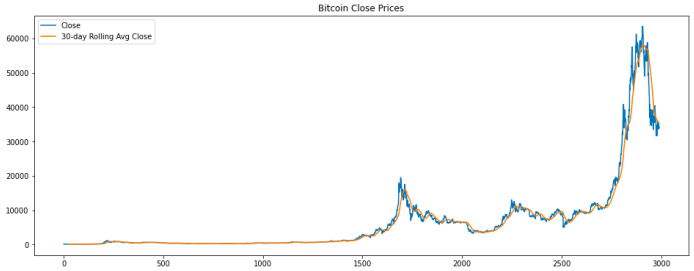

Fig. 13.

Shown above (Fig.13) is the variation in price of Bitcoin from 2013-04-29 to 2021-07-06

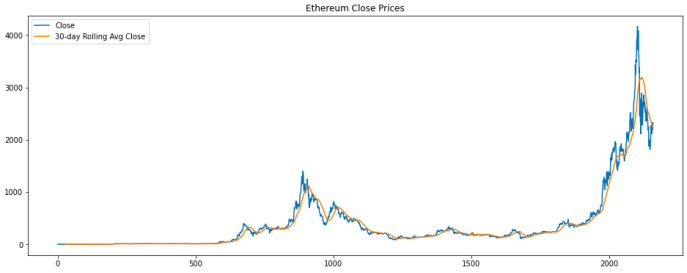

Fig. 14.

Shown above (Fig.14) is the variation in price of Ethereum from 2015-08-08 to 2021-07-06

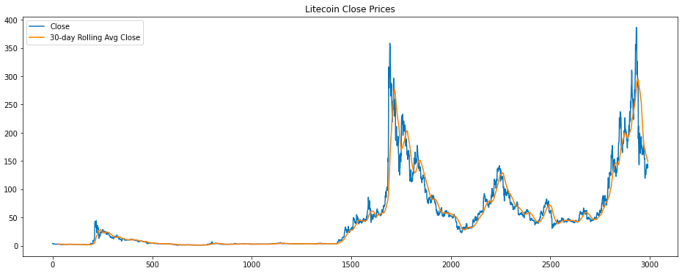

Fig. 15.

Shown above (Fig.15) is the variation in price of Litecoin from 2013-04-29 to 2021-07-06.

VII. Prediction of prices and Error Analysis

Firstly, to use data, we scaled it to normalised data using the **MinMaxScaler** feature from the **sklearn.preprocessing** module.
We used 80% of the data for training and 20% of the data will be predicted based on it. Training the model begins with using the **Sequential** model from **keras.models** library

How was it trained?
We used a method similar to **Markov chains** which uses the idea of prediction of (say) the 61^{st} value using the previous 60 values and trains the data in this manner and uses the same trained model to predict the data for the last 60 days of the given data for any cryptocurrency, to match the close price to that of the actual data.

This required the use of recurrent neural networks like **LSTM (Long Short Term Memory)** and **Dense** to create layers and make a model.

We used the **Adam optimization** which is a **stochastic gradient descent method** which is based on adaptive estimation of first-order and second-order moments.

For error analysis (loss), we simply used the **RMSE** error check i.e Root Mean Square Error

We predicted the values using this model and inverse transformed the predicted values so as to get the actual predicted price rather than in range 0-1 (normalised).

	Close Prices (Actual)	Date	Close Prices (Predictions)
2393	8577.98	2019-11-17 23:59:59	8259.73
2394	8309.29	2019-11-18 23:59:59	8213.09
2395	8206.15	2019-11-19 23:59:59	8145.72
2396	8027.27	2019-11-20 23:59:59	8065.04
2397	7642.75	2019-11-21 23:59:59	7966.29
...	...	...	...
2986	33897.05	2021-07-02 23:59:59	31754.39
2987	34668.55	2021-07-03 23:59:59	31758.98
2988	35287.78	2021-07-04 23:59:59	31829.27
2989	33746.00	2021-07-05 23:59:59	31999.98
2990	34235.19	2021-07-06 23:59:59	31963.36

598 rows × 3 columns

Fig. 16.
Predicted v/s Real Prices for Bitcoin

	Close Prices (Actual)	Date	Close Prices (Predictions)
1728	214.22	2020-05-01 23:59:59	230.84
1729	215.33	2020-05-02 23:59:59	234.56
1730	210.93	2020-05-03 23:59:59	237.89
1731	208.17	2020-05-04 23:59:59	239.95
1732	206.77	2020-05-05 23:59:59	240.71
...	...	...	...
2155	2150.04	2021-07-02 23:59:59	2173.38
2156	2226.11	2021-07-03 23:59:59	2202.89
2157	2321.72	2021-07-04 23:59:59	2233.71
2158	2198.58	2021-07-05 23:59:59	2275.30
2159	2324.68	2021-07-06 23:59:59	2293.00

432 rows × 3 columns

Fig. 17.
Predicted v/s Real Prices for Ethereum

	Close Prices (Actual)	Date	Close Prices (Predictions)
2393	59.57	2019-11-17 23:59:59	60.44
2394	56.80	2019-11-18 23:59:59	60.00
2395	55.62	2019-11-19 23:59:59	59.39
2396	55.29	2019-11-20 23:59:59	58.61
2397	50.88	2019-11-21 23:59:59	57.81
...	...	...	...
2986	136.94	2021-07-02 23:59:59	129.98
2987	140.28	2021-07-03 23:59:59	131.20
2988	144.91	2021-07-04 23:59:59	132.18
2989	138.07	2021-07-05 23:59:59	133.55
2990	138.99	2021-07-06 23:59:59	133.87

598 rows × 3 columns

Fig. 18.
Predicted v/s Real Prices for Litecoin

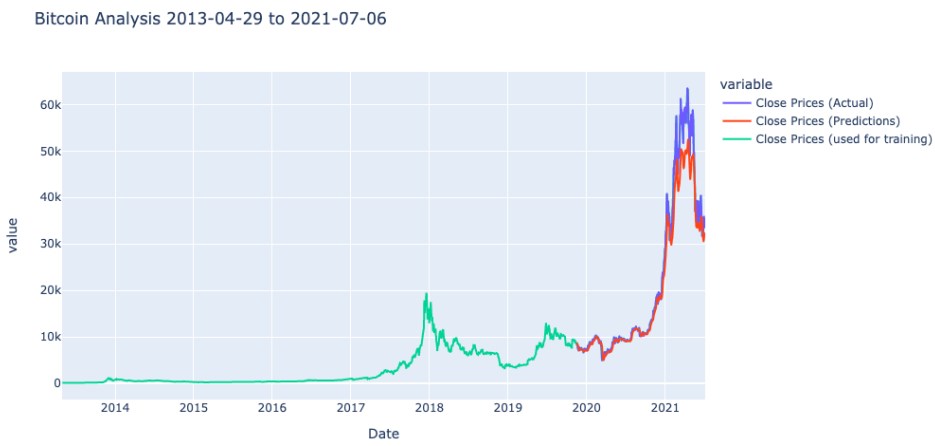

Fig. 19.
Predicted v/s Real Prices for Bitcoin

Fig. 20.
Predicted v/s Real Prices for Ethereum

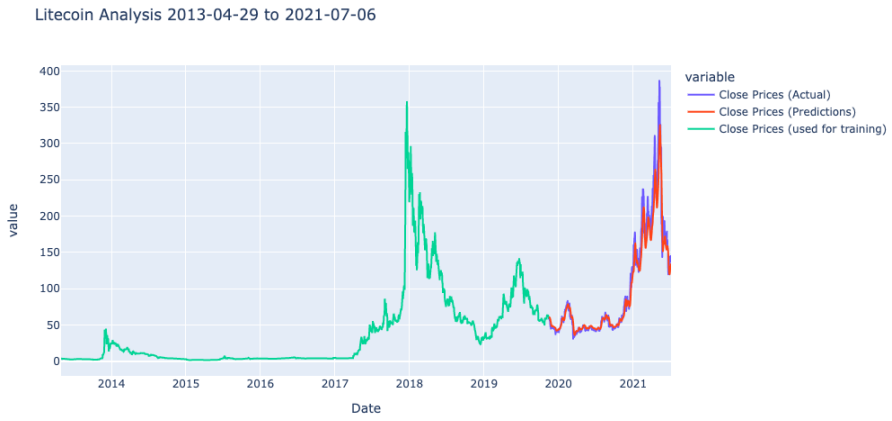

Fig. 21.
Predicted v/s Real Prices for Litecoin

REFERENCES

[1] community.plotly.com
[2] stackoverflow.com
[3] keras.io (for ML part)
[4] kaggle.com (for dataset and analysis ideas)
[5] coinbase.com (for cryptocurrency basics)
[6] https://www.kaspersky.com/resource-center/definitions/what-is-cryptocurrencykaspersky.com (for cryptocurrency basics)
[7] investing.com (for datasets)