



Crypto Analysis and Time Series Prediction

Aiswarya Kale

Naga Sireesha Chinthala

Data 690: Financial Data Science

Prof Abdullah Karasan

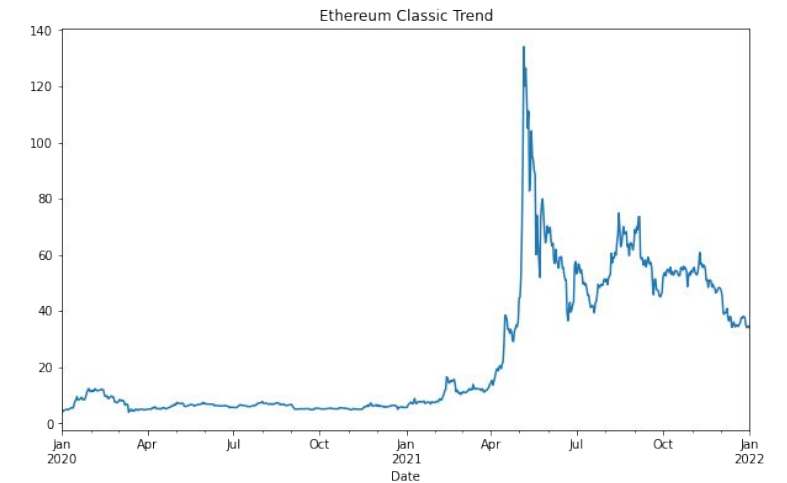
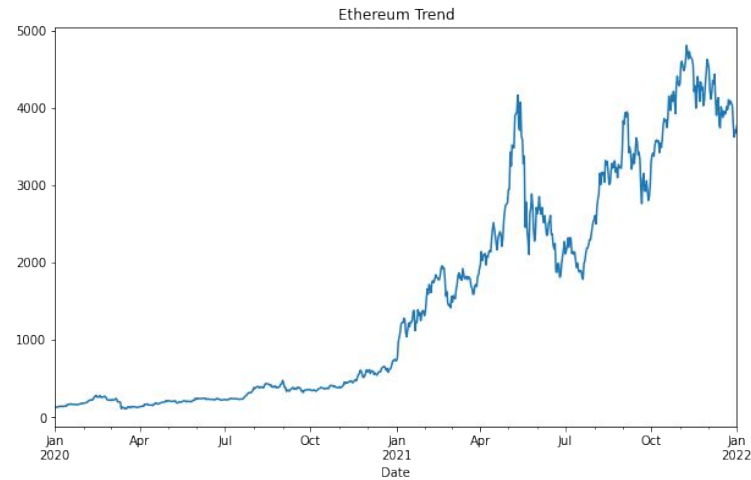
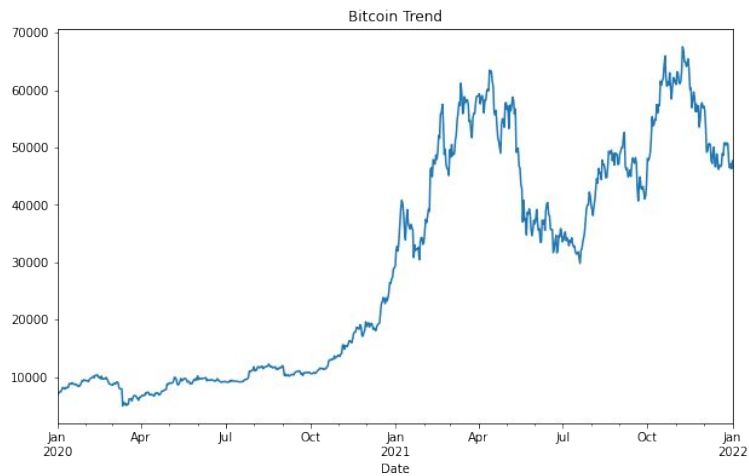
Introduction

- While the world was (and still is) in the grip of a pandemic in 2020, crypto's have had a stellar run. Especially, Bitcoin From its spectacular crash in 2018 and a relatively uneventful year in 2019, it started 2020 with a price of USD\$8,000 to an all-new record high of USD\$65,000. That's roughly more than an 800% return! It is one of the most successful investments of 2020.
- Ethereum and Ethereum classic also followed similar patterns during the pandemic.
- We will analyze trends of three popular crypto's Bitcoin, Ethereum, and Ethereum classic.
- We will predict the trend and price using the time series ARIMA model and RNN model LSTM.
- The data has been downloaded from yfinance API for all the three crypto's and closing price is considered throughout the analysis.

Exploratory Data Analysis

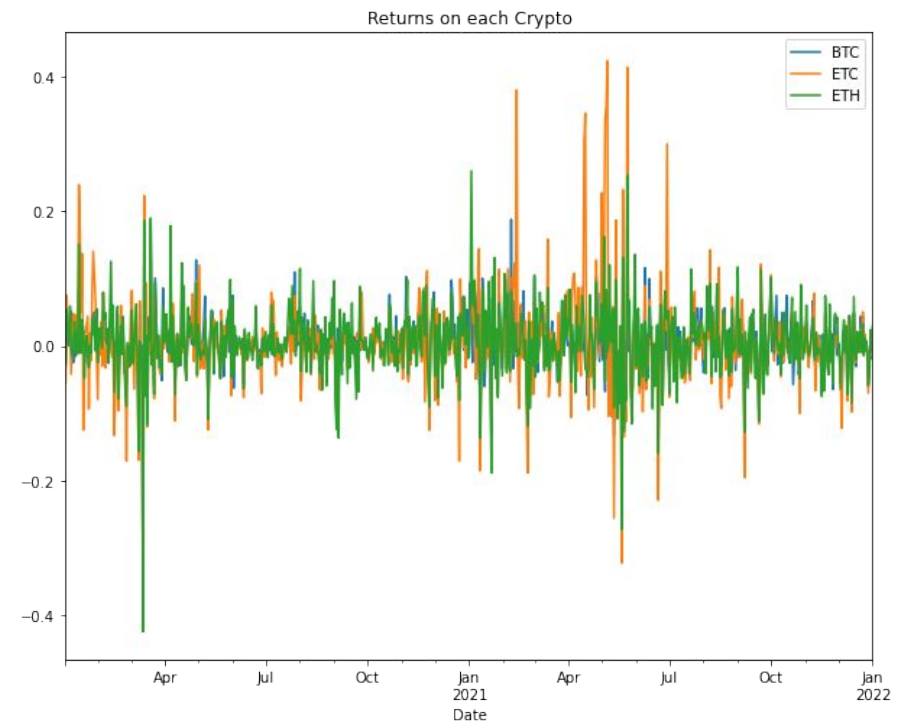
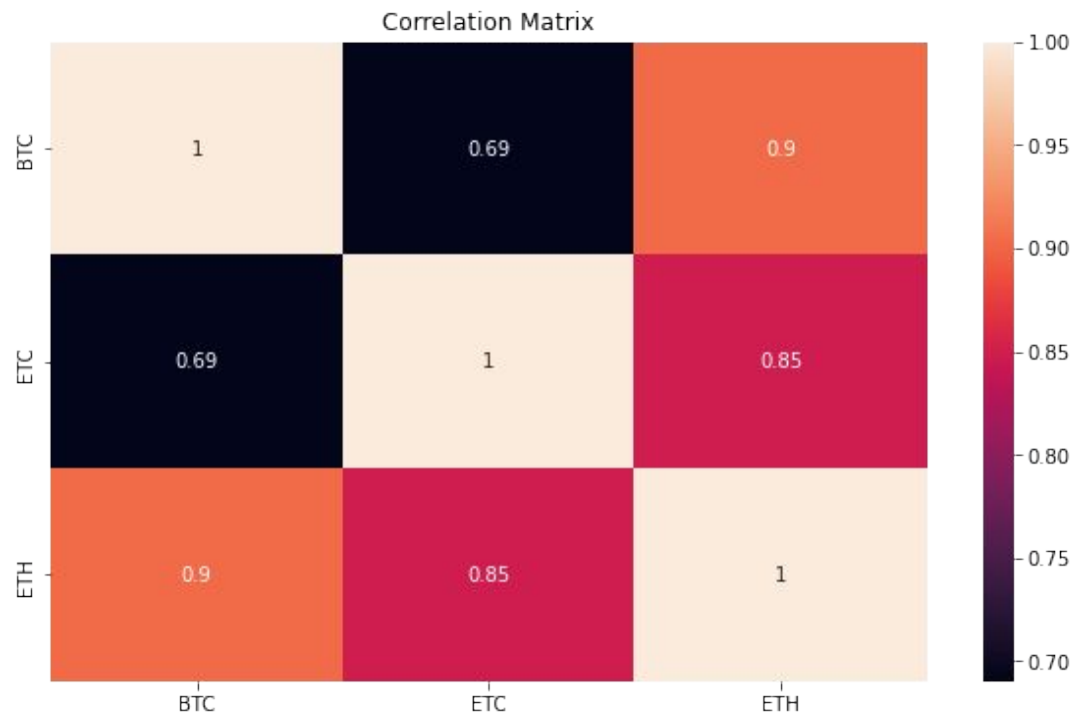
Bitcoin and Ethereum have an upward trend. Ethereum Classic also has a slight upward trend, but there was a sudden spike and drop in April 2021.

Data does not look stationary from the trend.



Exploratory Data Analysis (Cont..)

- All three cryptos has a very high positive correlation. Bitcoin and Ethereum are very highly correlated.
- Returns on each of these cryptos followed similar patterns and almost overlapping.



Exploratory Data Analysis (Cont..)

- To understand the correlation between these cryptos we have used the OLS stats model. The model results suggest that the changes in one crypto are considerably impacting the other.
- Based on the model results, Ethereum explains 63% of the variance in Bitcoin price changes and Ethereum Classic explains 46% of the variance in Ethereum price.
- We can say that these three cryptos are highly cointegrated.

OLS Regression Results

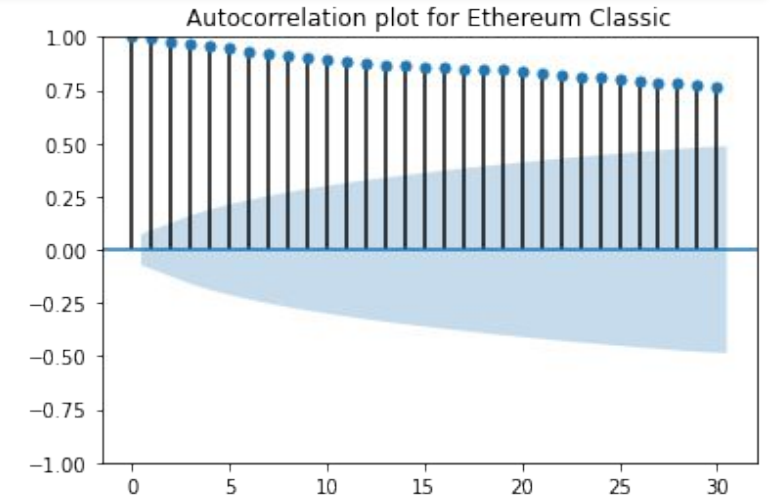
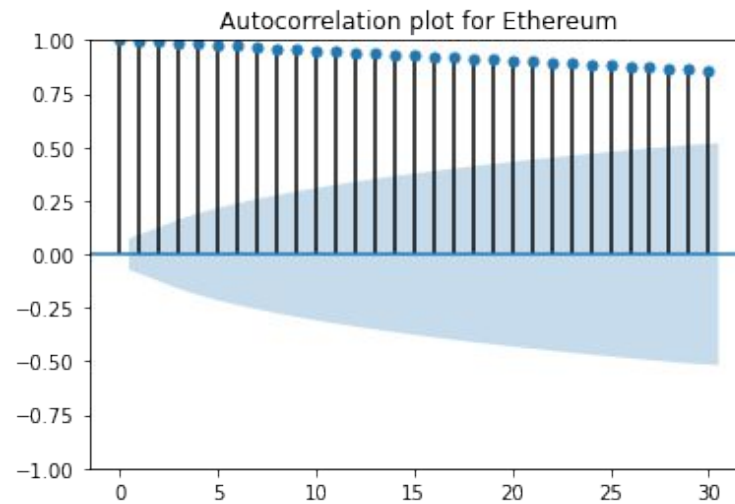
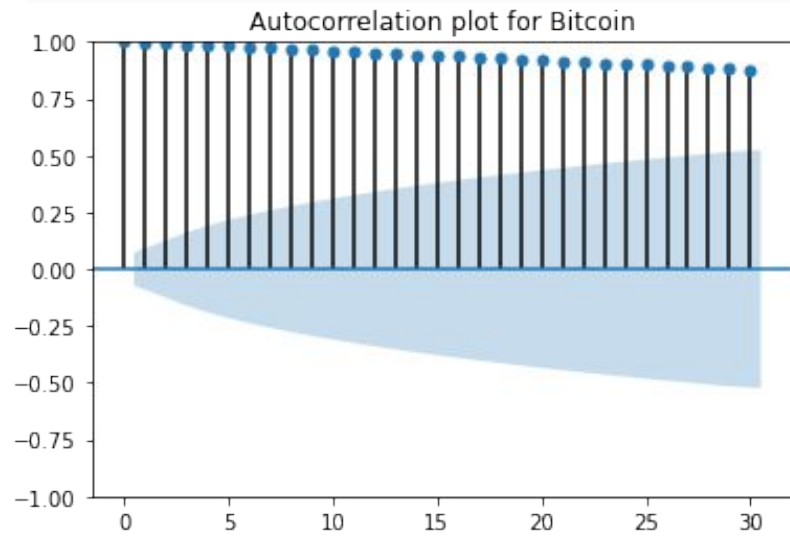
Dep. Variable:	BTC	R-squared:	0.632	
Model:	OLS	Adj. R-squared:	0.632	
Method:	Least Squares	F-statistic:	1253.	
Date:	Mon, 28 Mar 2022	Prob (F-statistic):	1.61e-160	
Time:	17:35:56	Log-Likelihood:	1682.6	
No. Observations:	731	AIC:	-3361.	
Df Residuals:	729	BIC:	-3352.	
Df Model:	1			
Covariance Type: nonrobust				
	coef	std err	t P> t [0.025 0.975]	
ETH	0.6022	0.017	35.403	0.000 0.569 0.636
const	-0.0002	0.001	-0.251	0.802 -0.002 0.002
Omnibus:	79.649	Durbin-Watson:	1.952	
Prob(Omnibus):	0.000	Jarque-Bera (JB):	587.684	
Skew:	0.046	Prob(JB):	2.43e-128	
Kurtosis:	7.392	Cond. No.	19.0	

OLS Regression Results

Dep. Variable:	ETH	R-squared:	0.457	
Model:	OLS	Adj. R-squared:	0.456	
Method:	Least Squares	F-statistic:	612.4	
Date:	Mon, 28 Mar 2022	Prob (F-statistic):	1.28e-98	
Time:	17:35:56	Log-Likelihood:	1336.7	
No. Observations:	731	AIC:	-2669.	
Df Residuals:	729	BIC:	-2660.	
Df Model:	1			
Covariance Type: nonrobust				
	coef	std err	t P> t [0.025 0.975]	
ETC	0.5217	0.021	24.747	0.000 0.480 0.563
const	0.0034	0.001	2.354	0.019 0.001 0.006
Omnibus:	222.314	Durbin-Watson:	1.887	
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1983.313	
Skew:	-1.100	Prob(JB):	0.00	
Kurtosis:	10.764	Cond. No.	14.6	

Stationarity Check

- Initial analysis suggests that the data is not stationary. We have used Auto-Correlation, Partial Auto-Correlation, and Augmented Dickey-Fuller test (ADF) to check the stationarity. Results of all three tests confirm that the data is not stationary.



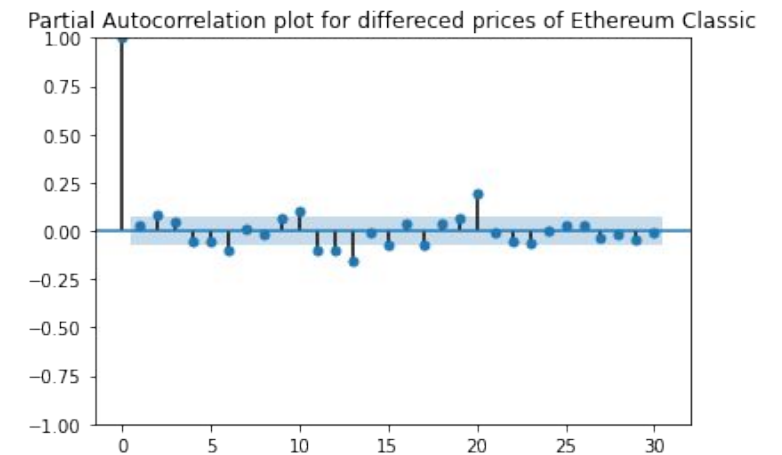
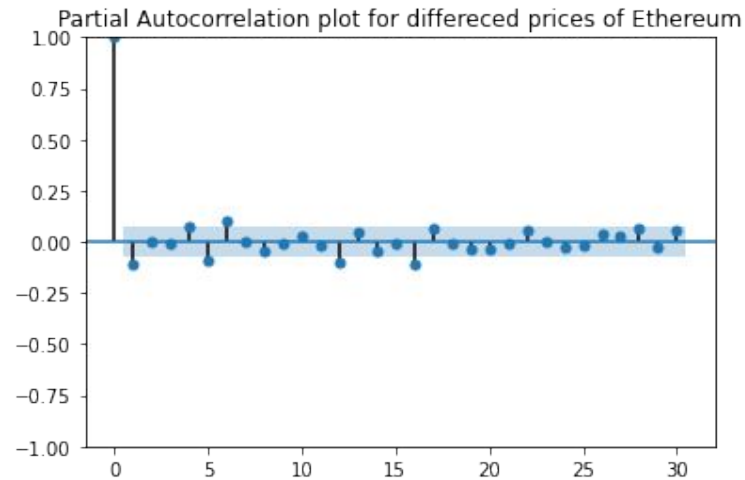
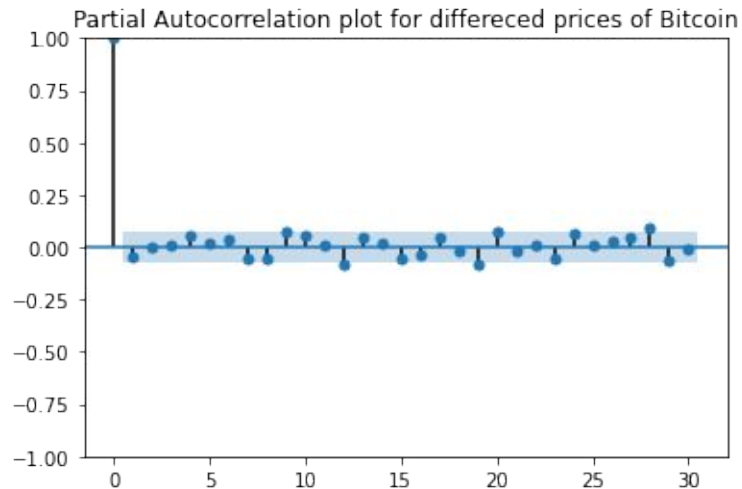
The p-value of ADF test for BTC 0.073

The p-value of ADF test for ETH 0.91

The p-value of ADF test for ETC 0.38

Stationarity Check

- To make the data stationary, we took the differenced prices of stocks.
- And we again verified the stationarity on differenced data using Auto-Correlation, Partial Auto-Correlation, and Augmented Dickey-Fuller test (ADF). Results of all three tests confirm that the differenced data is stationary.



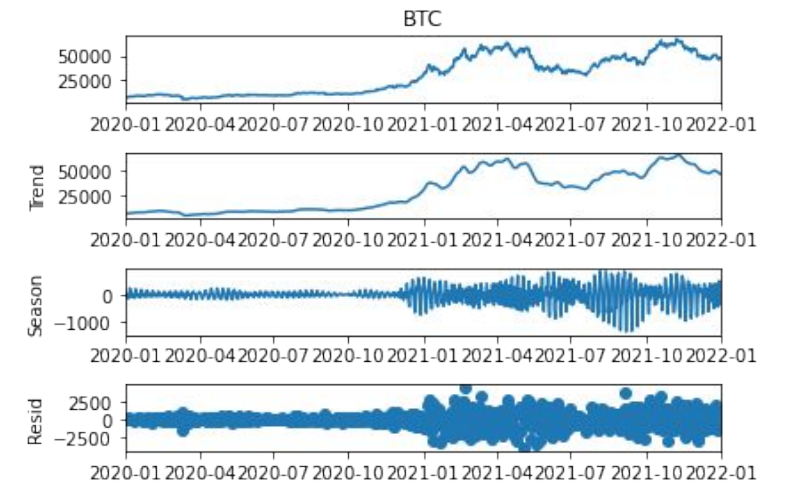
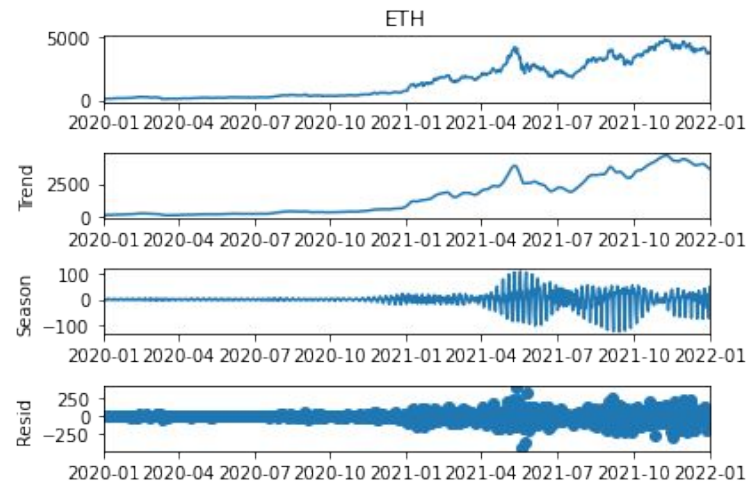
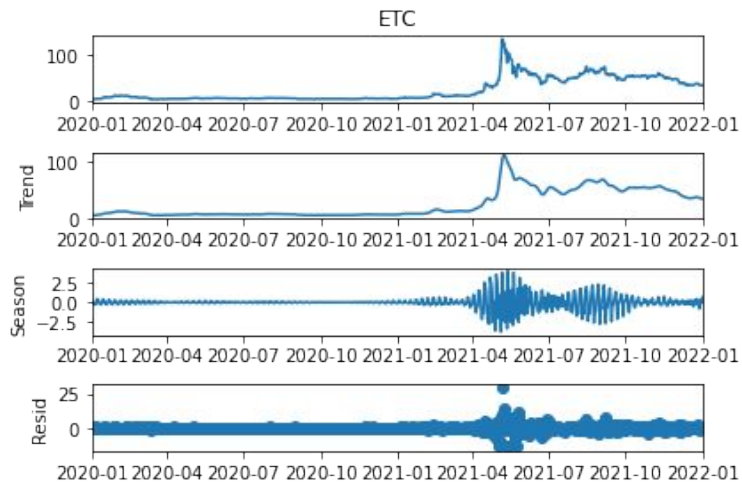
The p-value of ADF test for BTC 0.0

The p-value of ADF test for ETH 0.0

The p-value of ADF test for ETC 0.0

Seasonality Check

There is no seasonal behavior observed in any crypto prices.



Prediction Using ARIMA

- ARIMA (Autoregressive Integrated Moving Average Model) takes 3 parameters
 - p**: The number of lag observations included in the model also called the lag order.
 - d**: The number of times that the raw observations are differenced, also called the degree of difference.
 - q**: The size of the moving average window, also called the order of moving average.
- These parameters are selected based on low AIC values.
- The model has been run for all the three cryptos and forecasted values for out of sample data.

ARIMA Results - Bitcoin

- Parameters based on low AIC are : $p = 4, d = 2, q = 3$
- Based on p-values except for auto regressor lag 3, all the remaining lag coefficients are statistically significant.
- The out-of-sample forecast showed an upward trend, meaning it is a good idea to invest in Bitcoin.

SARIMAX Results

Dep. Variable: BTC No. Observations: 585

Model: ARIMA(4, 2, 3) Log Likelihood: -4978.795

Date: Mon, 28 Mar 2022 AIC: 9973.590

Time: 17:37:40 BIC: 10008.535

Sample: 01-01-2020 HQIC: 9987.211
- 08-07-2021

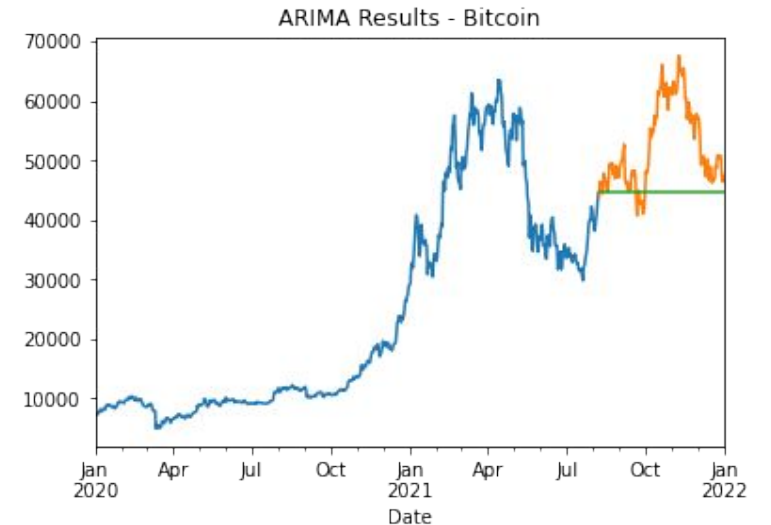
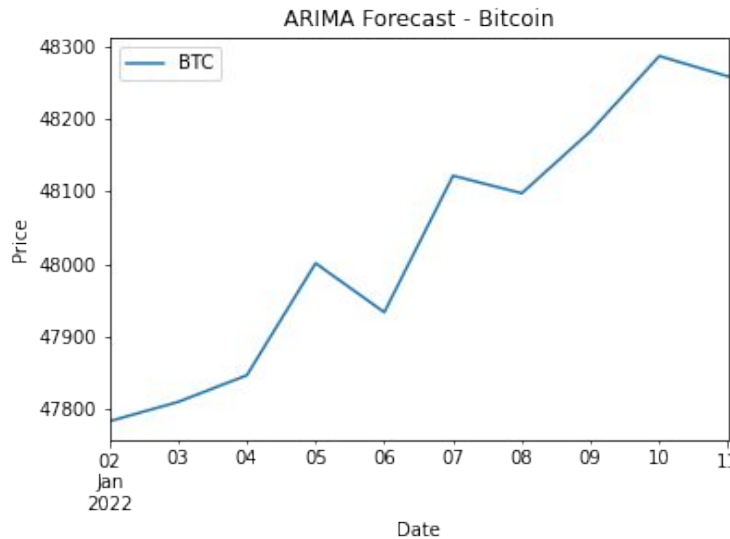
Covariance Type: opg

	coef	std err	z	P> z	[0.025	0.975]
ar.L1	-0.5007	0.037	-13.602	0.000	-0.573	-0.429
ar.L2	-0.8775	0.039	-22.332	0.000	-0.954	-0.800
ar.L3	0.0564	0.039	1.457	0.145	-0.019	0.132
ar.L4	0.0881	0.033	2.661	0.008	0.023	0.153
ma.L1	-0.5468	0.017	-31.590	0.000	-0.581	-0.513
ma.L2	0.5170	0.020	25.343	0.000	0.477	0.557
ma.L3	-0.9662	0.021	-46.734	0.000	-1.007	-0.926

sigma2 1.691e+06 6.01e+04 28.149 0.000 1.57e+06 1.81e+06

Ljung-Box (L1) (Q): 0.00 Jarque-Bera (JB): 893.96
Prob(Q): 0.95 Prob(JB): 0.00

Heteroskedasticity (H): 22.20 Skew: -0.11
Prob(H) (two-sided): 0.00 Kurtosis: 9.06



ARIMA Results - Ethereum

- Parameters: $p = 3$, $d = 2$, $q = 5$
- All the lag coefficients are statistically significant
- Ethereum also showed upward trend for out of sample forecast.

SARIMAX Results

Dep. Variable: ETH No. Observations: 585

Model: ARIMA(3, 2, 5) Log Likelihood: -3415.246

Date: Mon, 28 Mar 2022 AIC: 6848.492

Time: 17:39:01 BIC: 6887.806

Sample: 01-01-2020 HQIC: 6863.816

- 08-07-2021

Covariance Type: opg

	coef	std err	z	P> z	[0.025	0.975]
ar.L1	-1.2642	0.045	-28.397	0.000	-1.351	-1.177
ar.L2	-1.2304	0.051	-24.044	0.000	-1.331	-1.130
ar.L3	-0.6580	0.043	-15.283	0.000	-0.742	-0.574
ma.L1	0.1550	0.057	2.740	0.006	0.044	0.266
ma.L2	0.1683	0.042	3.978	0.000	0.085	0.251
ma.L3	-0.5066	0.048	-10.547	0.000	-0.601	-0.412
ma.L4	-0.6103	0.052	-11.716	0.000	-0.712	-0.508
ma.L5	-0.1933	0.028	-6.830	0.000	-0.249	-0.138

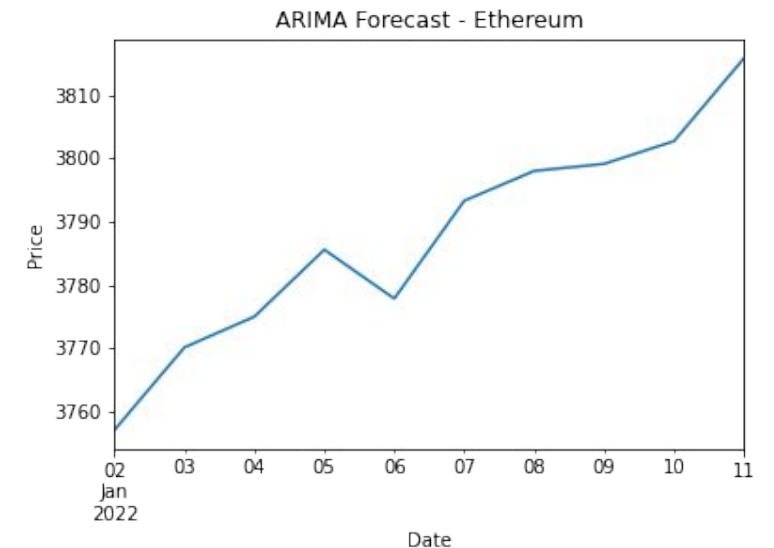
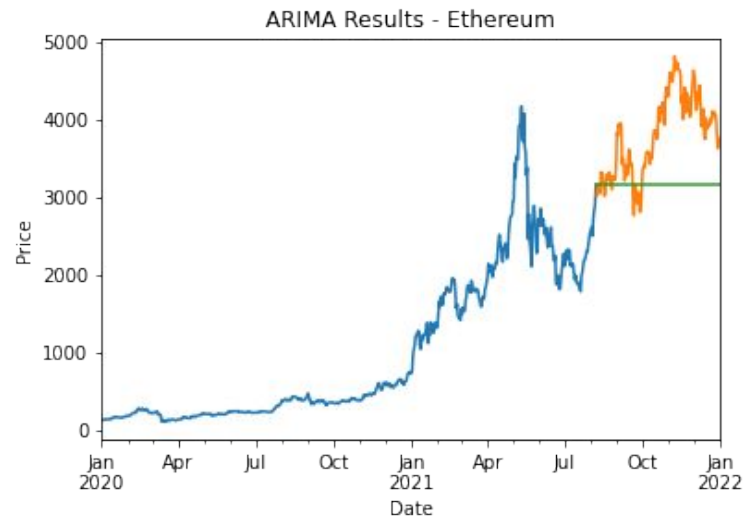
sigma2 7899.0684 205.195 38.495 0.000 7496.894 8301.242

Ljung-Box (L1) (Q): 0.01 Jarque-Bera (JB): 8038.24

Prob(Q): 0.93 Prob(JB): 0.00

Heteroskedasticity (H): 150.93 Skew: -1.03

Prob(H) (two-sided): 0.00 Kurtosis: 21.07



ARIMA Results – Ethereum Classic

- Parameters: $p = 5$, $d = 1$, $q = 5$
- Except for the moving average lag 5, all other coefficients are statistically significant
- Ethereum Classic has shown a downward trend for out-of-sample forecast.

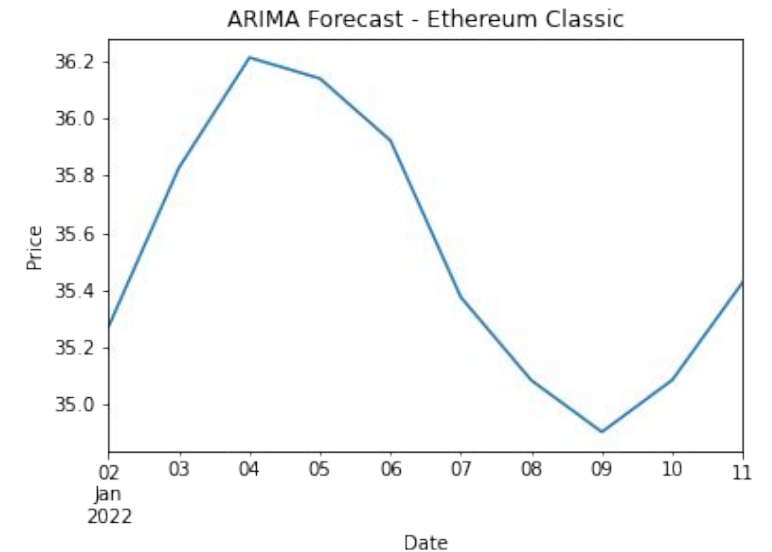
SARIMAX Results

Dep. Variable: ETC No. Observations: 585
Model: ARIMA(5, 1, 5) Log Likelihood: -1530.747
Date: Mon, 28 Mar 2022 AIC: 3083.494
Time: 23:43:16 BIC: 3131.563
Sample: 01-01-2020 HQIC: 3102.229
- 08-07-2021

Covariance Type: opg

	coef	std err	z	P> z	[0.025	0.975]
ar.L1	-0.5884	0.115	-5.105	0.000	-0.814	-0.363
ar.L2	-0.2478	0.120	-2.068	0.039	-0.483	-0.013
ar.L3	-0.6209	0.067	-9.282	0.000	-0.752	-0.490
ar.L4	-0.3827	0.092	-4.169	0.000	-0.563	-0.203
ar.L5	0.2954	0.091	3.244	0.001	0.117	0.474
ma.L1	0.6010	0.124	4.846	0.000	0.358	0.844
ma.L2	0.3878	0.117	3.329	0.001	0.159	0.616
ma.L3	0.8605	0.073	11.843	0.000	0.718	1.003
ma.L4	0.5780	0.122	4.754	0.000	0.340	0.816
ma.L5	-0.1360	0.115	-1.186	0.235	-0.361	0.089
sigma2	11.0475	0.262	42.116	0.000	10.533	11.562

Ljung-Box (L1) (Q): 0.09 Jarque-Bera (JB): 40562.76
Prob(Q): 0.76 Prob(JB): 0.00
Heteroskedasticity (H): 98.04 Skew: 1.68
Prob(H) (two-sided): 0.00 Kurtosis: 43.69

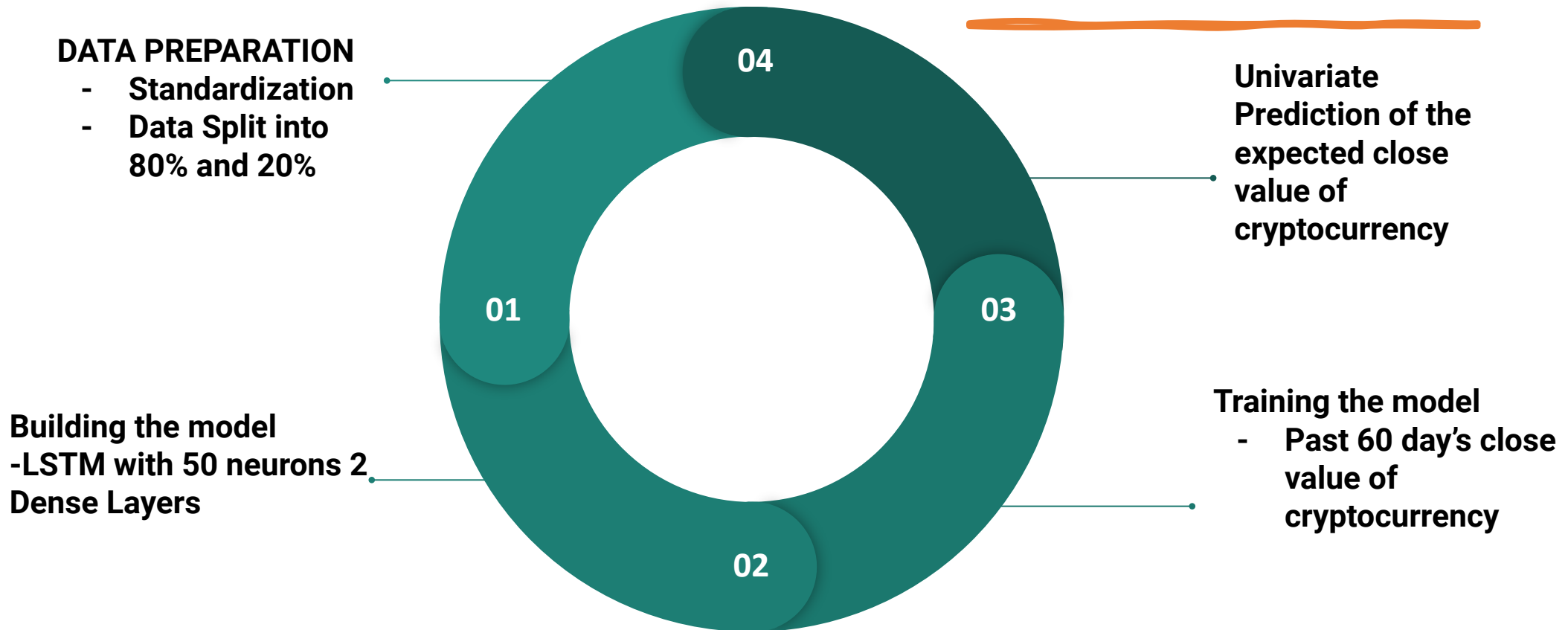




Prediction Using LSTM

- LSTM performs very well for sequential data such as time series. RNNs are able to form a much deeper understanding of a sequence and its context compared to other algorithms.
- In an RNN, the information goes through a cycle. When making a decision, it considers the current input and also what it has learned from the inputs it has received previously.
- In an LSTM you have three gates: input, forget and output gate. These gates determine whether or not to let new input in (input gate), delete the information because it isn't important (forget gate), or let it impact the output at the current timestep (output gate).
- Data is divided into 80% training and 20% testing.

Methodology



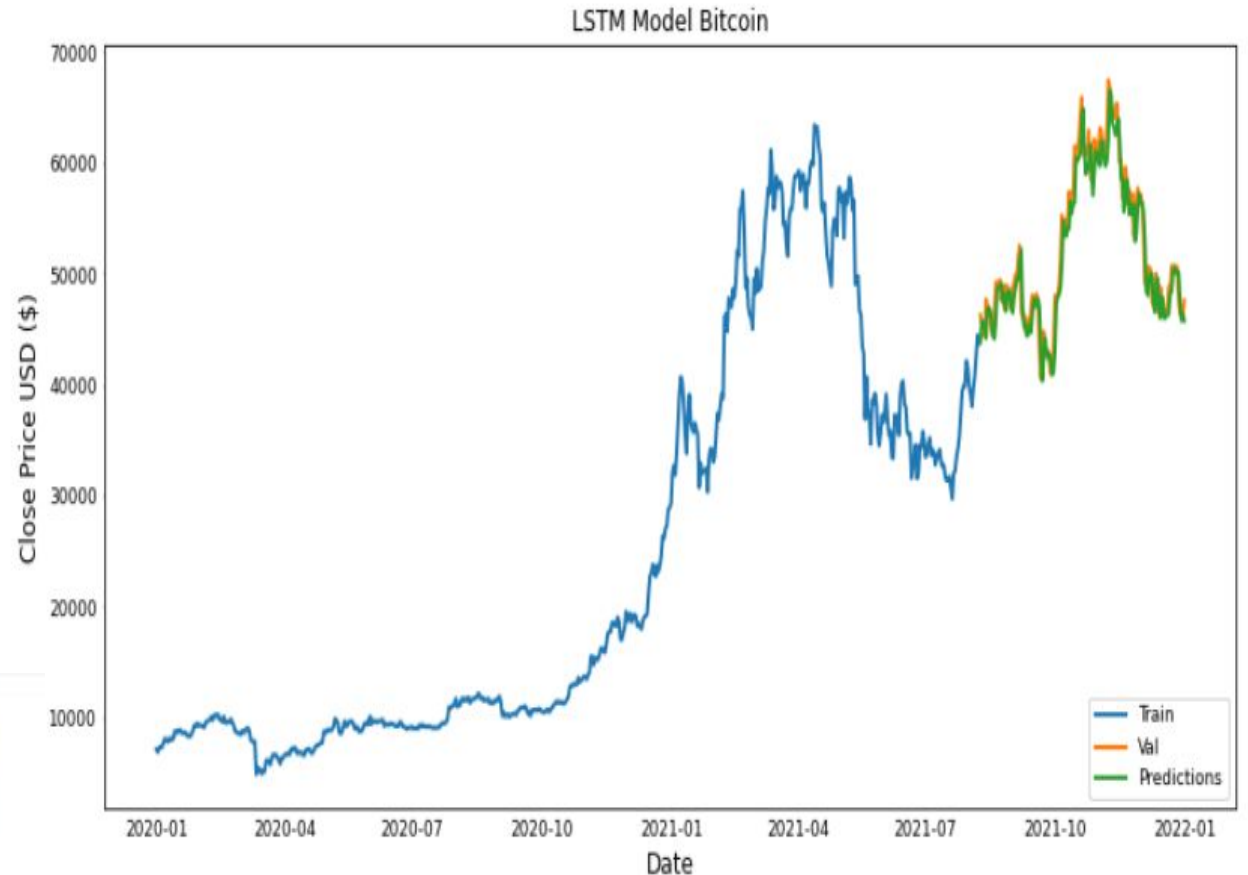
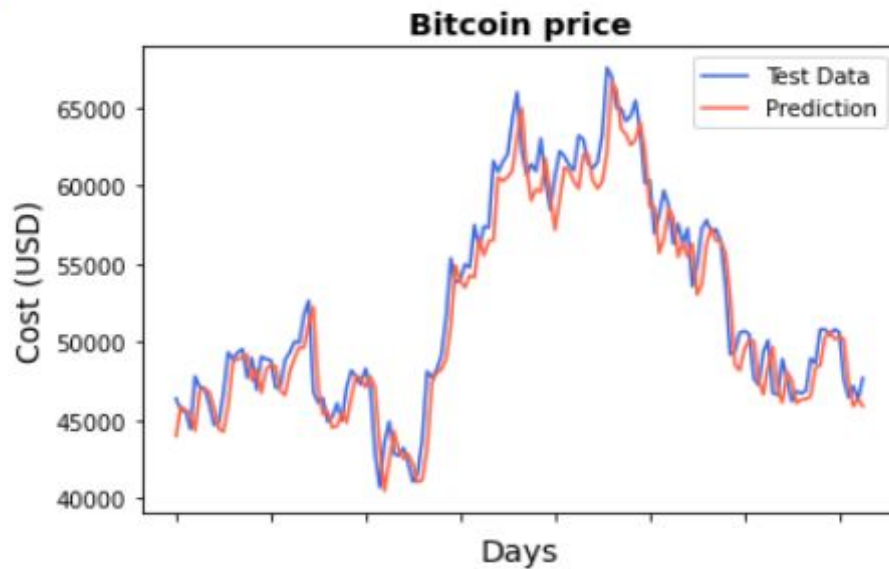
LSTM Results – Bitcoin

RMSE: 1944.0829271354098

MAE: 1516.2626016695206

R2 score: 0.9198929474919921

[]



```
[*****100%*****] 1 of 1 completed  
[*****100%*****] 1 of 1 completed
```

Prediction close price at 3-29-2022: \$[47412.38]

Actual price at 3-29-2022: \$47465.73046875

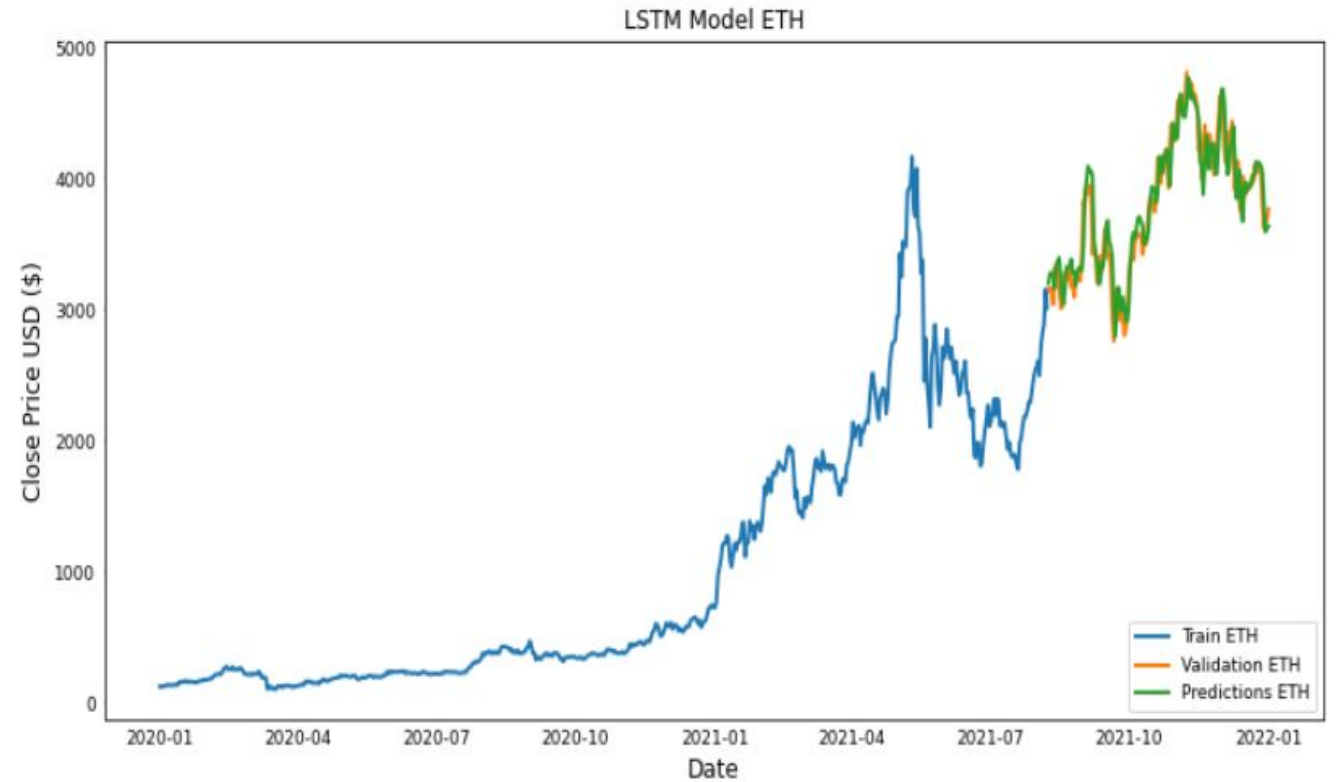
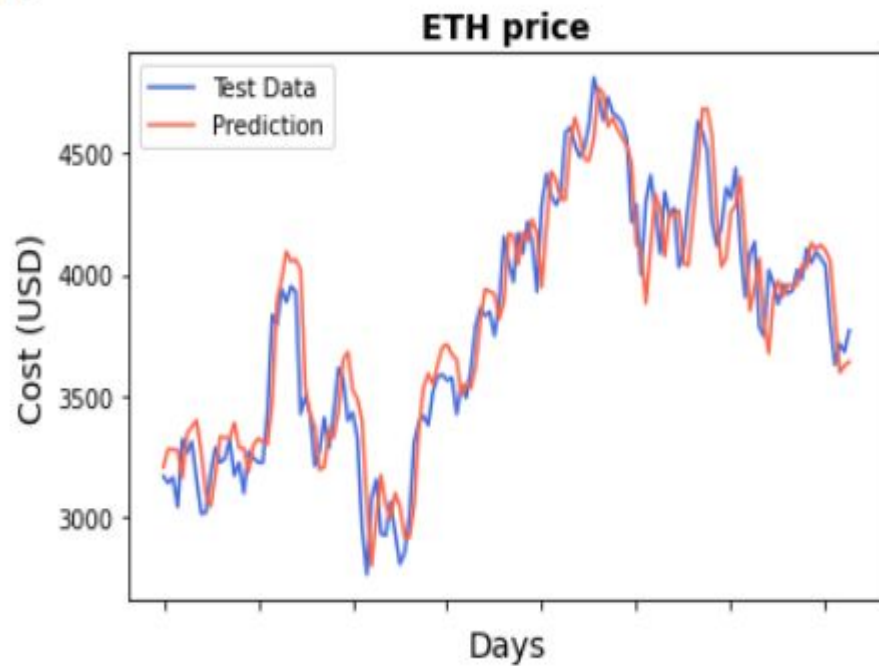
LSTM Results – Ethereum

RMSE: 171.84604520273467

MAE: 140.30372097067635

R2 score: 0.8890275882520761

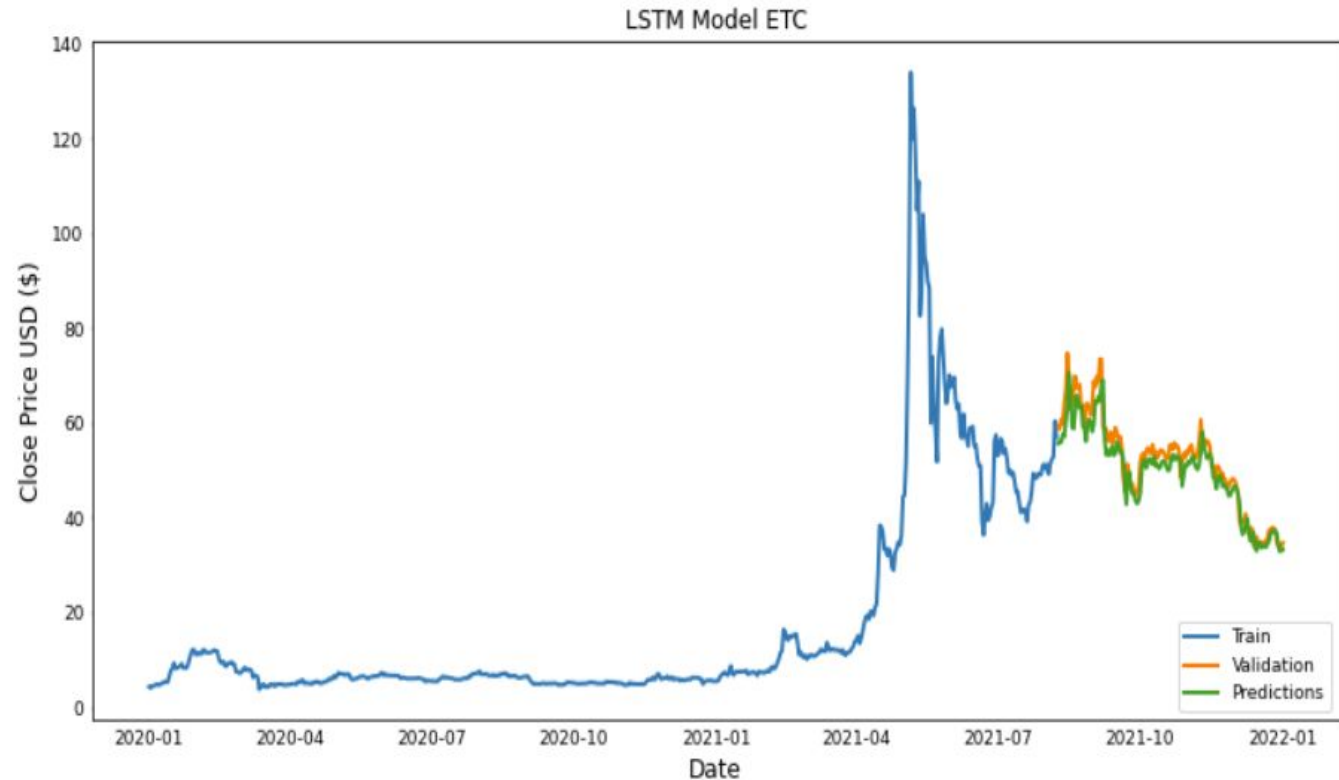
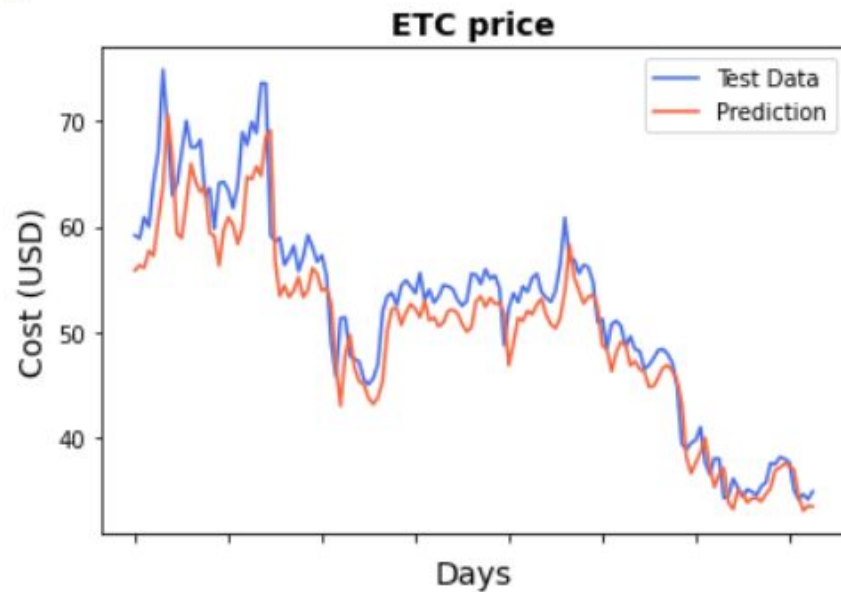
[]



LSTM Results – Ethereum Classic

RMSE: 3.5017023632002706
MSE: 12.26191944044236
MAE: 2.7699289975101
R2 score: 0.8759571893895505

LJ



```
[*****100%*****] 1 of 1 completed
[*****100%*****] 1 of 1 completed
Prediction close price at 3-29-2022: $[3305.8403]
Actual price at 3-29-2022: $3401.98779296875
```

Conclusion

- All the cryptos are highly cointegrated.
- Both models performed performed certainly well.
- Bitcoin and Ethereum showed an upward trend for forecasted data, meaning it is good idea to invest on them.
- Ethereum Classic has not showed certain upward trend.



References

- <https://machinelearningmastery.com/arima-for-time-series-forecasting-with-python/>
- <https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/>
- <https://www.analyticsvidhya.com/blog/2017/12/fundamentals-of-deep-learning-introduction-to-lstm/>
- <https://medium.com/analytics-vidhya/time-series-analysis-of-cryptocurrencies-using-deep-learning-fb-prophet-48abdb2e7ebf>
- <https://machinelearningmastery.com/time-series-prediction-lstm-recurrent-neural-networks-python-keras/>

Queries..?



THANK YOU

