

Progressive Education Society's
Modern College of Arts, Science and Commerce (Autonomous)
Shivajinagar, Pune – 411005

Project Title

**“CRYPTOCURRENCY PAIR FORECASTING & ANALYSIS BY USING
TIME SERIES MODELS & ML ALGORITHMS WITH INTRODUCTION TO
BLOCKCHAIN TECHNOLOGY”**

by

Aishwarya Sachine Mehendale

Abhijeet Balasaheb Talole

Akshata Gangadhar Musale

Guided by

Assistant Prof. Sarika Khirid

An Abstract of The
**“CRYPTOCURRENCY PAIR FORECASTING & ANALYSIS BY USING TIME SERIES MODELS & ML ALGORITHMS
WITH INTRODUCTION TO BLOCKCHAIN TECHNOLOGY”**

Submitted to the **Statistics Department**
as partial fulfillment of **P-305**
of the requirements for the **MSc. Statistics PG Degree**
in ***Progressive Education Society's Modern College of Arts, Science and Commerce (Autonomous)***
Shivajinagar, Pune – 411005,

The goal of this study is to predict prices for Cryptocurrencies using Time series analysis and machine learning techniques. The purpose of this project is to take a sneak peek into the future by **forecasting the next 30 days' average daily Realized Volatility (RV) of ETH-BTC** using 2 different approaches - the traditional econometric approach to volatility prediction of financial time series **GARCH** and state-of-the-art **LSTM Neural Networks**.

Quantitative research methodology was used in this study and the Khiri dataset Consist the historical data values of any Khiri crypto-pair such as Open/Close/High/Low prices of any interval such as 15-minutes, Hourly, 1-day interval weekly, monthly. Dataset were obtained using the Binance API .

INTRODUCTION

In recent times, the universe of virtual market trading has encountered impressive expansion. Online trading platforms have changed the manner by which speculators inquire about and execute orders. The idea of algorithmic trading presents the next phase of this development, and one that is as of now affecting entities of high volume and liquidity.

In this virtual world, Cryptocurrencies can no longer be ignored. “Cryptocurrencies are digital, straightforward and simple to use in comparison with the traditional currencies”, The individual investor, the big cooperation and even governments (who seek ways to control their influence in their financial markets) are interested in them.

Bitcoin is one of the oldest and biggest cryptocurrencies being traded as of now, in terms of volume being traded it is so big that even now, with the advent of thousands of new cryptocurrencies (Altcoins), Bitcoin has market share of more than 45% as compared to other cryptocurrencies being followed by Ethereum at 8.75%. This says a lot about why Bitcoin and Ethereum might be really interesting and important stock to predict. Also, Their prices fluctuate heavily.

Cryptocurrency is one of the most volatile markets today and has gained a lot of attention from investors across the globe. Cryptocurrency, being a novel technique for transaction system, has led to a lot of confusion among the investors.

In India Government imposed 30% flat tax on cryptocurrency and include it in the high risk financial/digital asset instead of completely banning it in India and also setup a government body CBDC (Central Bank Digital Currency) regulated by RBI, to regulate Cryptocurrencies. As a Indian crypto trader considering cryptocurrency related taxation structure we need to reduced loses and take better decisions while booking profit from Crypto-trading or investing. This is one of the most important reasons which attracted us to make statistical analysis of Cryptocurrencies.

KEYWORDS

Cryptocurrency

Blockchain

Bitcoin

Altcoin

Etherium

Stablecoin

Binance API

Ticker

MOTIVATION

In India Government imposed 30% flat tax on cryptocurrency and include it in the high risk financial/digital asset instead of completely banning it in India and also setup a government body CBDC (Central Bank Digital Currency) regulated by RBI, to regulate Cryptocurrencies. As a Indian crypto trader considering cryptocurrency related taxation structure we need to reduced loses and take better decisions while booking profit from Crypto-treading or investing.

We know that Cryptocurrency is one of the most volatile markets today and has gained a lot of attention from investors across the globe.

Those are the most important reasons which attracted us to make statistical analysis of Cryptocurrencies.

OBJECTIVES

- Finding best trading strategy for Cryptocurrency Trading'
- Find best Crypto-pair for Binance Spot trading
- To Find appropriate model for forecasting of cryptocurrency Volatility.
- The goal of this study is to find the best Time series model and predict prices for Cryptocurrencies using Time series analysis and machine learning techniques.
- The purpose of this project is to take a sneak peek into the future by **forecasting the next 7 days' average daily Realized Volatility (RV) of ETH-BTC** using 2 different approaches - the traditional econometric approach to volatility prediction of financial time series **GARCH** and state-of-the-art **LSTM Neural Networks**.

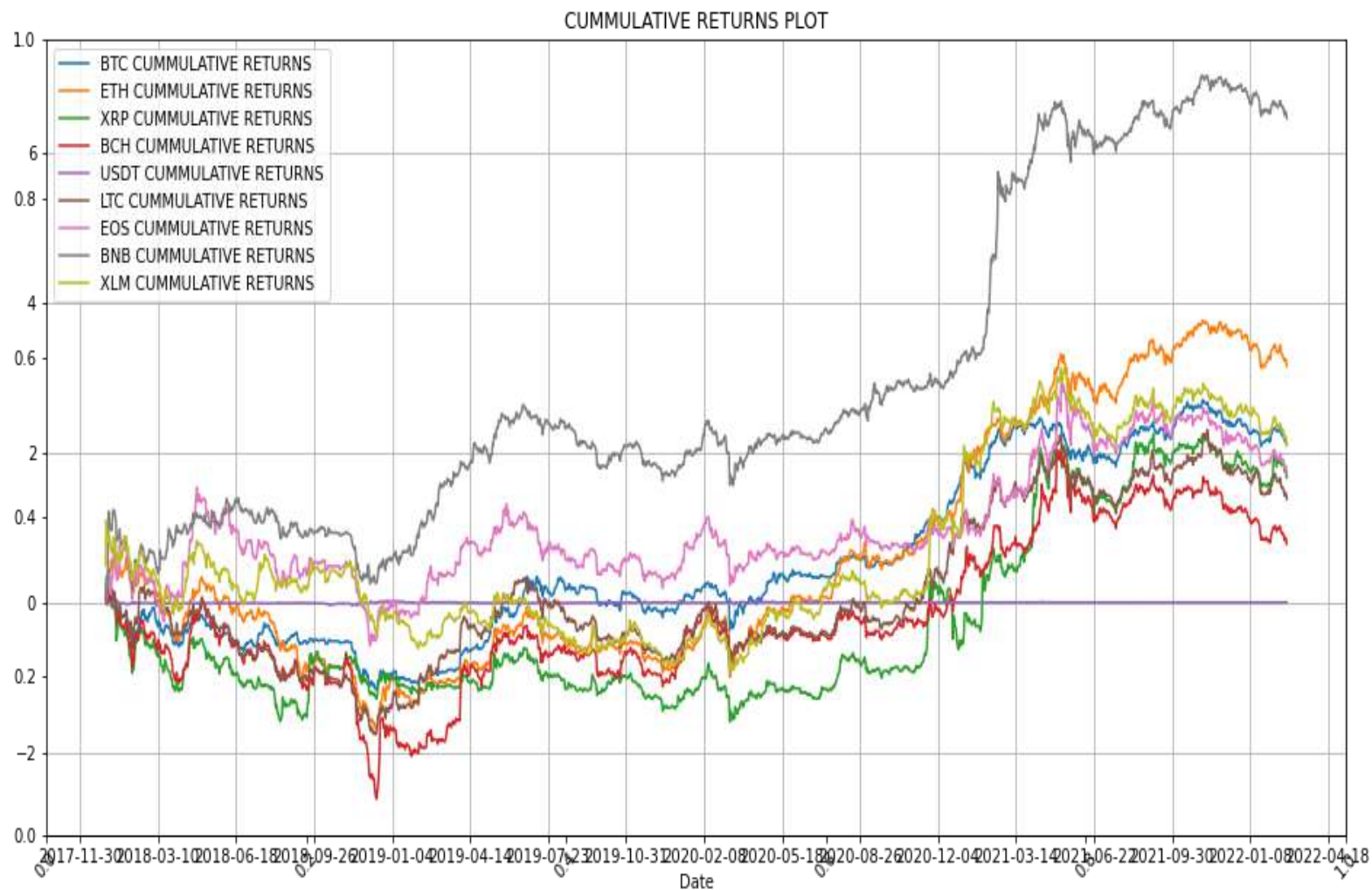
Additionally, there will be two secondary objectives:

- Spread awareness about new Fintech term Cryptocurrency
- Introduce the blockchain technology

STATISTICAL TOOLS USED FOR DATA ANALYSIS

- Normalization
- Exploratory Data Analysis
- Time Series Analysis:
- Baseline Models
 - Mean Baseline Model
 - Random Walk Naïve Forecasting
- GARCH
 - GJR-GARCH
 - TARCH
 - Simulation
 - Bootstrap
- Machine Learning
- Neural Network
 - LSTM

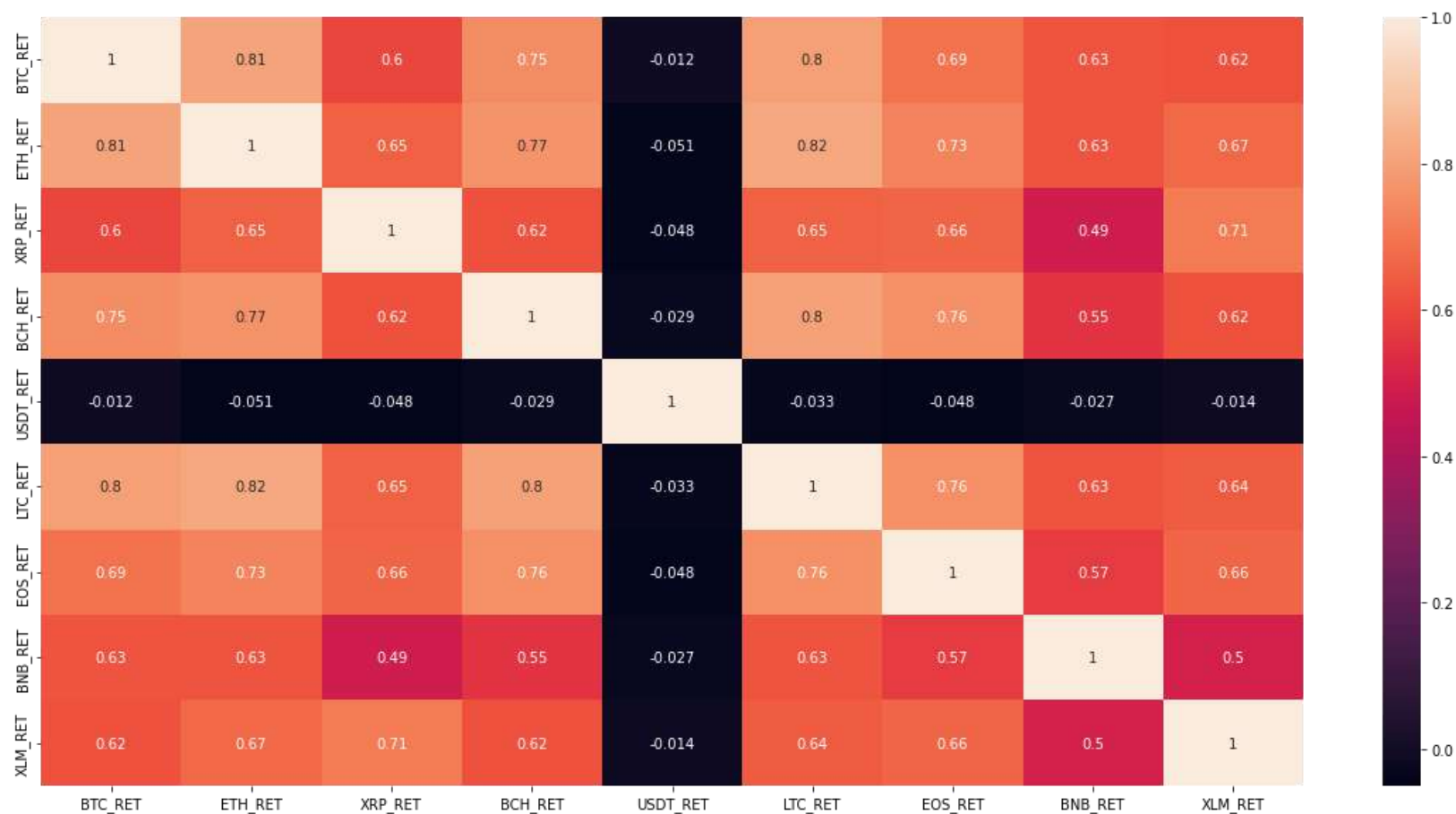
CUMMULATIVE RETURNS PLOT



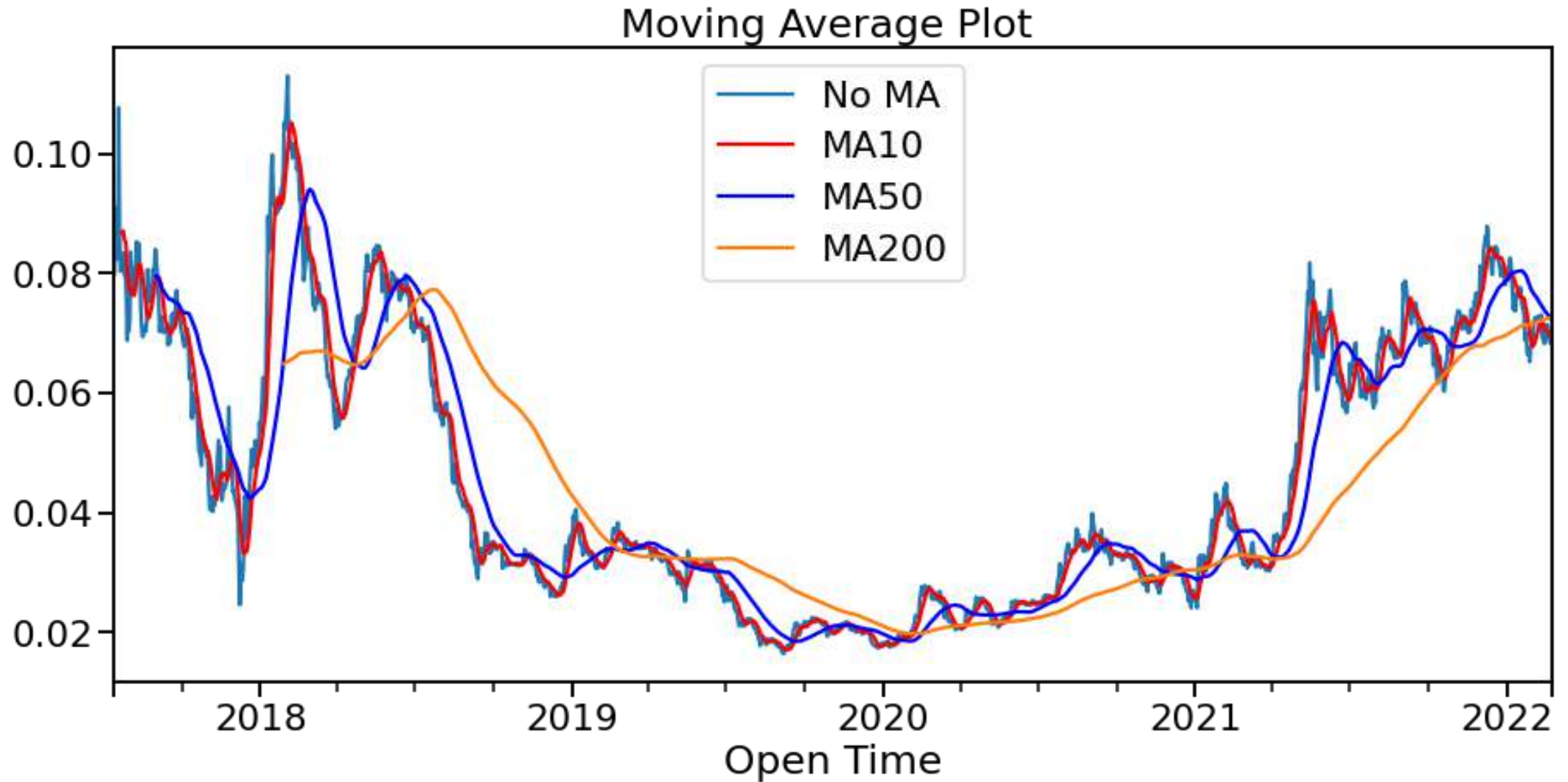
COINTEGRATION

- BTC_RET and ETH_RET: p-value = 0.0
- BTC_RET and XRP_RET: p-value = 2.470976042943848e-29
- BTC_RET and BCH_RET: p-value = 1.9958557257613114e-24
- BTC_RET and USDT_RET: p-value = 1.8460483551336274e-23
- BTC_RET and LTC_RET: p-value = 7.855462705098844e-28
- BTC_RET and EOS_RET: p-value = 1.0541411510509593e-22
- BTC_RET and BNB_RET: p-value = 0.0

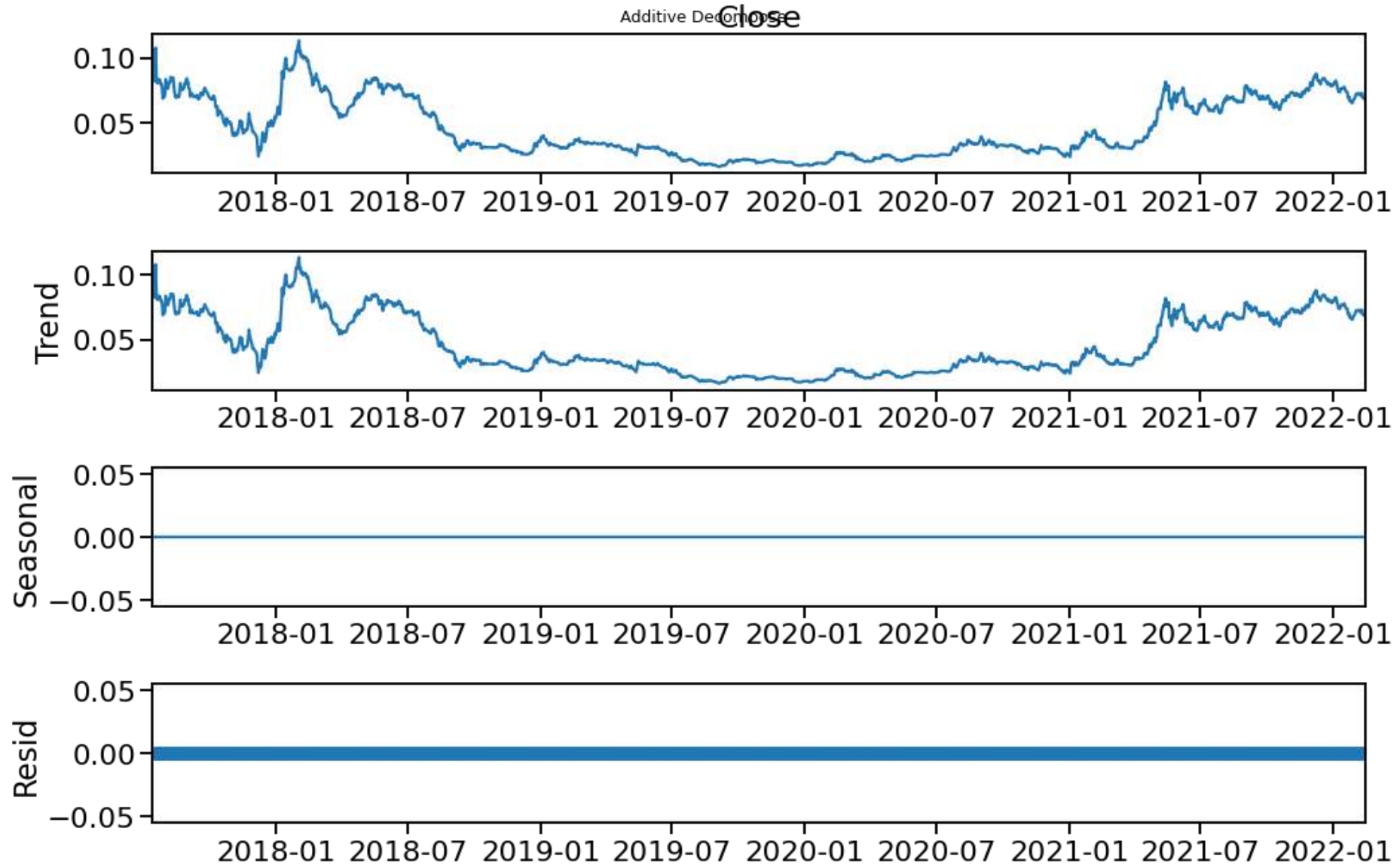
CORRELATION MATRIX HEATMAP



MOVING AVERAGE PLOT



DECOMPOSITION OF CLOSE PRICE OF ETH-BTC



Stationarity Checking of Close Price by Augmented Dickey-Fuller (ADF) test

ADF Statistic: -2.6280922757920044

p-value: 0.08732904025578841

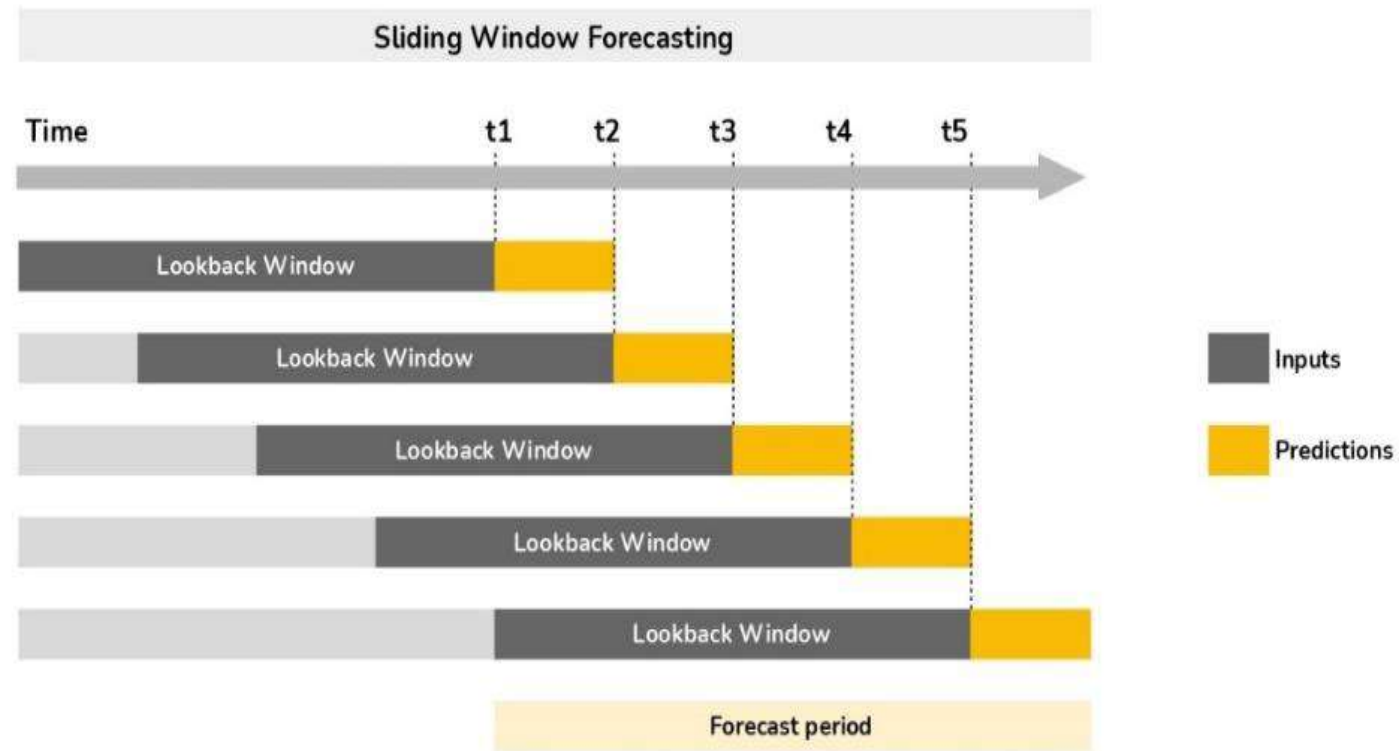
Critical Values:1%: -3.4343

5%: -2.8633

10%: -2.5677

Here **p-value** is too Large than **alpha** Hence **accept H0** and conclude that **process is not stationary**.

SLIDING WINDOW



Validation of Prediction of Close Price Test Data by LSTM

Validation of, predictions of ETH-BTC's Close price By LSTM



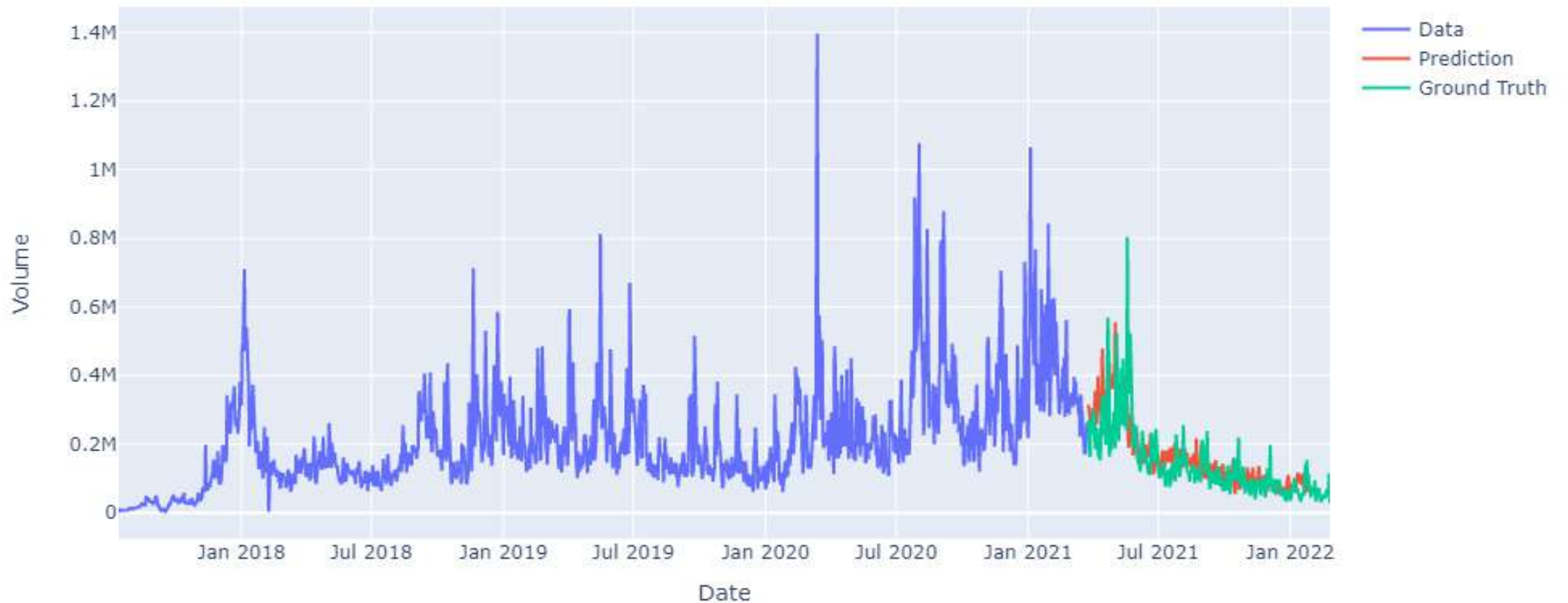
LSTM Forecast for Close Values

Closing price plot Forecast for next 30 days by LSTM



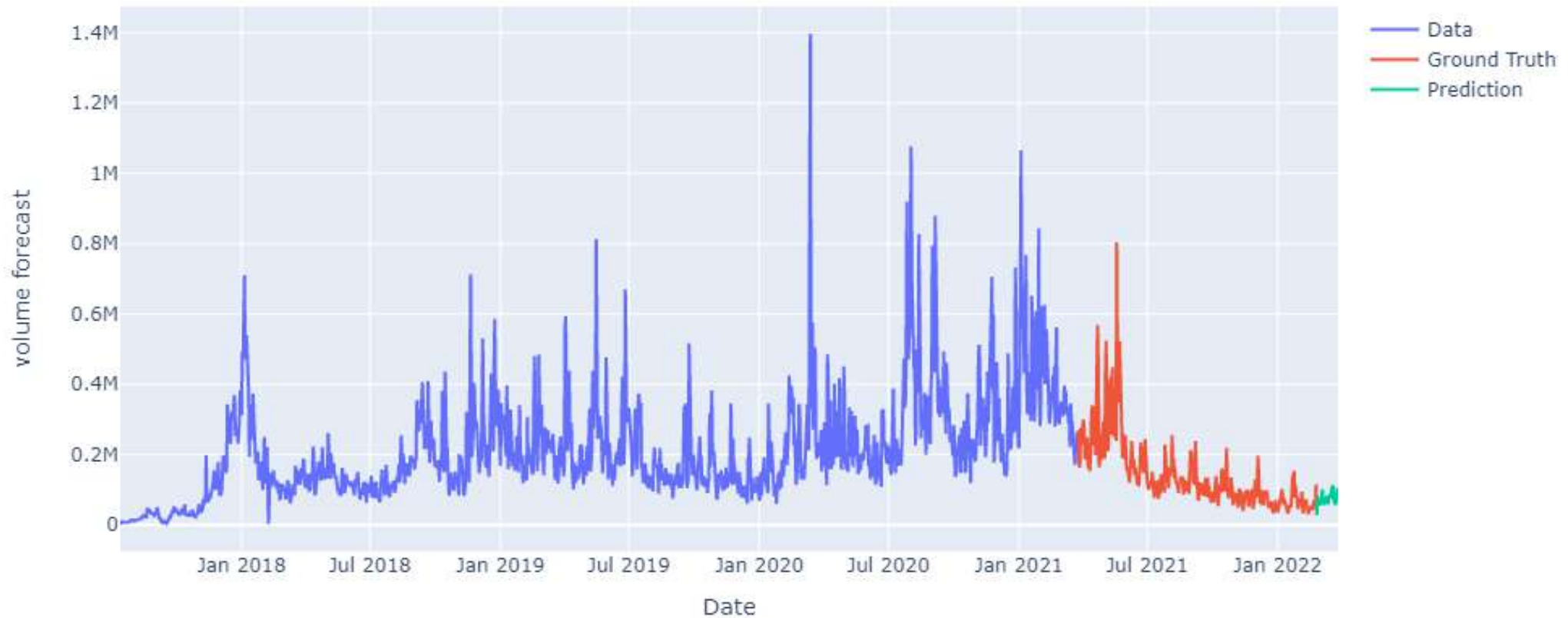
Validation of Prediction of Volume Test Data by LSTM

Validation of scaled test data of ETH-BTC's volume by LSTM

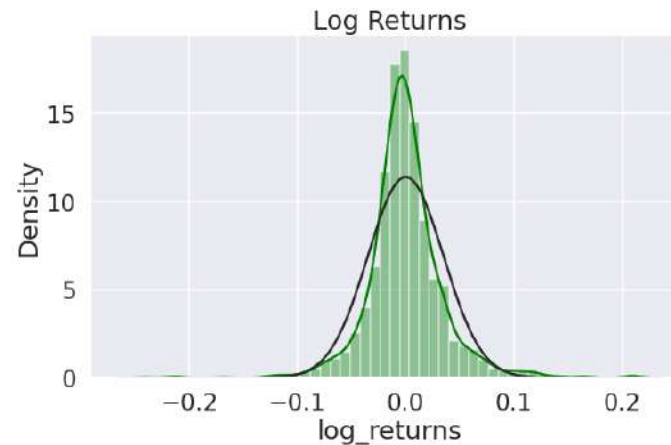
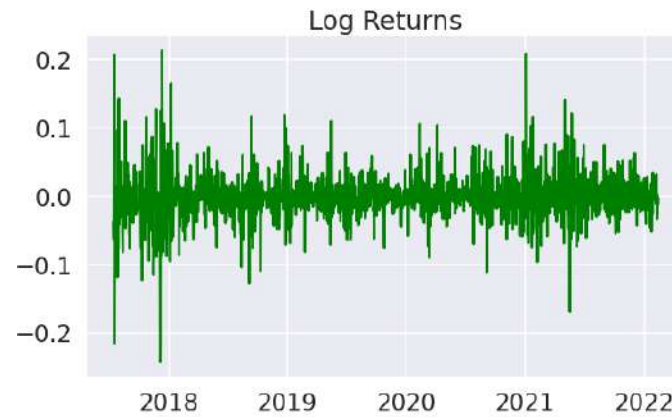
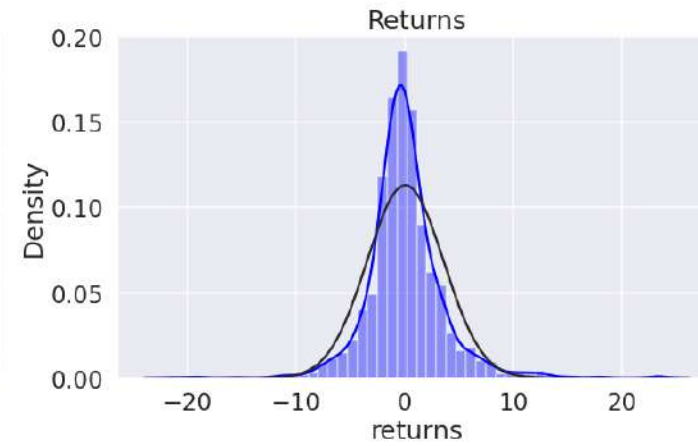
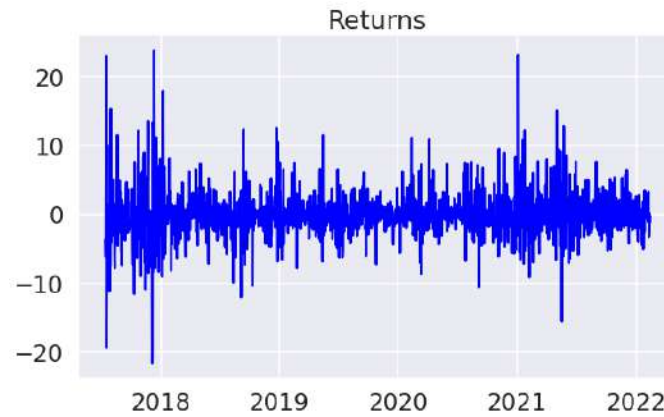


LSTM Forecast for Volume

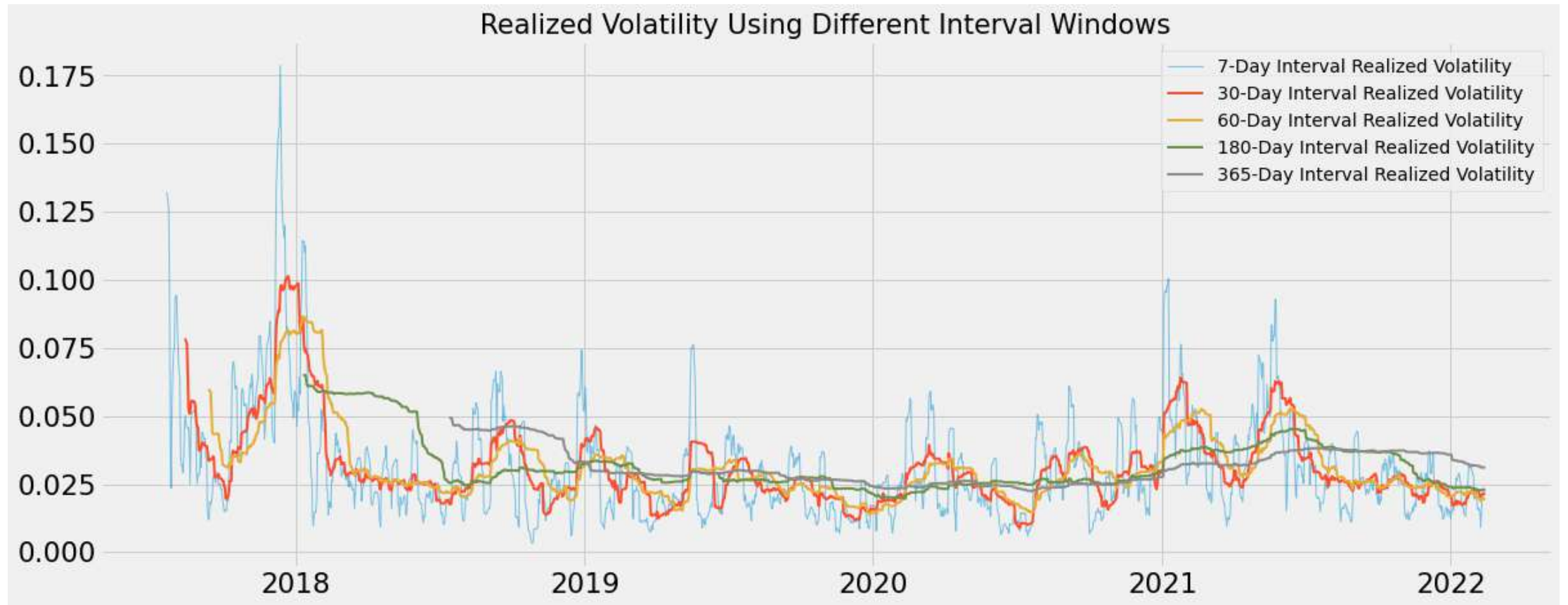
Forecast of volume of ETH-BTC for next 30 Days



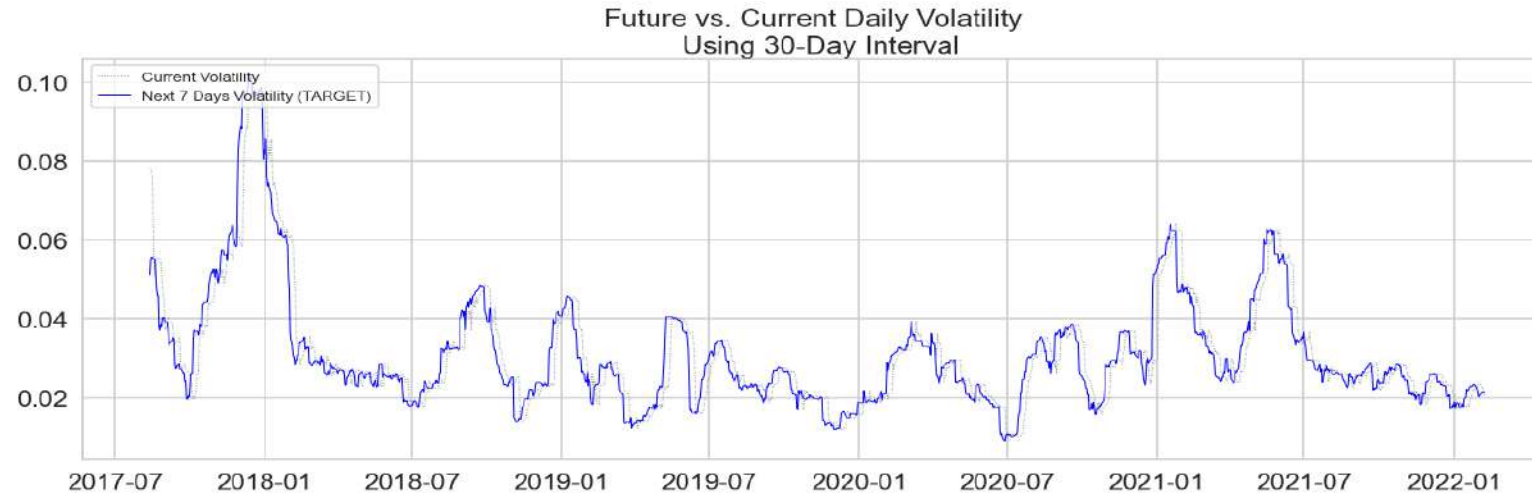
Distribution Plot of Returns and Log Returns



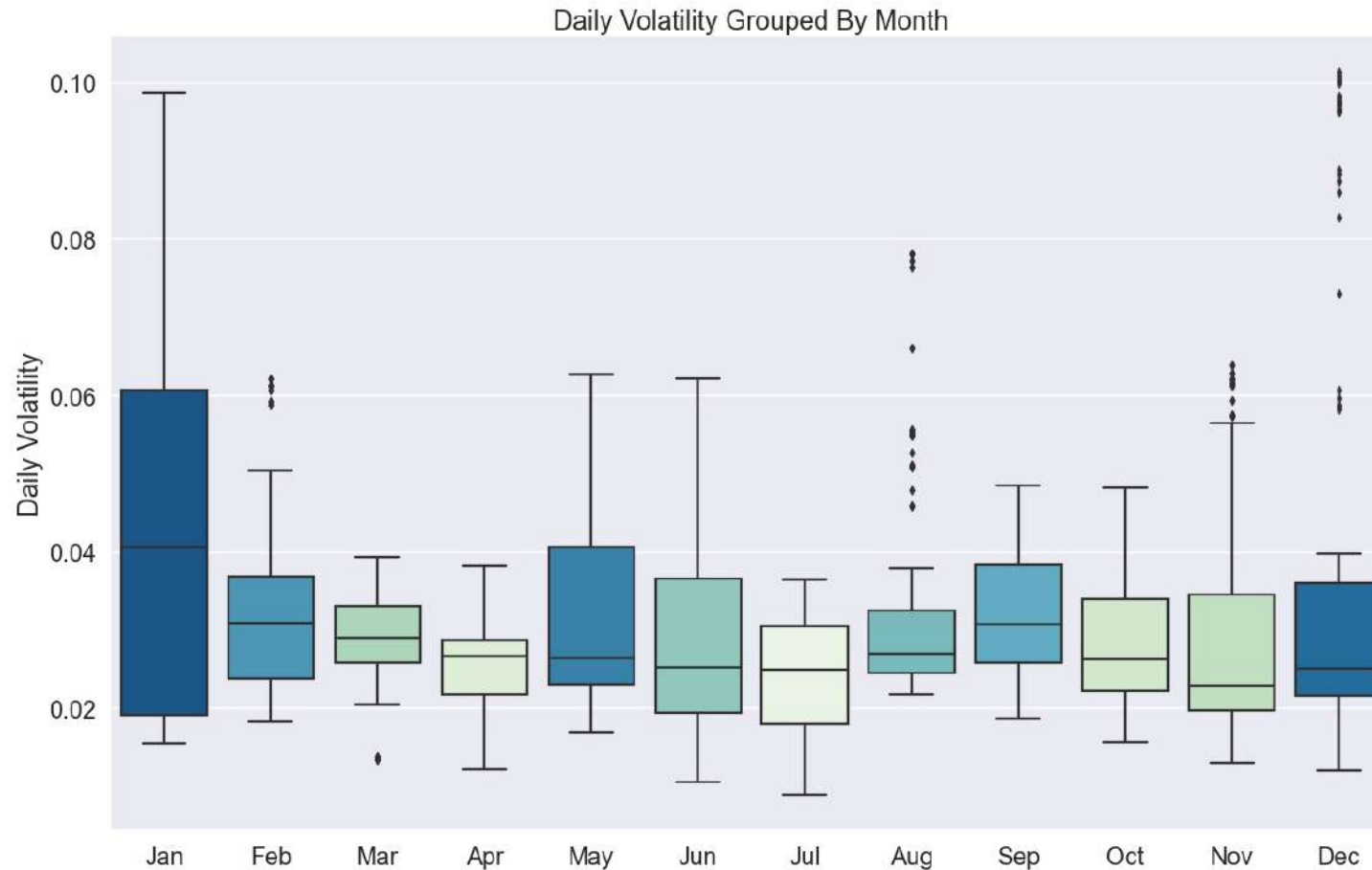
Interval Window Selection



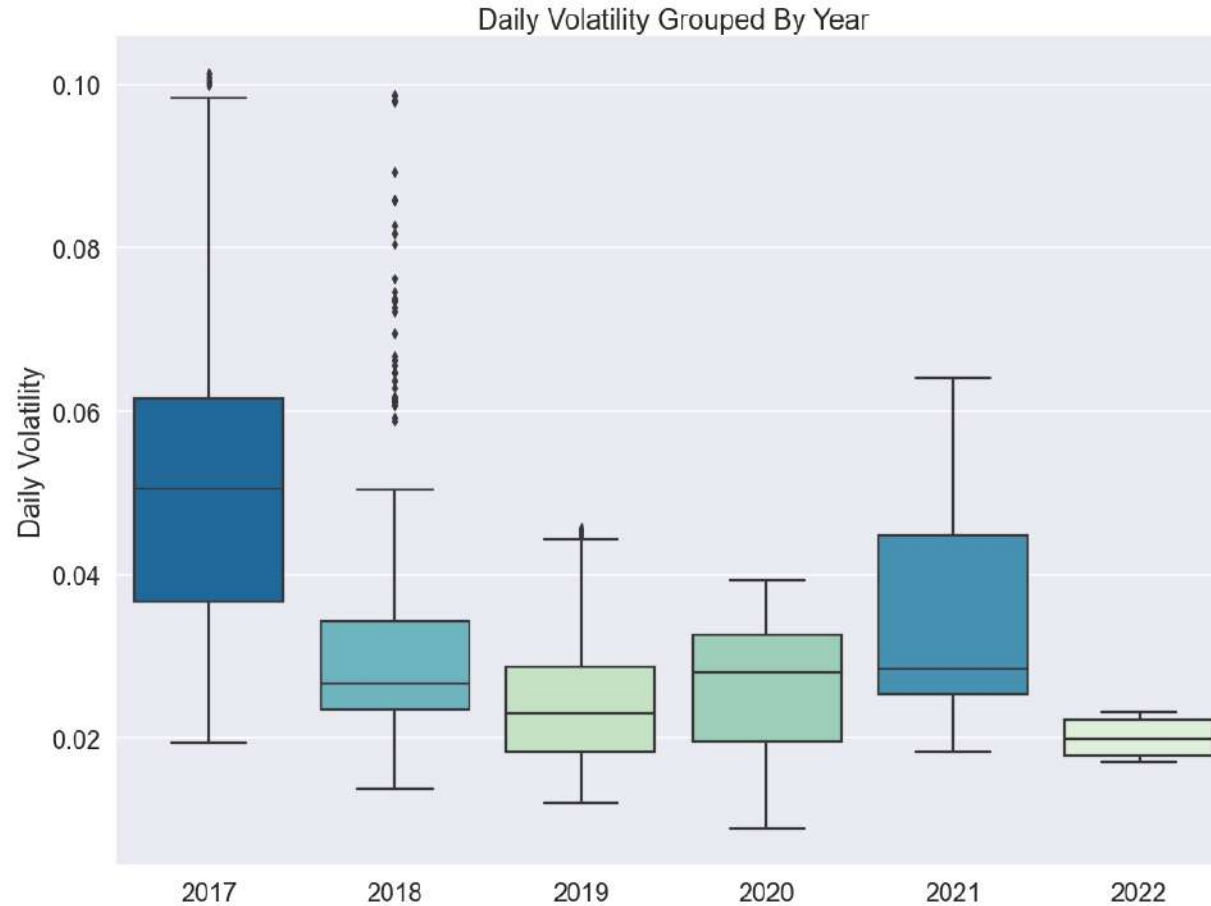
Plot of Future volatility vs Current volatility (using 30 day interval)



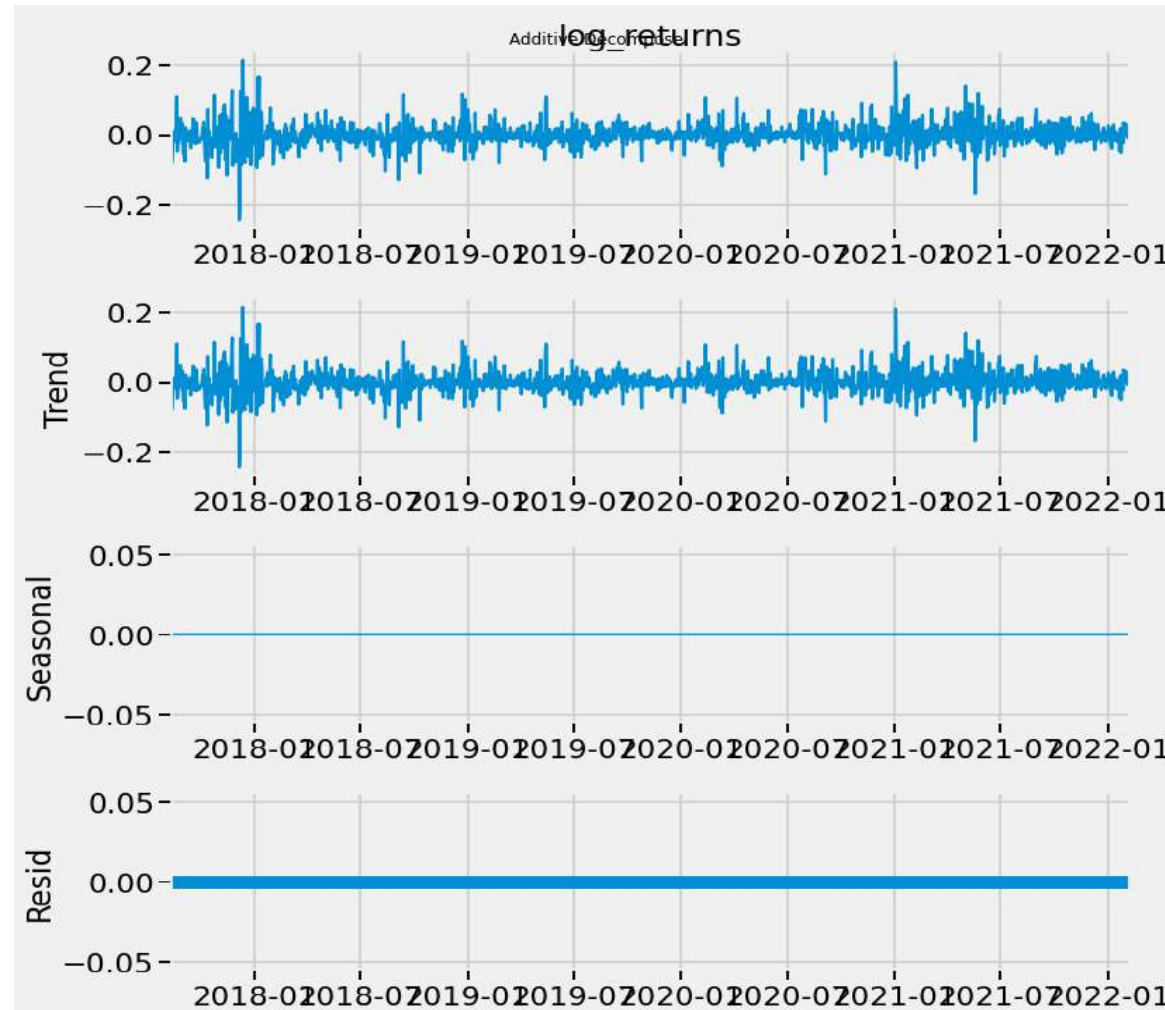
Plot of Daily Volatility Grouped by Month



Plot of Daily Volatility Grouped by Year



Returns and Log Returns Decomposition:



Returns and Log Returns Stationarity Checking

Returns

ADF Statistic: -7.156171946822528

p-value: 3.0506955923818777e-10

Critical Values:

1%: -3.4344

5%: -2.8633

10%: -2.5677

Log Returns

ADF Statistic: -7.28402390437071

p-value: 1.4748762010608668e-10

Critical Values:

1%: -3.4344

5%: -2.8633

10%: -2.5677

- Here p-value too small than alpha Hence reject H0 and conclude that process is stationary.
- --> Returns and Log returns are both not dependent on time/trend

Fitting of Auto ARIMA

Here our Best model: ARIMA(0,0,1)(0,0,0)[0] In our fitted SARIMA Model $p=d=P=D=Q=0$ so it is MA(1) model

but we don't consider this model because our best fitted model is changing every time due to incoming data and Heteroskedasticity

Heteroskedasticity(H) : 0.45

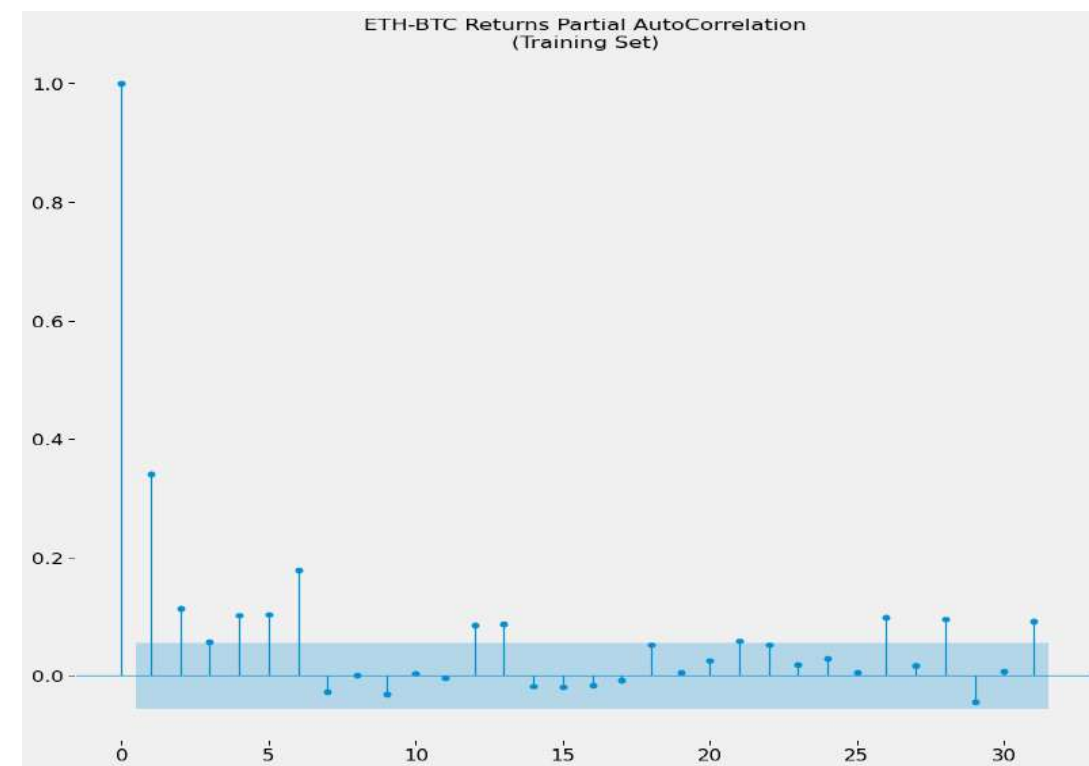
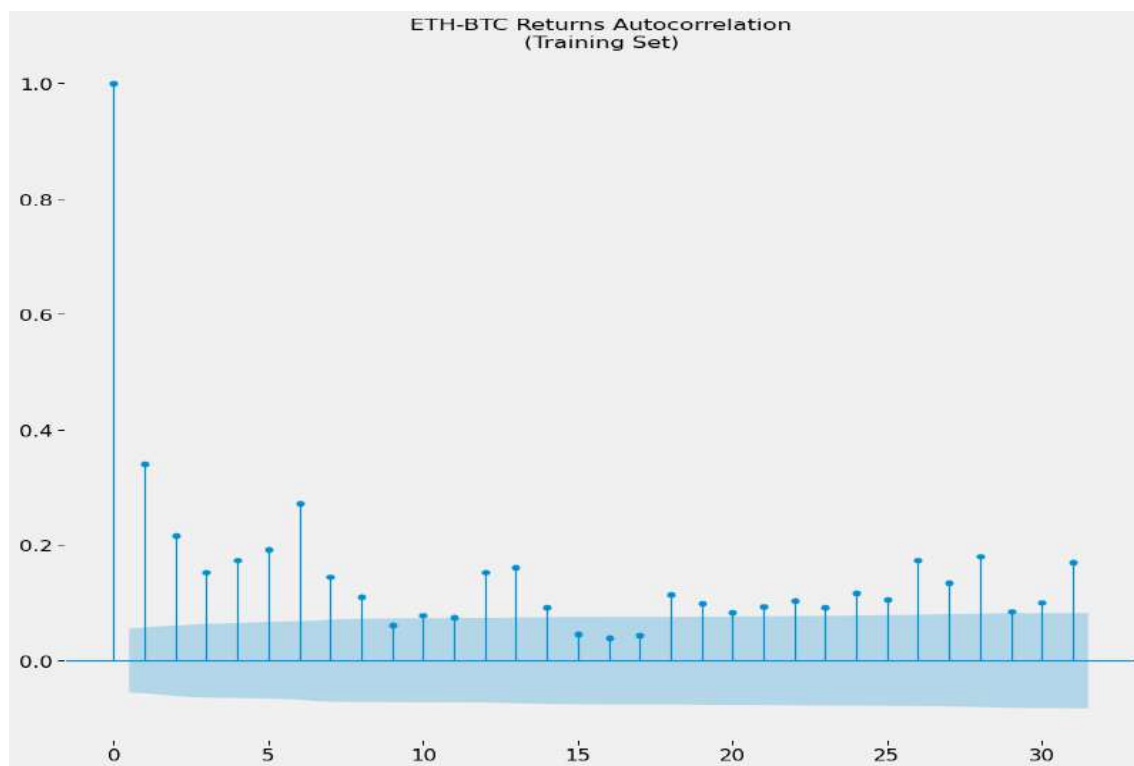
H0 : Homoskedasticity is present present (residuals are equally scattered) v/s
H1 :

Heteroskedasticity is present present (residuals are not equally scattered)

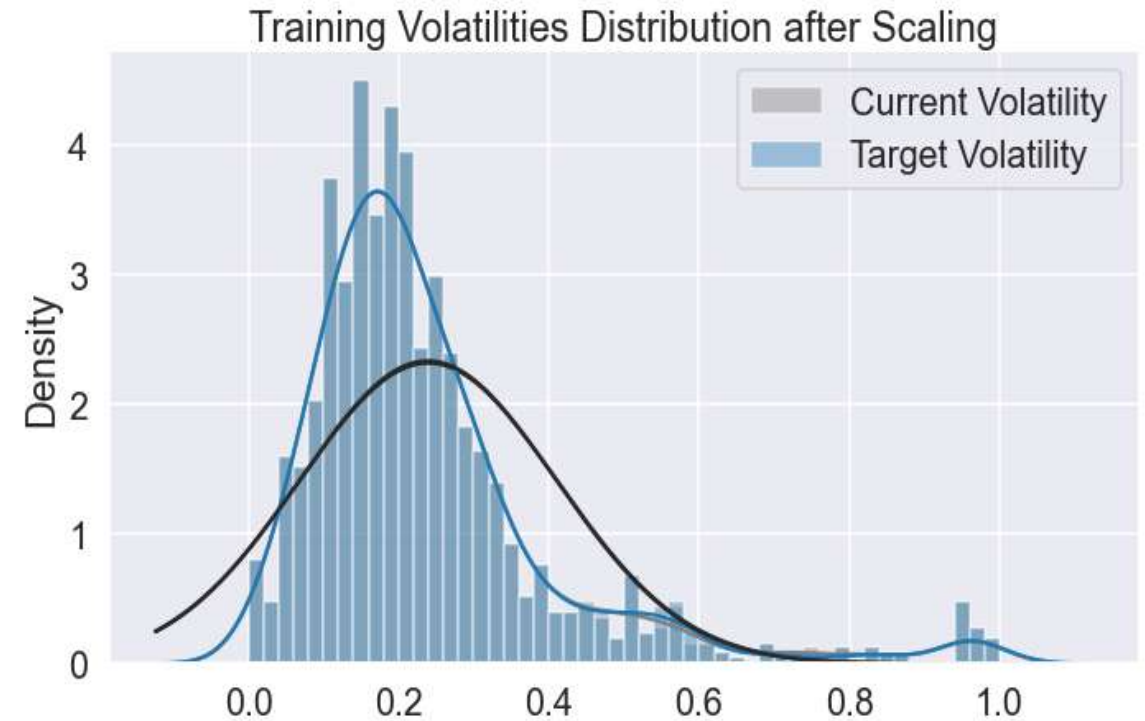
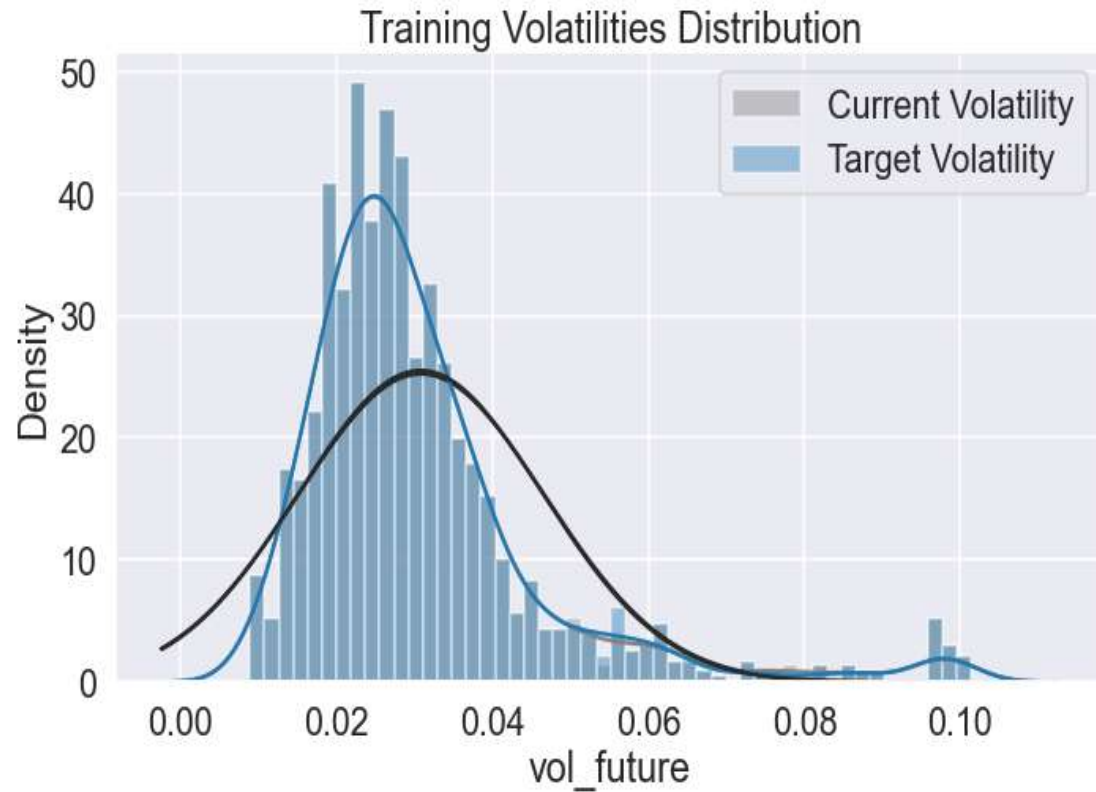
Prob(H) (two-sided) : 0.00 < Alpha (0.05)

Hence we Reject H0 and conclude that significant Heteroskedasticity is present (0.45)

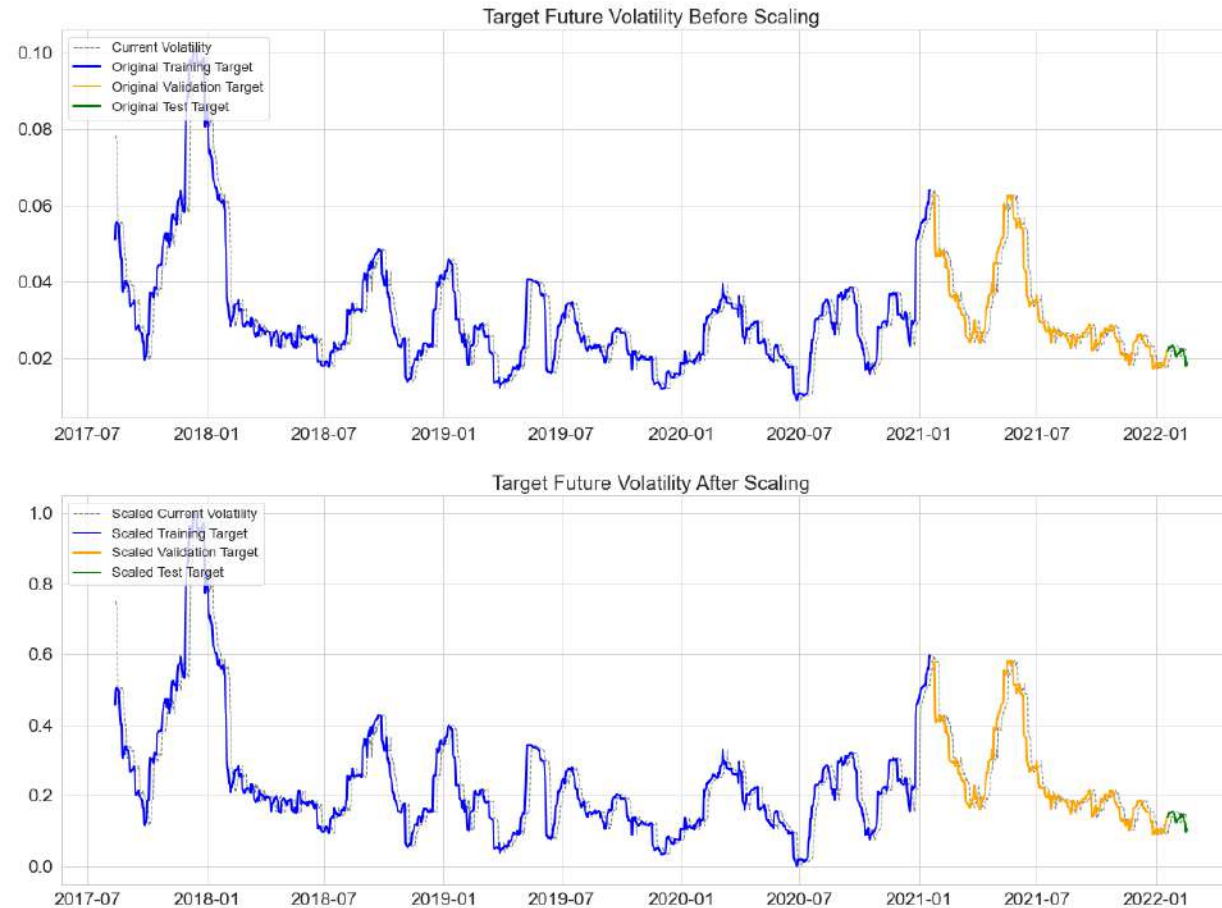
ACF & PACF Plots



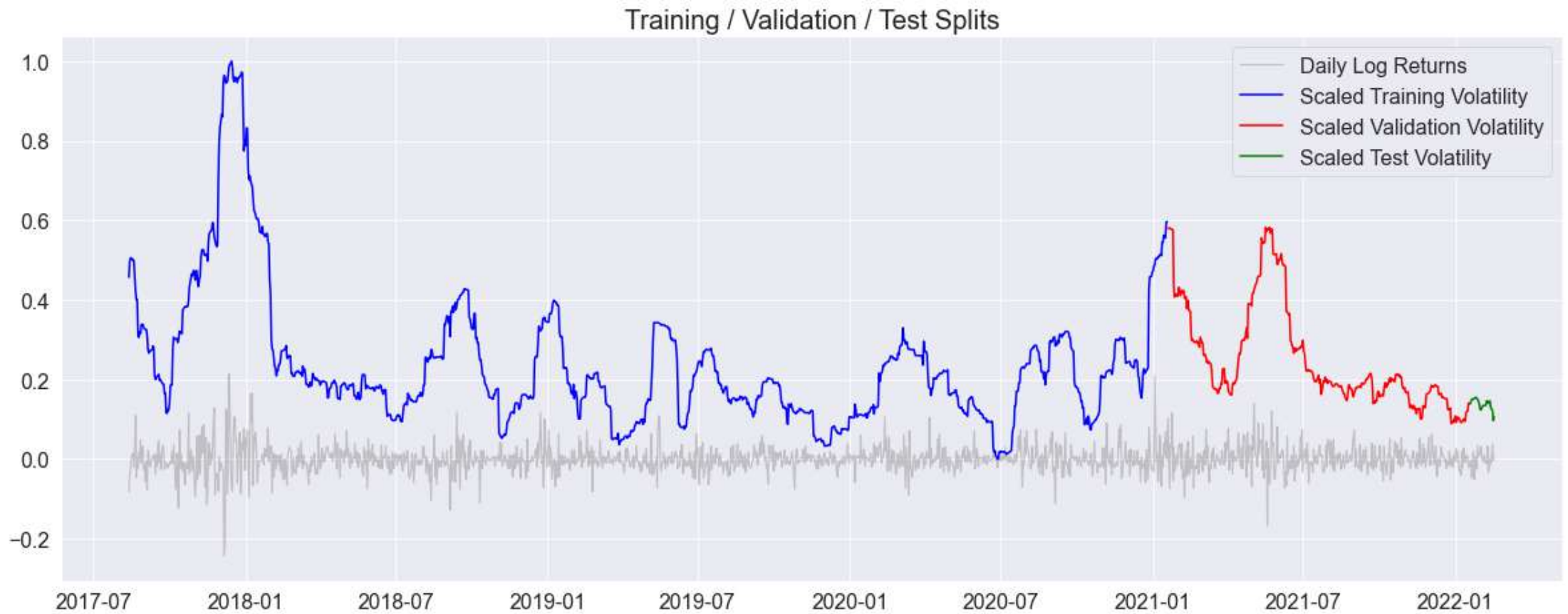
Training Volatility Before and After Scaling



Train-Validation-Test Visualization



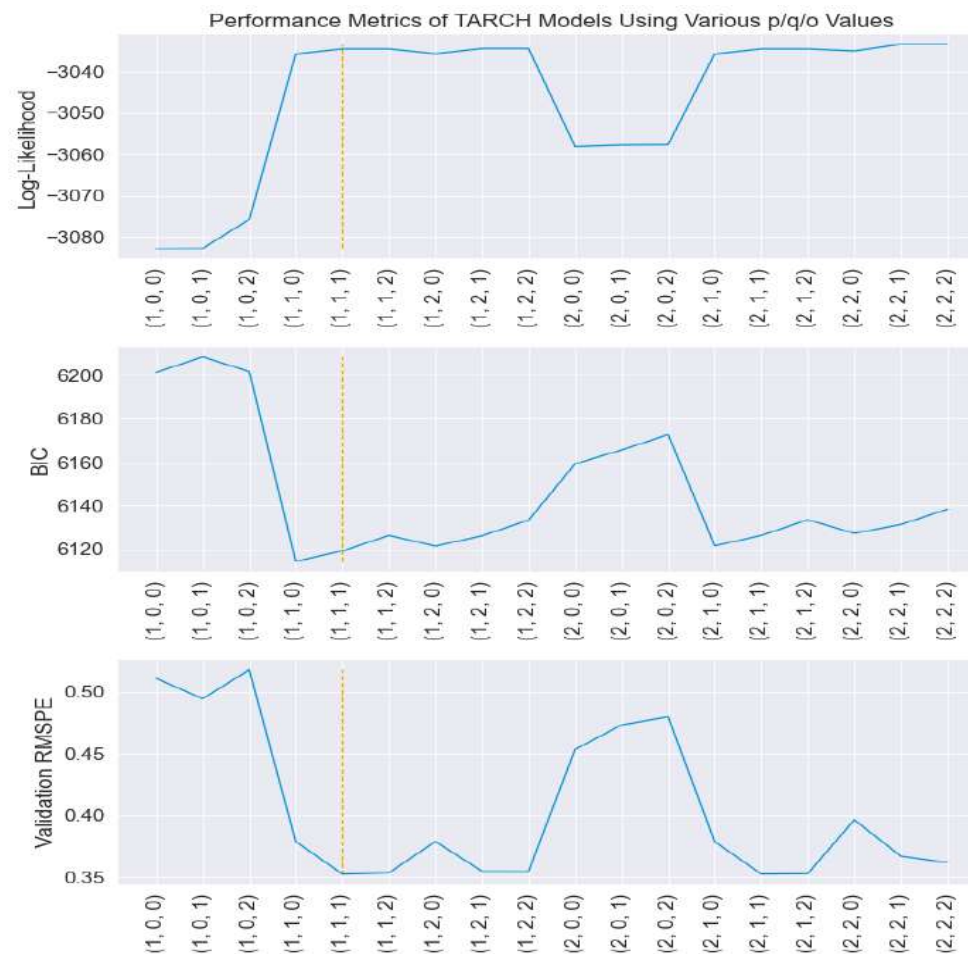
Scaled Train-Validation-Test Visualization



Models used for Volatility Analysis

- Baseline Models
 - Mean baseline model
 - Random Walk Naïve forecasting
- GARCH Models
 - Basic GARCH Model
 - GJR GARCH model
 - TARCH
 - Simulation based TARCH
 - Bootstrap Based TARCH

Hyperparameter Tuning for TARCH

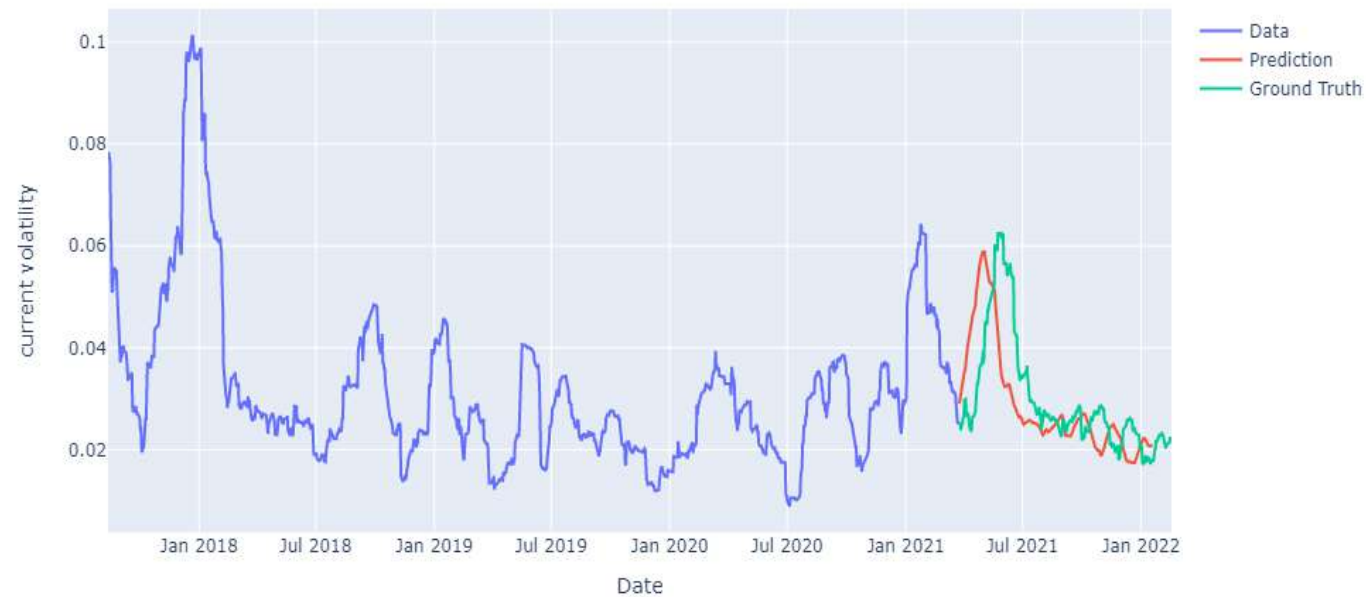


Performance Metrix

<i>Model</i>		<i>Validation RMSPE</i>	<i>Validation RMSE</i>
0	<i>Mean Baseline</i>	0.544414	0.110840
1	<i>Random Walk Naive Forecasting</i>	0.198058	NaN
2	<i>GARCH(1,1), Constant Mean, Normal Dist</i>	0.456065	0.322710
3	<i>Analytical GJR-GARCH(1,1,1), Constant Mean, Sk...</i>	0.480666	0.338645
4	<i>Simulation TARARCH(1,1), Constant Mean, Skewt Dist</i>	0.351545	0.278526
5	<i>Bootstrap TARARCH(1,1), Constant Mean, Skewt Dist</i>	0.352703	0.279399
6	<i>Bootstrap TARARCH(1, 1, 1), Constant Mean, Skewt...</i>	0.352394	0.279253
7	<i>LSTM</i>	0.112561	0.101456

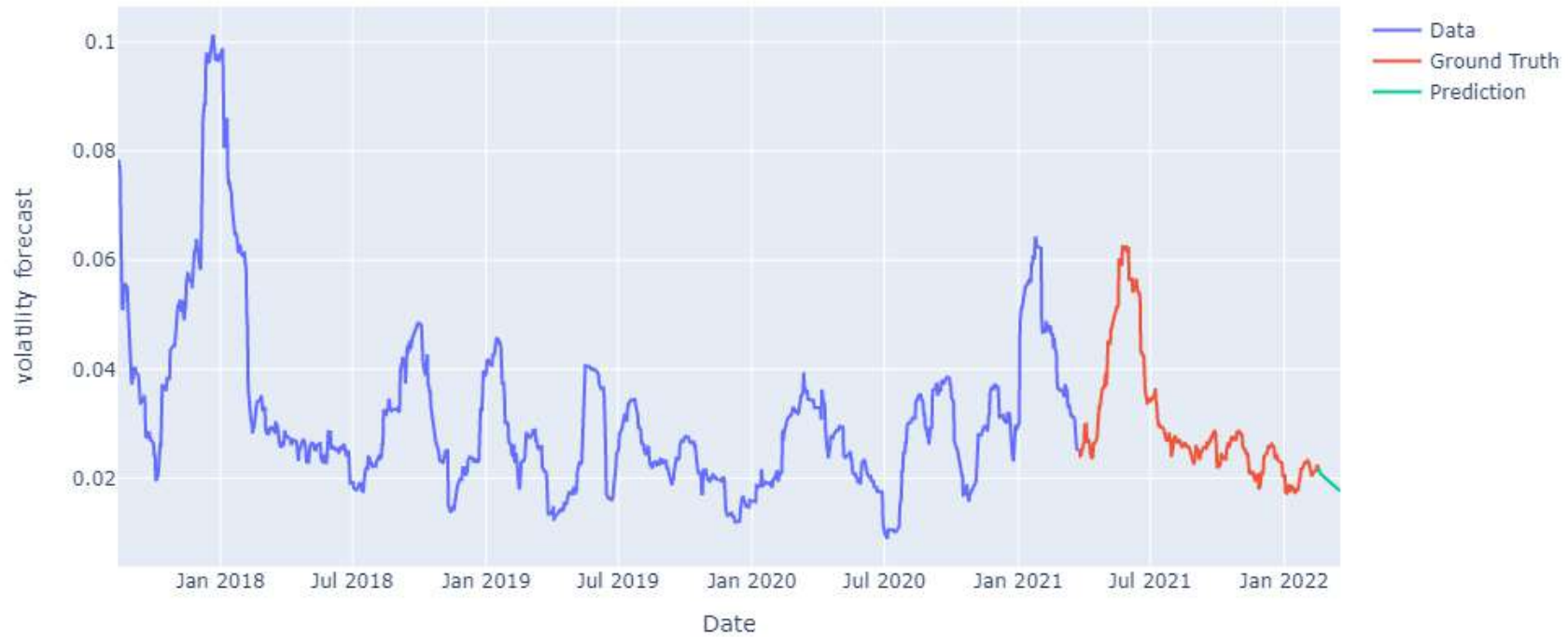
LSTM Validation for Test Data of Volatility

Validation of scaled test data of ETH-BTC's volatility by LSTM



LSTM Forecast for Volatility

Forecast of ETH-BTC for next 30 Days



SUMMARY

From Heatmap we select best pair of cryptocurrency ETH-BTC (Base currency-Quote Currency) for further analysis because both are top 2 Cryptocurrencies of Crypto market.

We fit MA on close price of ETH-BTC when MA is too large it not capable to capturing some spike and dips and when MA is small its again there is again noisy observation problem so there is compromise and we need to looking for some other methods like AR, ARMA, ARIMA, ACF, EWMA

We check Stationarity of close price so Our Close prices is not stationary hence instead of Calculating volatility of close price we choose to formulate returns and then volatility of returns for further analysis.

But before that we forecast Close prices by LSTM and they are slightly decreasing.

we plot volume plot and then forecast it by LSTM so our forecast shows rate of change of decreasing volume is reduced, so its good indicator.

We calculate returns and then log returns For practicality purposes, it's generally preferable to use the log returns especially in mathematic modeling, because it helps eliminate non-stationary properties of time series data, and makes it more stable. There's another advantage to log returns, which is that they're additive across time:

Then we compare the distribution plots and stationarity plot of Returns and Log Returns and their distribution is normal and stationarity plots are also same.

Then we fit auto ARIMA model to log returns but there is high heteroskedasticity so instead of going to traditional time series model we have to choose time series models for High Volatility such as Baseline models and GARCH model

We fit Baseline models, GARCH model and LSTM machine learning model and compare them based on RMSPE & RMSE so our LSTM model have lowest RMSPE & RMSE values.

So we forecast volatility by LSTM model

Our LSTM forecast for Volatility are decreasing in future for next one month.

Overall conclusion is that its good time to tread in BTC Instead of ETH.

Close price forecast by LSTM shows that trend of future close price is decreasing hence Price of Bitcoin is increasing or price of Ethereum is decreasing. So we can transfer our Ethereum portfolio into Bitcoin.

References

- **Wikipedia**
- Géron, A. (2019). *In Hands-on machine learning with Scikit-Learn & TensorFlow: concepts, tools, and techniques to build intelligent systems*. O'Reilly Media, Inc.
- Sinclair, E. (2020). *Positional option trading: An advanced guide*. John Wiley & Sons.
- <https://algotrading101.com/learn/yfinance-guide/>
- <https://www.coursera.org/learn/tensorflow-sequences-time-series-and-prediction/supplement/DM4fi/convolutional-neural-networks-course>
- <https://insights.deribit.com/options-course/>
- https://arch.readthedocs.io/en/latest/univariate/univariate_volatility_forecasting.html
- <https://www.investopedia.com/terms/v/vix.asp>
- <https://www.hindawi.com/journals/complexity/2021/6647534/>

- <https://github.com/ritvikmath/Time-Series-Analysis/blob/master/GARCH%20Stock%20Modeling.ipynb>
- <https://github.com/ritvikmath/Time-Series-Analysis/blob/master/GARCH%20Model.ipynb>
- <https://www.kaggle.com/c/optiver-realized-volatility-prediction>
- <https://www.youtube.com/watch?v=NKHQIN-08S8>
- https://goldinlocks.github.io/ARCH_GARCH-Volatility-Forecasting/
- <https://towardsdatascience.com/time-series-analysis-on-multivariate-data-in-tensorflow-2f0591088502>
- <https://deepmind.com/blog/article/wavenet-generative-model-raw-audio>
- <https://github.com/philipperemy/keras-tcn>
- <http://users.metu.edu.tr/ozancan/ARCHGARCHTutorial.html>
- <https://towardsdatascience.com/8-commonly-used-pandas-display-options-you-should-know-a832365efa95>

THANK YOU

