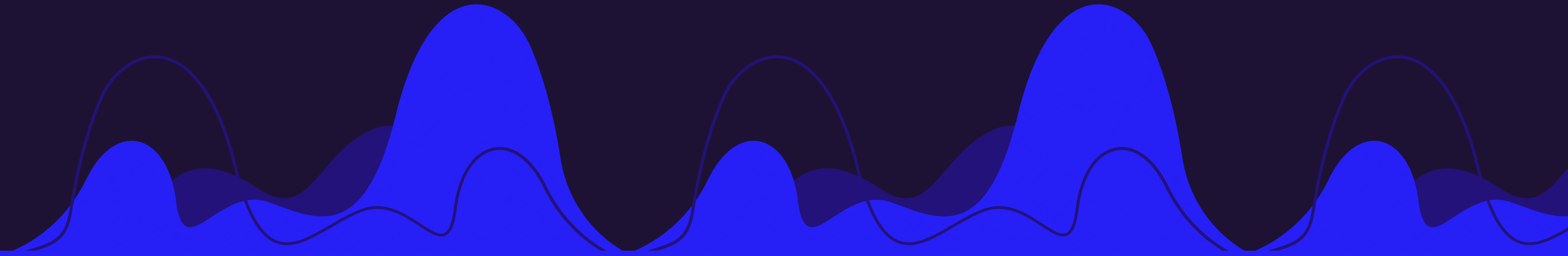


# Stock Prediction using ARIMA



Elona Zharri

SpringBoard Data Science Capstone Project



# Thanks to my mentors



**Eleanor Thomas**

Advisor



**Chris Esposito**

Advisor



# Problem

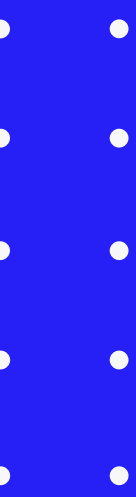
The domain of financial time series data is complicated due to its random walk process, unpredictable day to day variations and being time dependable. Due to the fact that stock markets are affected by many highly interrelated economic, political and even psychological factors, it is difficult to forecast the movement of the stock market.



In the earlier stage, under the assumption of efficient market investors believed that the movement of stock price presents a state of random walk. That means it is impossible to predict the change of stock price by its historical data. Nevertheless, some researchers who did empirical studies applying investment portfolios found historical information is actually useful in prediction.



This project is focused on answering “How close can we get in predicting the stock market?”, by utilizing historical data, a variety of exploratory techniques and different supervised machine learning algorithms.



# Data

I have used the TESLA Dataset ranging from 2010 to 2020, consisting of 6193 rows and 7 columns.

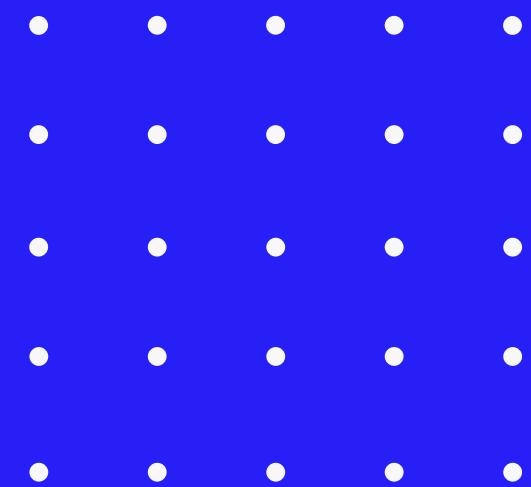
Dataset features:

- Date
- Open
- Low
- High
- Close
- Adjusted Close
- Volume



- + **Data Wrangling**
- + **Exploratory Data Analysis**
- + **Feature Engineering**
- + **Scaling Splitting Data**
- + **Models & Predictions**
- + **Forecasting**
- + **Model Evaluation**

## Work Flow



# Data Wrangling



## Missing Data

No missing or duplicated data was detected.



## Date to DatetimeIndex

Changing the 'Date' column into DatetimeIndex and assigning it as the index of the dataframe, provides better flexibility in resampling data.

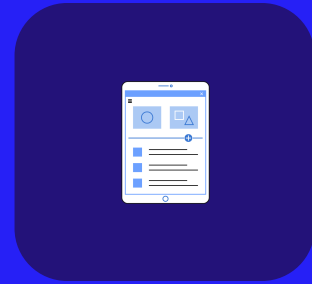


## Adjusted Close to Close

The 'Close' column was dropped, and the 'Adjusted Close' column was renamed to 'Close', as it will give a better idea of the overall value of the stock.



# Exploratory Data Analysis



## Stationarity

Checked for stationarity by plotting moving average, and identified an increasing trend of the data.



## Data Transformation

Removed data trend by taking the log of the 'Close' column. Did not take the difference of the series as the `auto_arima` function will determine the most optimal for the difference.



## ADF Test

The ADF test was used to show that data is not stationary, which in time series analysis is a very important component.



# Feature Engineering



## Moving Average

The reason for calculating the moving average of a stock is to help smooth out the price data over a specified period of time by creating a constantly updated average price.



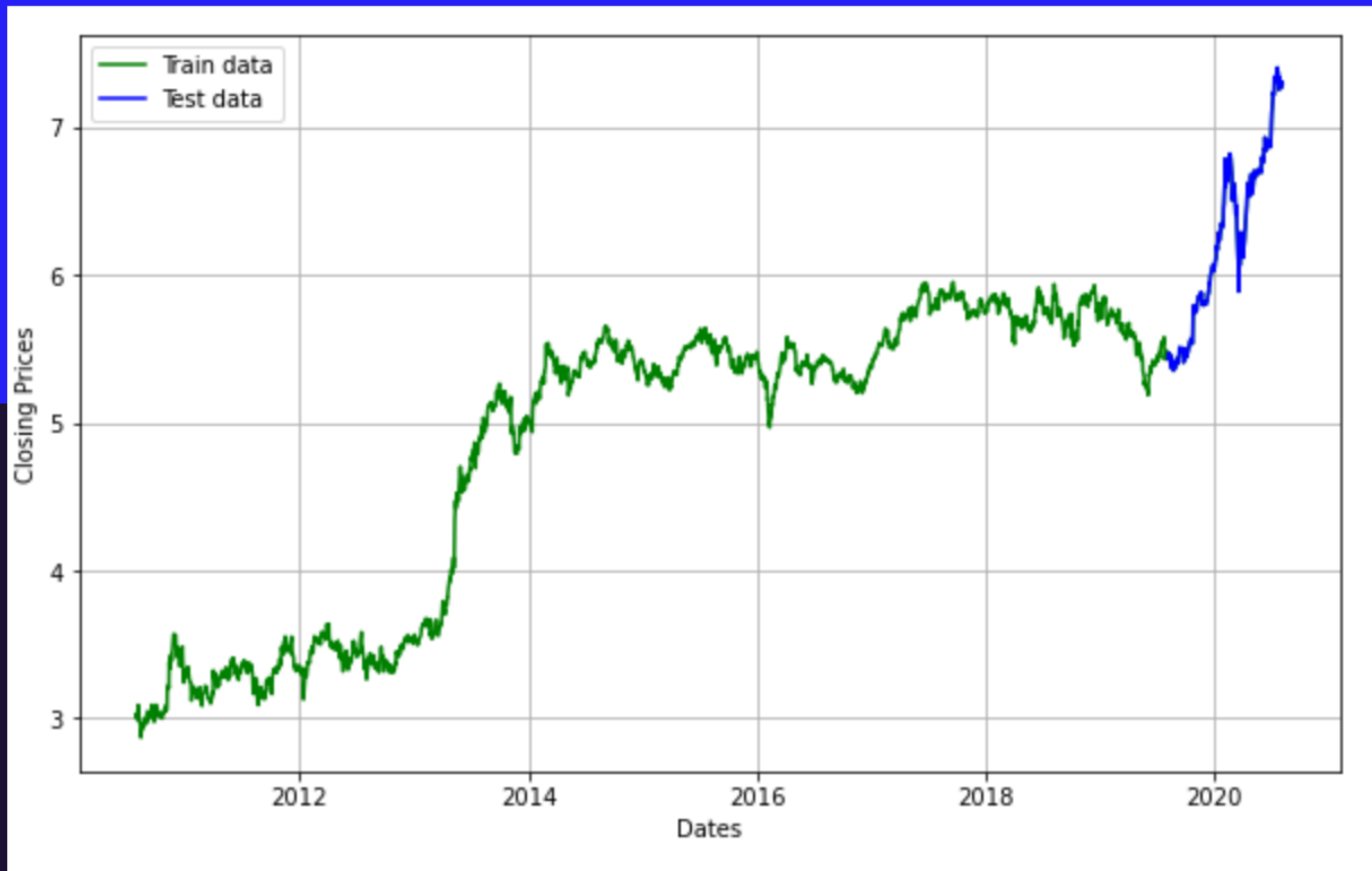
## Bollinger Band

The reason for calculating BB is to generate oversold or overbought signals.





# Splitting Data



# Model & Predictions

Best model: ARIMA(0,1,0)(0,0,0)[0] intercept  
Total fit time: 1.595 seconds

## SARIMAX Results

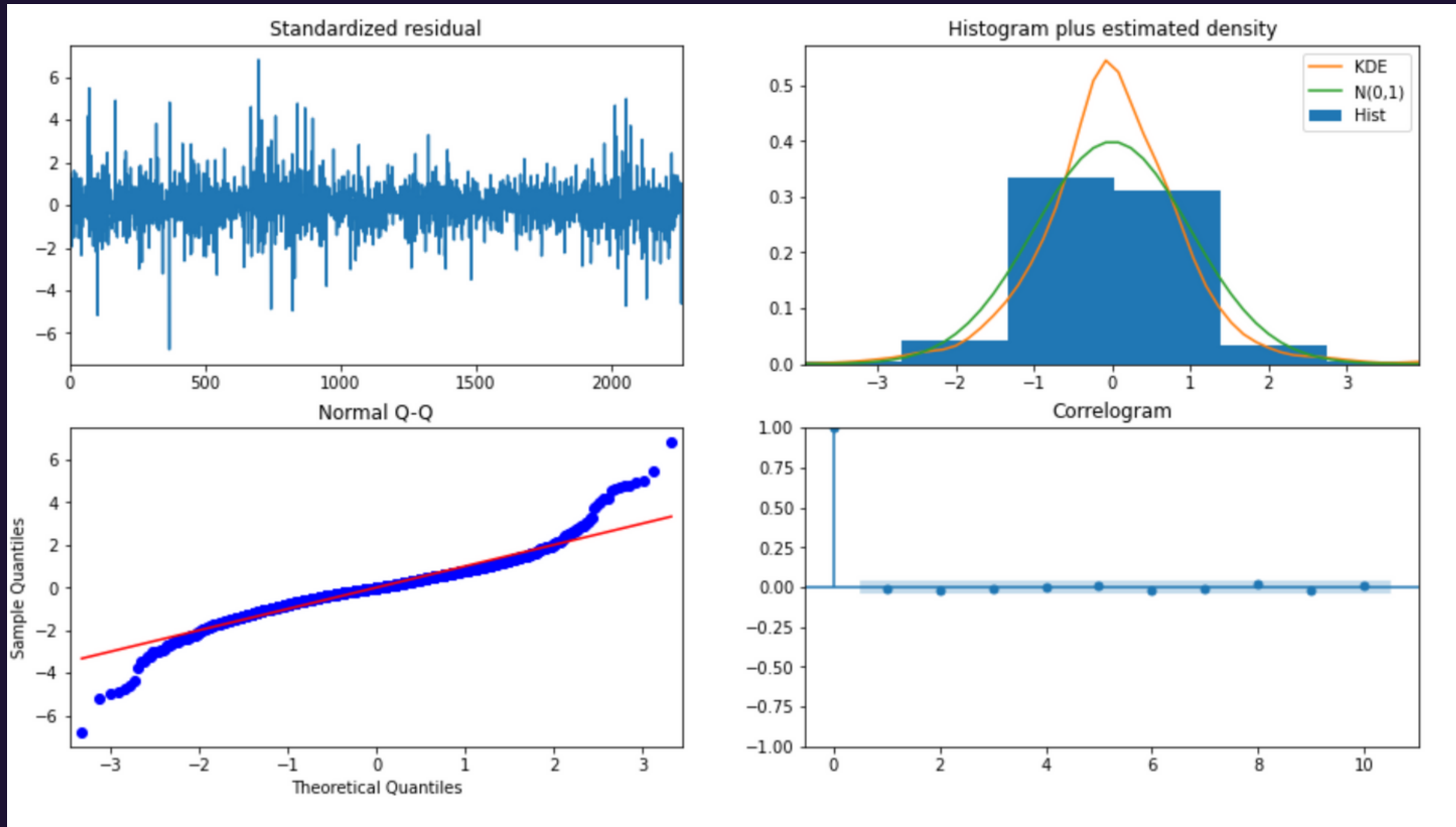
```
=====
Dep. Variable:          y      No. Observations:      2268
Model:                SARIMAX(0, 1, 0)      Log Likelihood      4594.006
Date:                Thu, 23 Sep 2021      AIC      -9184.012
Time:                22:47:18      BIC      -9172.559
Sample:                0      HQIC      -9179.833
                  - 2268
Covariance Type:      opg
=====
```

	coef	std err	z	P> z	[0.025	0.975]
intercept	0.0011	0.001	1.619	0.105	-0.000	0.002
sigma2	0.0010	1.54e-05	65.888	0.000	0.001	0.001

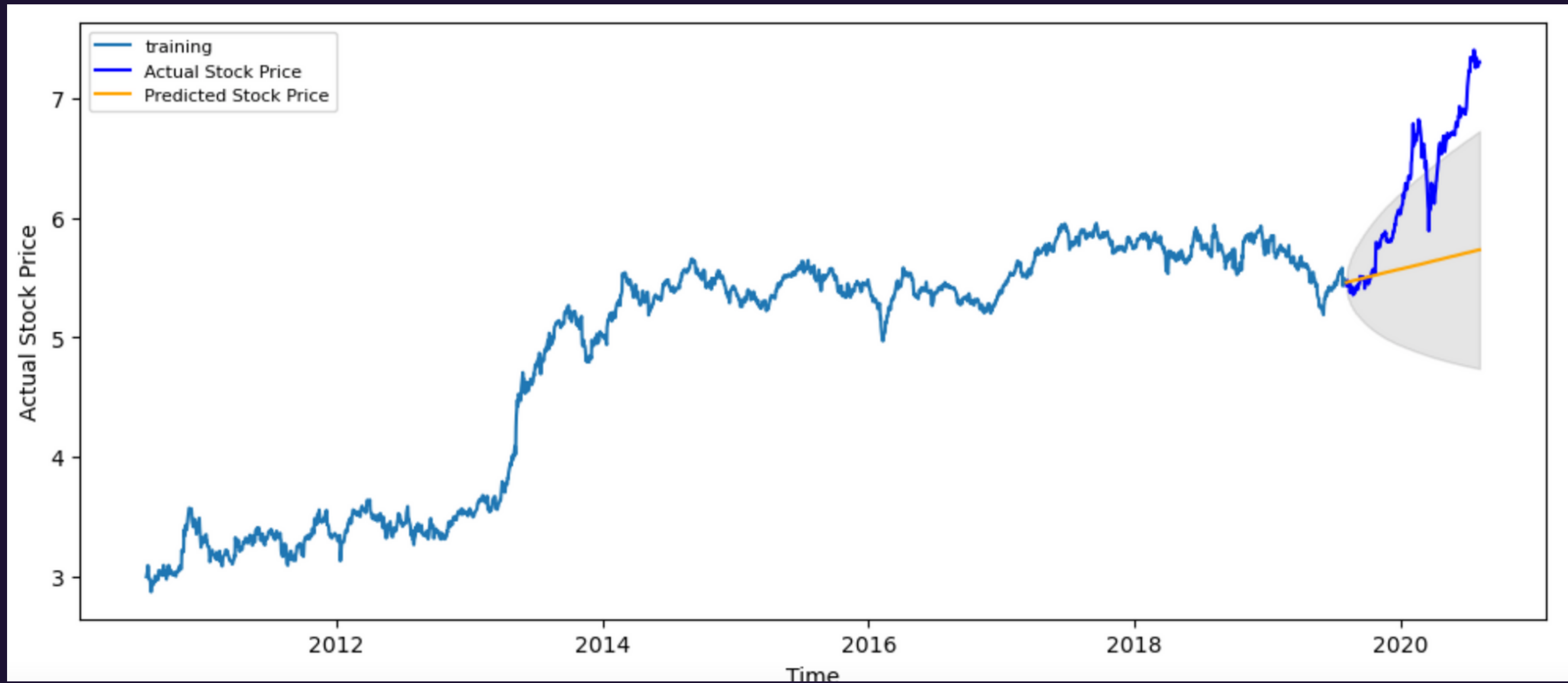
```
=====
Ljung-Box (L1) (Q):      0.52      Jarque-Bera (JB):      3074.65
Prob(Q):                0.47      Prob(JB):                0.00
Heteroskedasticity (H):  0.65      Skew:                0.18
Prob(H) (two-sided):    0.00      Kurtosis:             8.69
=====
```



# Model & Predictions



# Forecasting



# Model Evaluation

MSE: 0.6857742935340155

MAE: 0.6638883342454152

RMSE: 0.8281149035816319

MAPE: 0.09982834603158602



- 
- 
- 
- 
- 
- 

# Future Research

Next steps and goals.

1

## Perform Multivariate Analysis using ARIMA

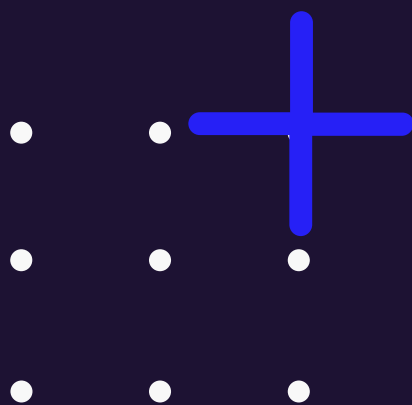
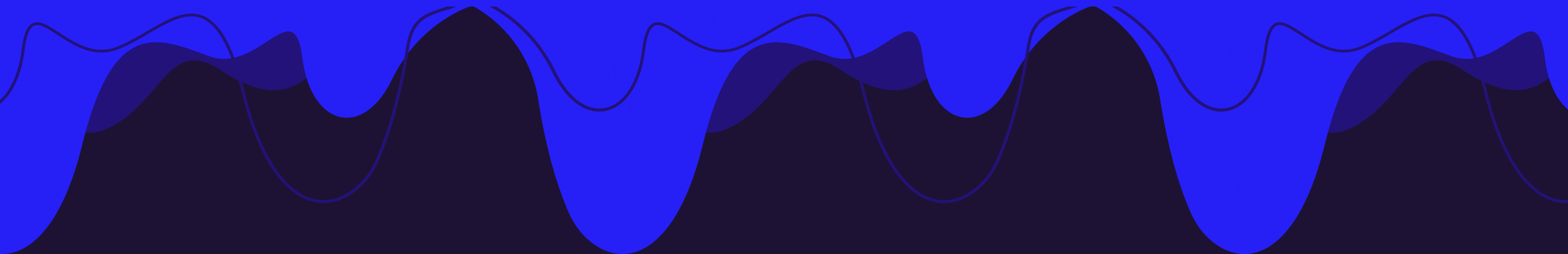
Use the technical indicators as exogenous features, and to analyze model performance.

## LSTM

2

Use the same technical indicators to analyze LSTM model prediction and compare with SARIMAX.





**Thank you!**