Stock Forecast

Analysis and Time-Series Forecasting Stock Price Directional Movements using Linear Regression, ANN and LSTM

Kng Yin Chew

Data Science, FCSIT, UM

Jln Profesor Diraja Ungku Aziz, Kuala Lumpur 50603

yckng00@gmail.com

012-9933855

Introduction

Best Site for Stock Analysis

Stock Market

A significant part in building up the economy of the nation

Technical Analysis

Forecasted by historical price + technical indicators

Forecast

Tomorrow stock market directional movement (up/down)

Problem

the problem i am trying to solve here is...

Problem 1

Uncertain stock price movement

Problem 2

Intricacy and non-linearlity are two fundamental difficulties that cause instability in stock market.

Problem 3

Investing in shares can be highly lucrative or can lead to huge losses

Objectives

Best Site for Stock Analysis

Objective 1

Investigate the relevant techniques in forecasting the stock prices.

Objective 2

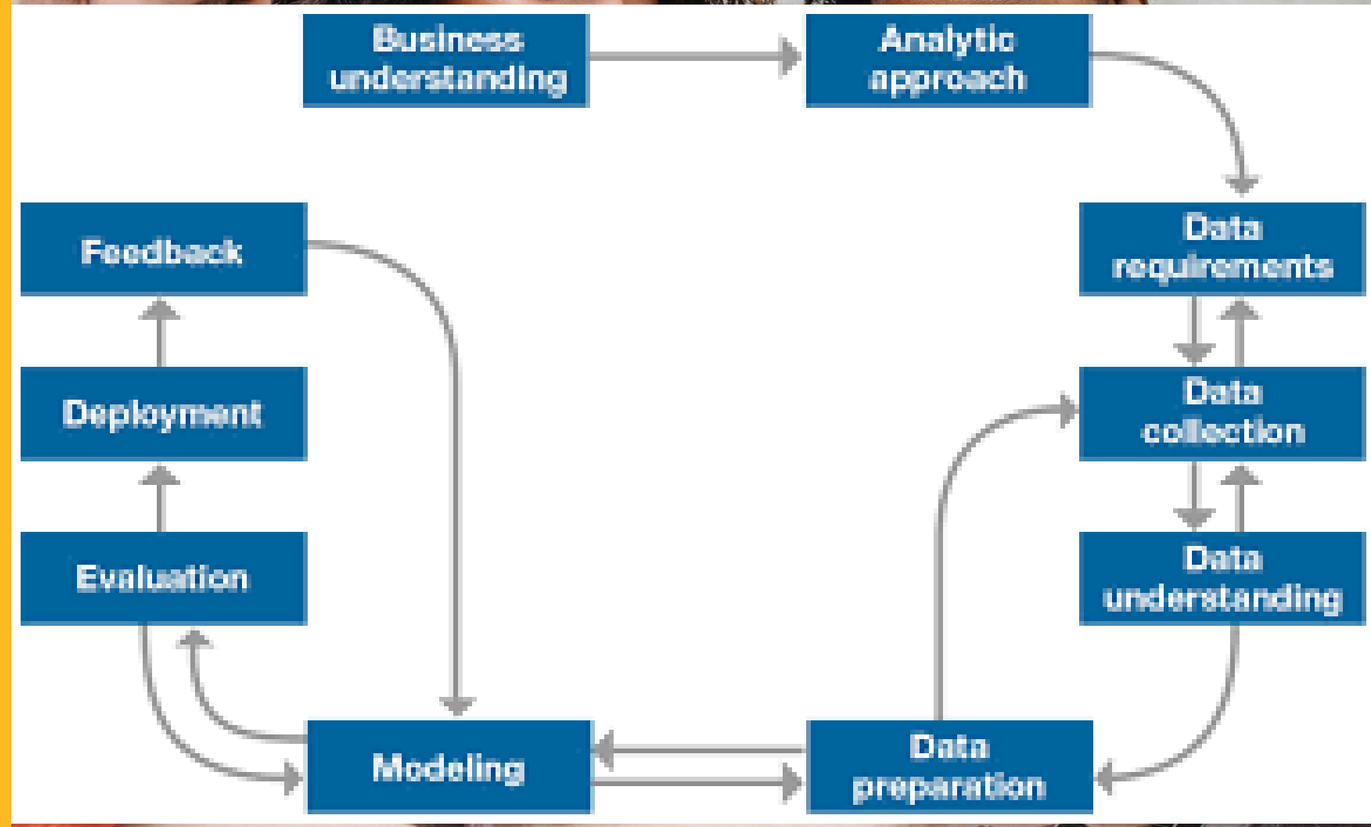
Investigate the use of technical indicators in forecasting the stock prices.

Objective 3

Develop the data product of the stock price forecasting system

Data Science Methodology

Introduce the competitive landscape in your market.



Data

Best Site for Stock Analysis

Historical stock price data (TSLA)

- from year 2010 to year 2021
- downloaded from yahoo! Finance
- using yfinance python libraries

Date	Open	High	Low	Close	Adj Close	Volume
2010-12-16	6.000000	6.182000	5.930000	6.162000	6.162000	3950500
2010-12-17	6.268000	6.308000	6.142000	6.272000	6.272000	4065000
2010-12-20	6.328000	6.438000	6.252000	6.340000	6.340000	2617000

```
# scrape data from yahooFinance
import yfinance as yf
data = yf.download(tickers='TSLA', start='2010-06-01', end='2021-06-01')
```

Data preprocessing

Best Site for Stock Analysis

Standardization

Standardize the whole dataset because the numerical values in the dataset is varied in scales

Remove NaN rows

After doing feature engineering, due to the technical indicators needs history data to calculate its value hence there will be missing data occurs

Feature Engineering

Best Site for Stock Analysis

Historical stock price data (TSLA)

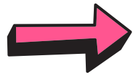
60 previous day's stock closing price

Technical Indicators

15 technical indicators calculated from the stock open, close, high, low price

Historical stock price data

Date	Open	High	Low	Close	Adj Close	Volume
2010-12-16	6.000000	6.182000	5.930000	6.162000	6.162000	3950500
2010-12-17	6.268000	6.308000	6.142000	6.272000	6.272000	4065000
2010-12-20	6.328000	6.438000	6.252000	6.340000	6.340000	2617000



✓ 1s

```
[10] forecasting_step = 60

     for i in range(1, forecasting_step):
         data['Close-%i' % (i)] = data['Close'].shift(i)
```

Close-1	Close-2	Close-3	Close-4	Close-5	Close-6	...	Close-57	Close-58	Close-59
5.920000	5.706000	6.110000	6.304000	6.410000	6.474000	...	4.106000	4.020000	3.912000
6.162000	5.920000	5.706000	6.110000	6.304000	6.410000	...	4.280000	4.106000	4.020000
6.272000	6.162000	5.920000	5.706000	6.110000	6.304000	...	4.396000	4.280000	4.106000
6.340000	6.272000	6.162000	5.920000	5.706000	6.110000	...	4.082000	4.396000	4.280000

Technical Indicators

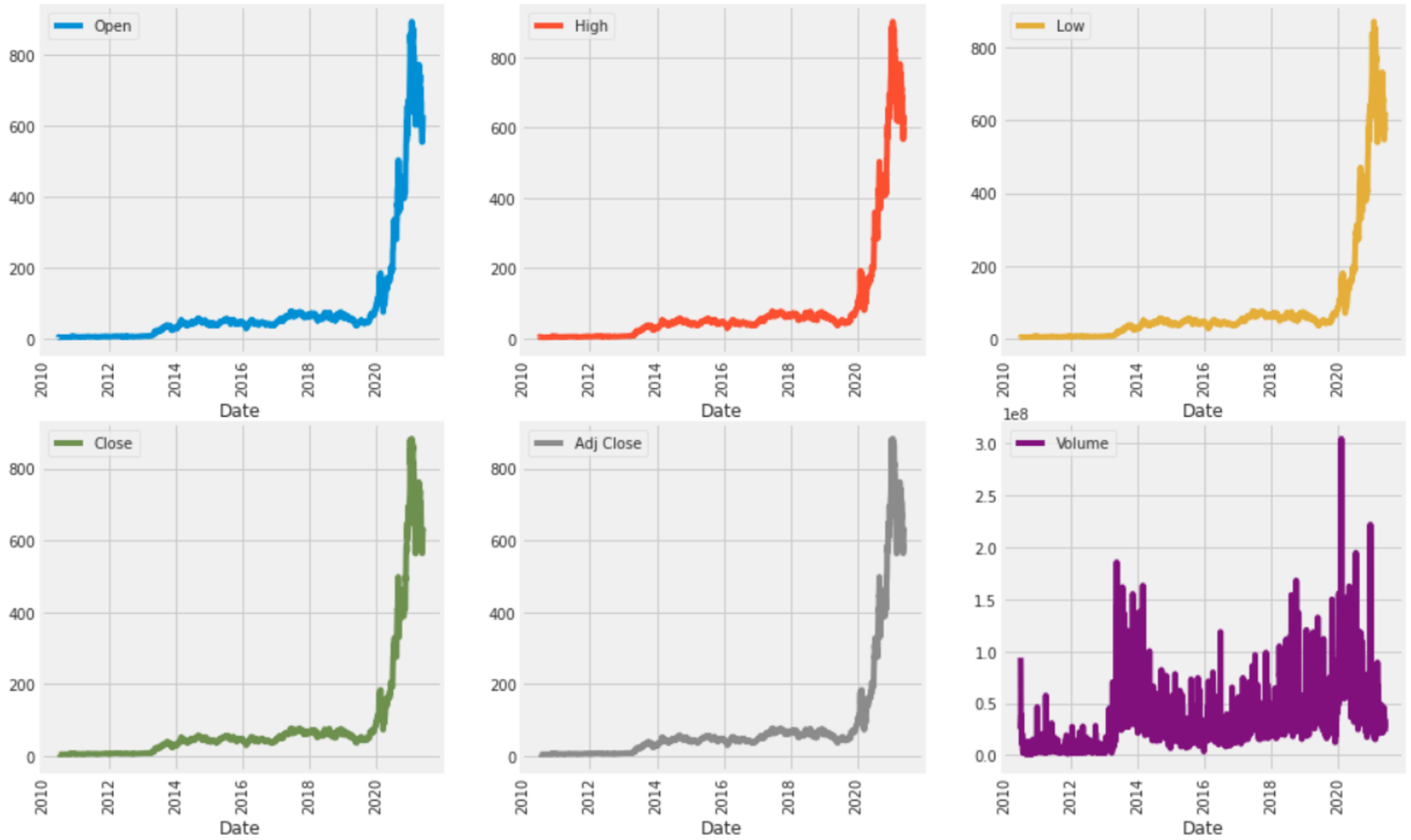
```
# Add  Trading Indicator
# Add RSI(Relative Strength Index)
n = 60 # windows size
data['RSI'] = ta.RSI(np.array(data['Close']),timeperiod = n)

# ATR
data['ATR'] = ta.ATR(np.array(data['High']), np.array(data['Low']), np.array(data['Close']),timeperiod = n)
data

# Create a column by name, SMA and assign the SMA calculation to it
data['SMA'] = data['Close'].rolling(window=n).mean()

# Create a column by name, Corr and assign the calculation of correlation to it
#data['Corr'] = data['Close'].rolling(window=n).corr(data['SMA'])

# Creata a column by name, ADX and assigin the ADX calculation to it
data['ADX'] = ta.ADX(np.array(data['High']),np.array(data['Low']),
                     np.array(data['Close']), timeperiod=n)

# Create a column by name, SAR and assign the SAR calculation to it
data['SAR'] = ta.SAR(np.array(data['High']), np.array(data['Low']),
                     0.2,0.2)

# Directional Movement Index
data['DX'] = ta.DX(np.array(data['High']),np.array(data['Low']),
                     np.array(data['Close']), timeperiod=n)

# Commodity Channel Index
data['CCI'] = ta.DX(np.array(data['High']),np.array(data['Low']),
                     np.array(data['Close']), timeperiod=n)
```

```
# Commodity Channel Index
data['CCI'] = ta.DX(np.array(data['High']),np.array(data['Low']),
                    np.array(data['Close']), timeperiod=n)

# Two window size parameters for calculating EMAs and period for smoothing the signal line
# can make into buy sell signal (point the lines interchange)
data['MACD'], data['MACD_SIGNAL'], data['MACD_HIST'] = ta.MACD(np.array(data['Close']), fastperiod=3, slowperiod=20, signalperiod=3)

# Slow Stochastic
data['SlowK'], data['SlowD'] = ta.STOCH(np.array(data['High']),np.array(data['Low']),
                    np.array(data['Close']), fastk_period=5, slowk_period=3, slowk_matype=0, slowd_period=3, slowd_matype=0)

# Fast Stochastic
data['FastK'], data['FastD'] = ta.STOCHF(np.array(data['High']),np.array(data['Low']),
                    np.array(data['Close']), fastk_period=5, fastd_period=3, fastd_matype=0)

# Williams' %R
data['WILLR'] = ta.WILLR(np.array(data['High']),np.array(data['Low']),
                    np.array(data['Close']), timeperiod=n)
```

	RSI	ATR	SMA	ADX	SAR	DX	CCI
	55.186997	0.304135	5.147767	19.006803	6.212480	18.429832	18.429832
	55.704978	0.301833	5.187100	19.016938	5.552000	19.614954	19.614954
	56.024528	0.299902	5.225767	19.047022	5.703200	20.821959	20.821959
	56.549559	0.298204	5.264867	19.091925	5.850160	21.741186	21.741186

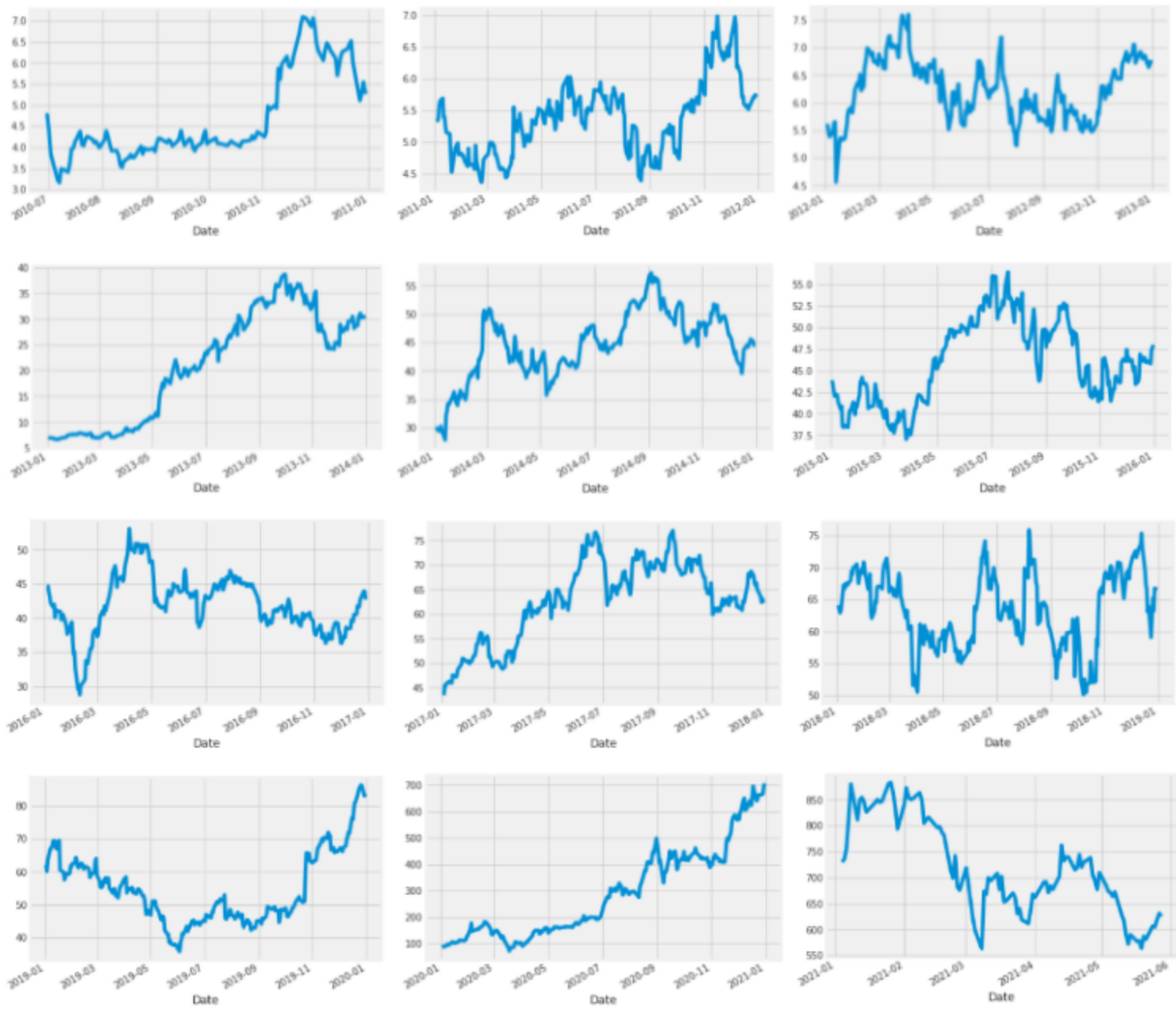
	MACD	MACD_SIGNAL	MACD_HIST	SlowK	SlowD	FastK	FastD	WILLR
	-0.094270	-0.112265	0.017995	36.563304	29.818914	59.108536	36.563304	-33.156022
	0.003331	-0.054467	0.057798	61.514327	40.248429	89.775530	61.514327	-30.172932
	0.074849	0.010191	0.064658	79.274371	59.117334	88.939048	79.274371	-28.745426
	0.148972	0.079581	0.069390	89.459349	76.749349	89.663469	89.459349	-25.334951

Exploratory Data Analysis



- A trends of increasing from year 2010 to year 2021
- There is a down trends starts from Feb 2021

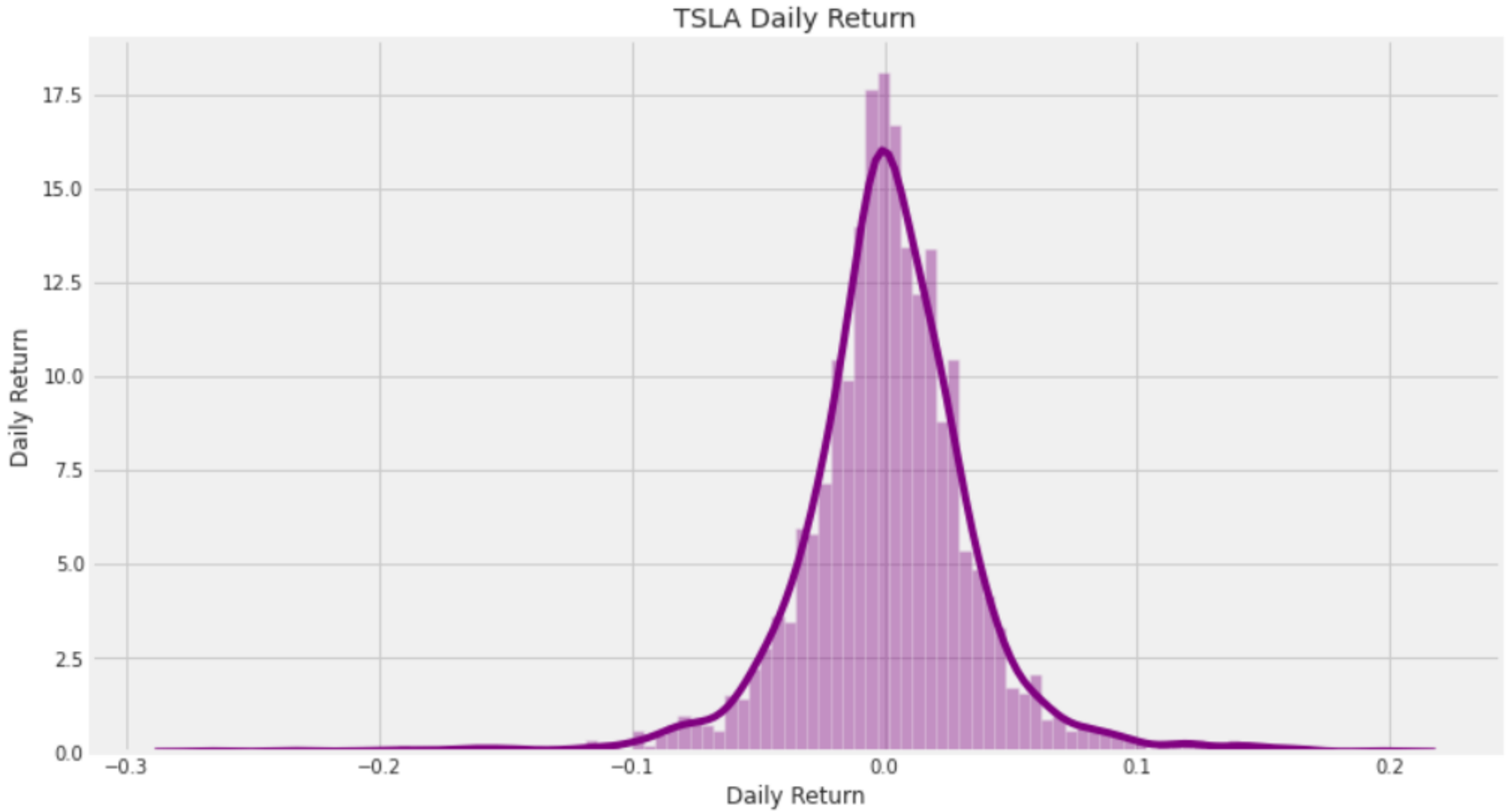
Exploratory Data Analysis



- Stock price from year 2010 to year 2021
- No seasonal pattern observed
- affected by various factor: financial news, situation of the company
- Unpredictable

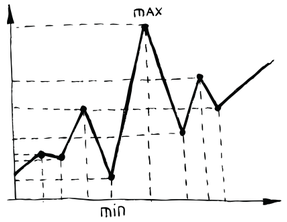
Exploratory Data Analysis

```
[ ]  data['Daily Return'] = (data['Adj Close'] - data['Adj Close'].shift(1))/data['Adj Close']
```



- -0.3 -> 0.2

Analytics Approach

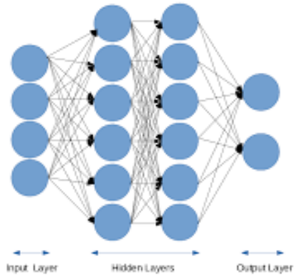


Linear Regression

Linear regression is a model that assumes a linear relationship between the input variable (x) and the single output variable (y)

Y = a*X + b

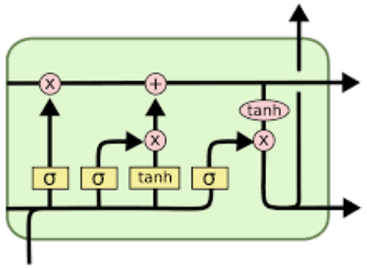
(Jason, 2016)



ANN

Adaptive system that learns by using interconnected nodes or neurons in a layered structure that resembles a human brain.

(MathWorks, n.d.)



LSTM

A type of recurrent neural network capable of learning order dependence in sequence prediction problems

(Jason, 2017)

Linear Regression

Without TIs

```
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score,mean_squared_error

lr = LinearRegression()
lr.fit(x_train, y_train)
predictions_1 = lr.predict(x_test)
r2 = r2_score(predictions_1,y_test)
rmse = np.sqrt(mean_squared_error(predictions_1,y_test))
print('R2 Score of Multiple Linear Regression Model',r2)
print('RMSE of Multiple Linear Regression Model',rmse)
```

```
R2 Score of Multiple Linear Regression Model 0.9008841304118193
RMSE of Multiple Linear Regression Model 0.21209126006918713
```

With TIs

```
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score,mean_squared_error

lr_2 = LinearRegression()
lr_2.fit(x_train, y_train)
predictions_4 = lr_2.predict(x_test)
r2 = r2_score(predictions_4,y_test)
rmse_4 = np.sqrt(mean_squared_error(predictions_4,y_test))
print('R2 Score of Multiple Linear Regression Model',r2)
print('RMSE of Multiple Linear Regression Model',rmse_4)
```

```
R2 Score of Multiple Linear Regression Model 0.8746170416992565
RMSE of Multiple Linear Regression Model 0.21863547713920797
```

Artificial Neural Network

Without TIs

```
from keras.models import Sequential
from keras.layers import Dense, Dropout

ann = Sequential()
ann.add(Dense(128, input_dim=x_train.shape[1], kernel_initializer='normal', activation='relu'))
ann.add(Dropout(0.3))
ann.add(Dense(64, kernel_initializer='normal', activation='relu'))
ann.add(Dropout(0.3))
ann.add(Dense(1, kernel_initializer='normal'))

# Compile model
ann.compile(loss='mean_squared_error', optimizer='adam')
```

```
history = ann.fit(x=x_train, y=y_train, batch_size=32, epochs=50, verbose=1)
```

With TIs

```
from keras.models import Sequential
from keras.layers import Dense, Dropout

ann_2 = Sequential()
ann_2.add(Dense(128, input_dim=x_train.shape[1], kernel_initializer='normal', activation='relu'))
ann_2.add(Dropout(0.3))
ann_2.add(Dense(64, kernel_initializer='normal', activation='relu'))
ann_2.add(Dropout(0.3))
ann_2.add(Dense(1, kernel_initializer='normal'))

# Compile model
ann_2.compile(loss='mean_squared_error', optimizer='adam')
```

```
history = ann_2.fit(x=x_train, y=y_train, batch_size=32, epochs=50, verbose=1)
```

Long short-term memory

Without TIs

```
from keras.models import Sequential
from keras.layers import Dense, LSTM

# reshape dataframe
x_train = np.reshape(x_train, (x_train.shape[0], 1, x_train.shape[1]))

#Process the data for LSTM
trainX =np.array(x_train)
testX =np.array(x_test)
#X_train = trainX.reshape(x_train.shape[0], 1, x_train.shape[1])
#X_test = testX.reshape(x_test.shape[0], 1, x_test.shape[1])

#Building the LSTM Model
lstm = Sequential()
lstm.add(LSTM(32, input_shape=(1, trainX.shape[2]), activation='relu', return_sequences=False))
lstm.add(Dense(1))
lstm.compile(loss='mean_squared_error', optimizer='adam')

#Model Training
history=lstm.fit(x_train, y_train, epochs=100, batch_size=8, verbose=1, shuffle=False)
```

With TIs

```
# reshape dataframe
x_train = np.reshape(x_train, (x_train.shape[0], 1, x_train.shape[1]))

#Process the data for LSTM
trainX =np.array(x_train)
testX =np.array(x_test)
#X_train = trainX.reshape(x_train.shape[0], 1, x_train.shape[1])
#X_test = testX.reshape(x_test.shape[0], 1, x_test.shape[1])

#Building the LSTM Model
lstm_2 = Sequential()
lstm_2.add(LSTM(32, input_shape=(1, trainX.shape[2]), activation='relu', return_sequences=False))
lstm_2.add(Dense(1))
lstm_2.compile(loss='mean_squared_error', optimizer='adam')

#Model Training
history=lstm_2.fit(x_train, y_train, epochs=100, batch_size=8, verbose=1, shuffle=False)
```

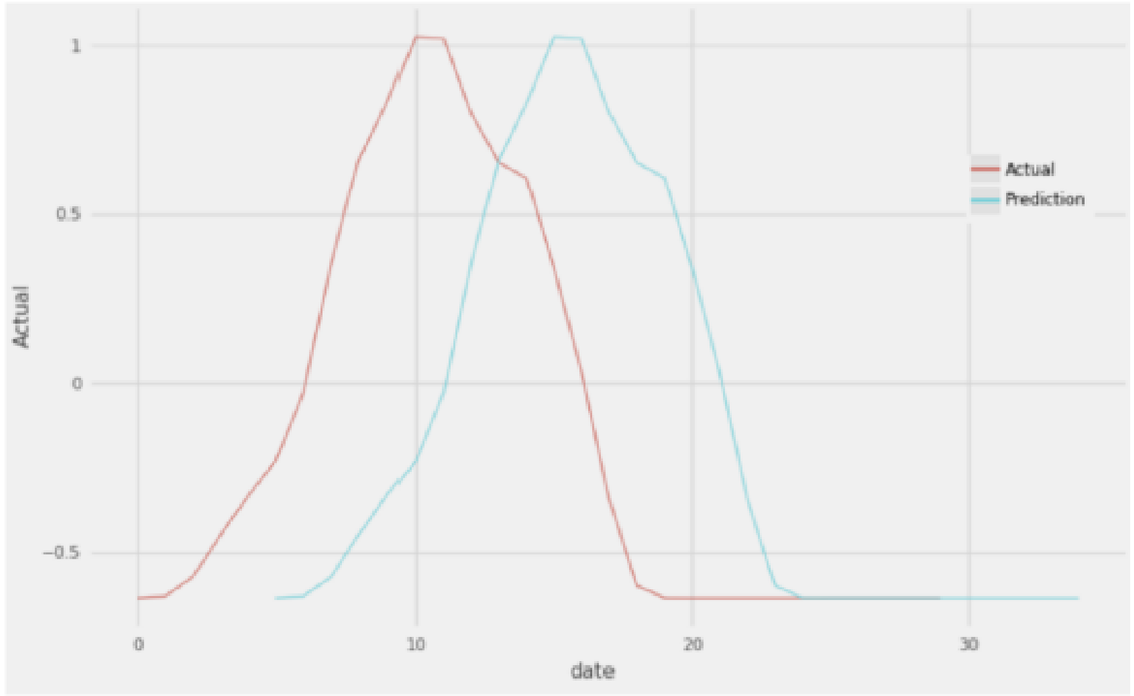
Model Evaluation

RMSE, R2 score

Use to evaluate the linear regression model performance and the strength between the response and predictor variables

(Willie, 2021)

Cross-correlation lag & Spearmans Correlation



Example — Lag Correlation

(Oguntayo, 2020)

Model Evaluation

Linear Regression

Without Technical Indicators perform better

	Without	With
RMSE	0.21	0.21
CC_Lag0	0.95	0.94
Spearman	0.93	0.92

ANN

With Technical Indicators perform better

	Without	With
RMSE	0.95	0.99
CC_Lag0	0.88	0.95
Spearman	0.83	0.93

LSTM

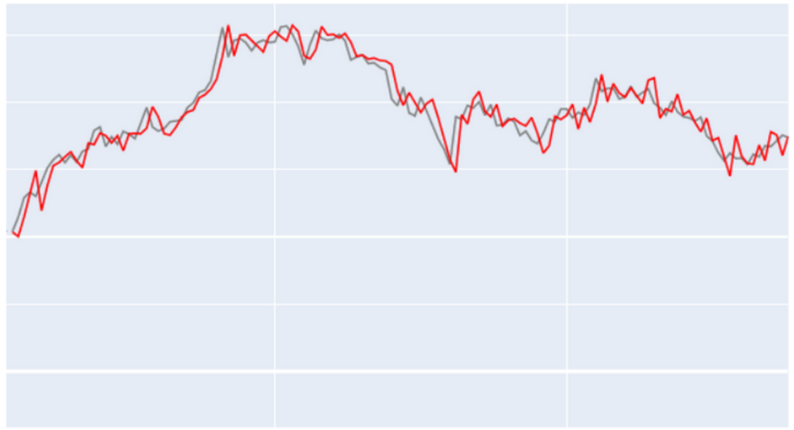
With Technical Indicators perform better

	Without	With
RMSE	1.21	0.79
CC_Lag0	0.78	0.92
Spearman	0.68	0.91

Result & Discussion

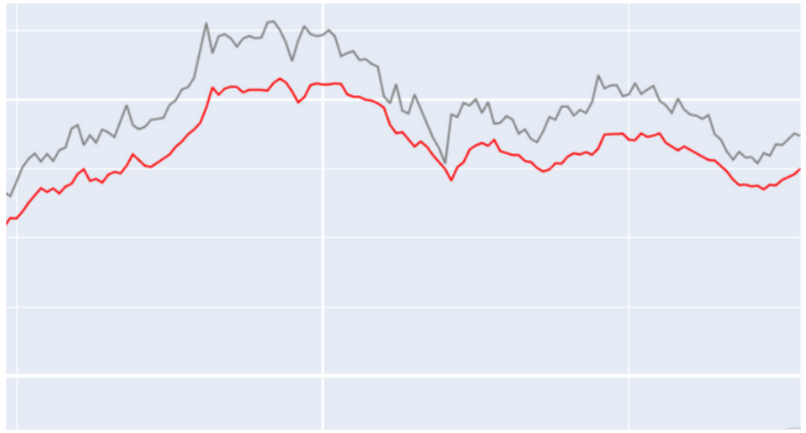
Linear Regression (without TIs)

low RMSE, low CC_Lag and spearmans



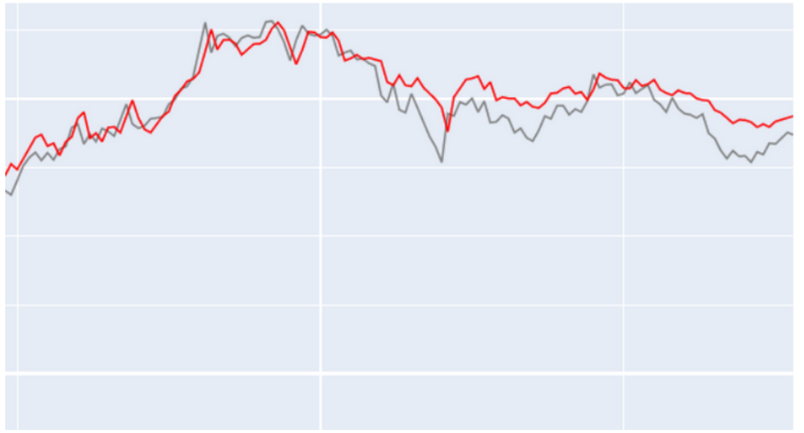
ANN (with TIs)

high RMSE, high CC_Lag and spearmans



LSTM (with TIs)

medium RMSE, CC_Lag and spearmans



Conclusion

ANN

Best model

Technical Indicators

Technical Indicators help in forecasting the future directional movement

Research Objectives

Reached !

Tools and coding

Google Colabotary

A Python development environment that runs in the browser using Google Cloud.

(Google, 2022)

Python Dash

A Python framework created by plotly for creating interactive web applications.

(Plotly, 2017)

Heroku app

A PaaS that enables developers to build, run, and operate applications entirely in the cloud.

(Heroku, n.d.)

TradingViz

Best Site for Stock Analysis

Solution 1

Present relevant techniques in forecasting stock price movement

Solution 2

Present the use of technical indicators in forecasting the stock price movement

Solution 3

Develop the data product of the stock price forecasting system

Target Stakeholders

Day Trader

A day trader is a type of trader who executes a relatively large volume of short and long trades to capitalize on intraday market price action. The goal is to profit from very short-term price movements. Day traders can also use leverage to amplify returns, which can also amplify losses.

A day trader is primarily concerned with the price action characteristics of a stock. This is unlike investors, who use fundamental data to analyze the long-term growth potential of a company to decide whether to buy, sell or hold its stock.
(James, 2021)

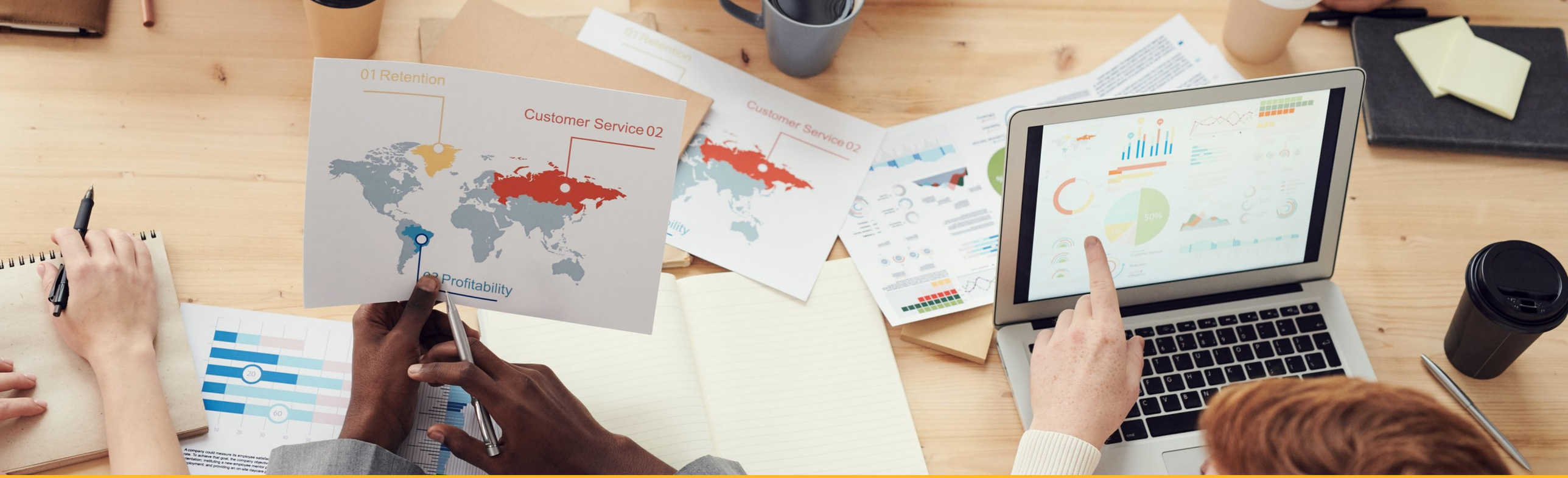
How we help them

Day Trader

Give them insight of what is the stock price movement tomorrow

If the model forecasted tomorrow stock price will be moving DOWN, then the trader can sell the stock

If the model forecasted tomorrow stock price will be moving UP, then the trader can sell the stock later to increase the profit.



TradingViz

Demonstration of Data Product:
https://stock-market-forecast.herokuapp.com/

Home page: TradingViz

User manual

TradingViz

Tommorow's Stock Directional Movement Forecast

Time Series Forecasting Stock Price Directional Movement: Forecast the next direction! This application is created for Day Trader to decide when should they buy/sell the stock by observing the graph of the directional (UP/DOWN) forecasted by the model.

Ticker TSLA

Chart Type candlestick

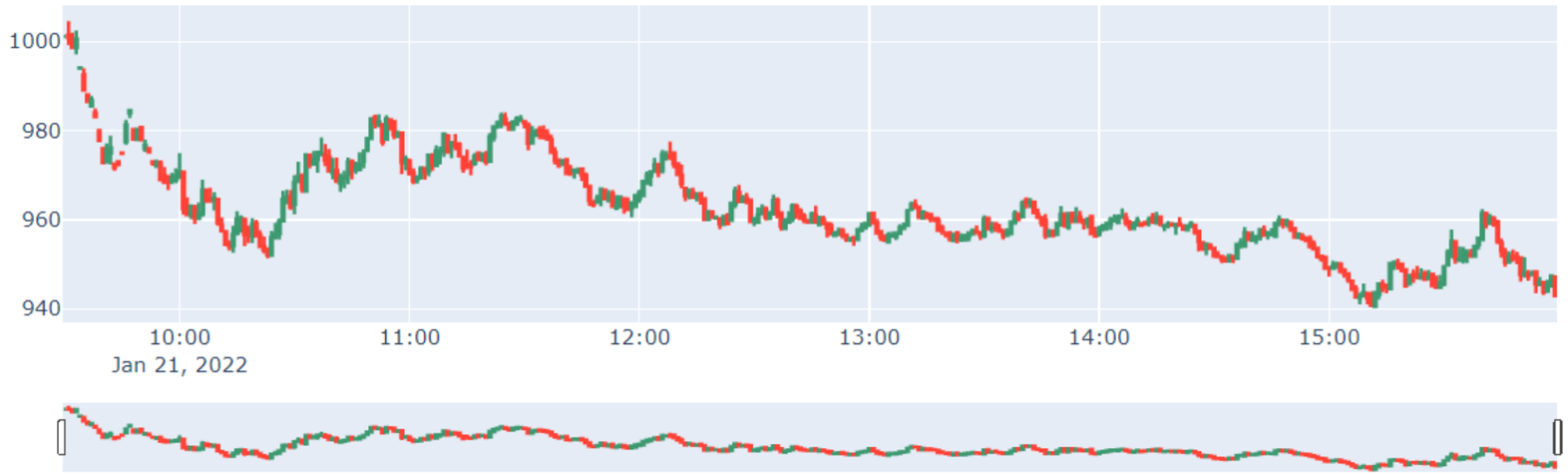
Indicators Select...

Date Range 01/21/2022 → 01/21/2022

Malaysia time: 15:20:13

US time: 02:20:13

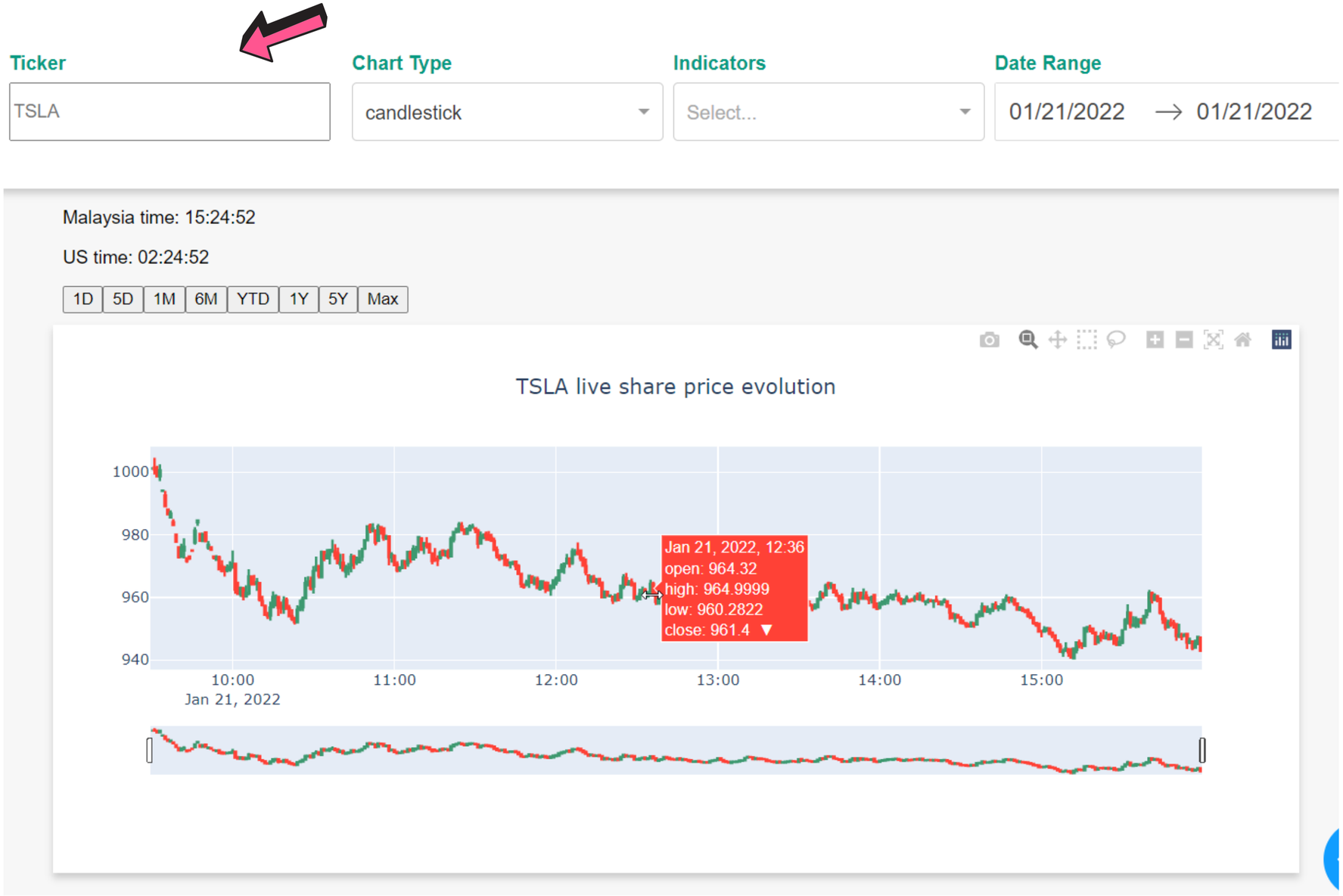
1D | 5D | 1M | 6M | YTD | 1Y | 5Y | Max



Home page

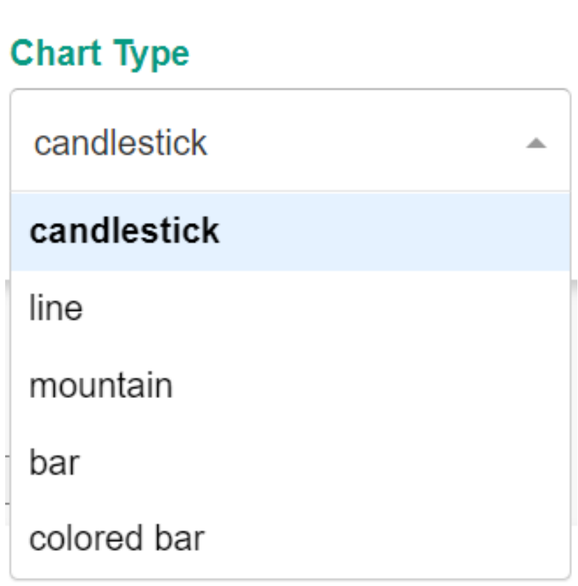
- Input box for user to key the symbol of the stock they wish to visualize, for example AAPL for Apple Inc.

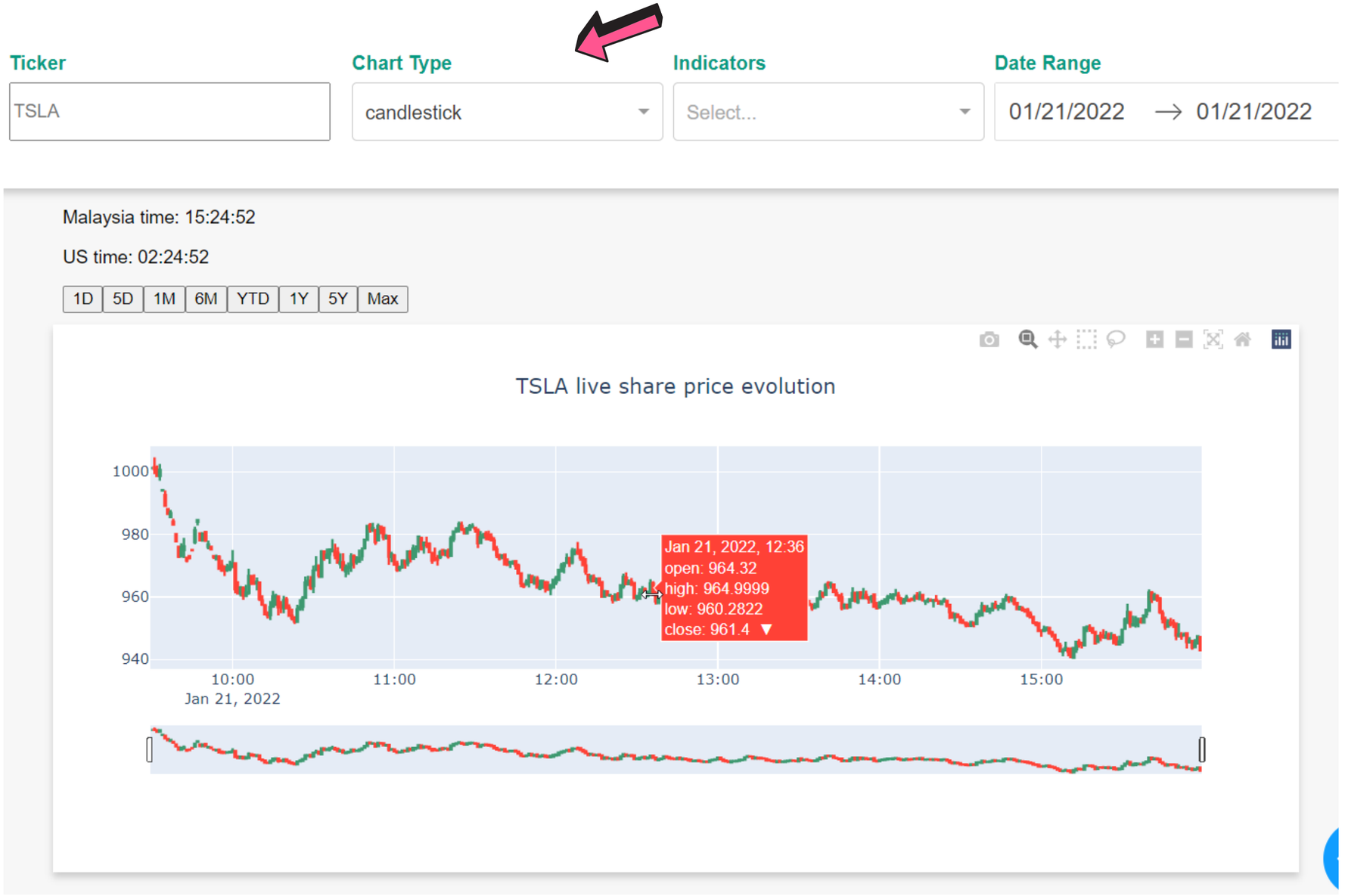
Symbol	Name
AAPL	Apple Inc. Common Stock
MSFT	Microsoft Corporation Common Stock
GOOGL	Alphabet Inc. Class A Common Stock
GOOG	Alphabet Inc. Class C Capital Stock
AMZN	Amazon.com, Inc. Common Stock
TSLA	Tesla, Inc. Common Stock
FB	Meta Platforms, Inc. Class A Common Stock



Home page

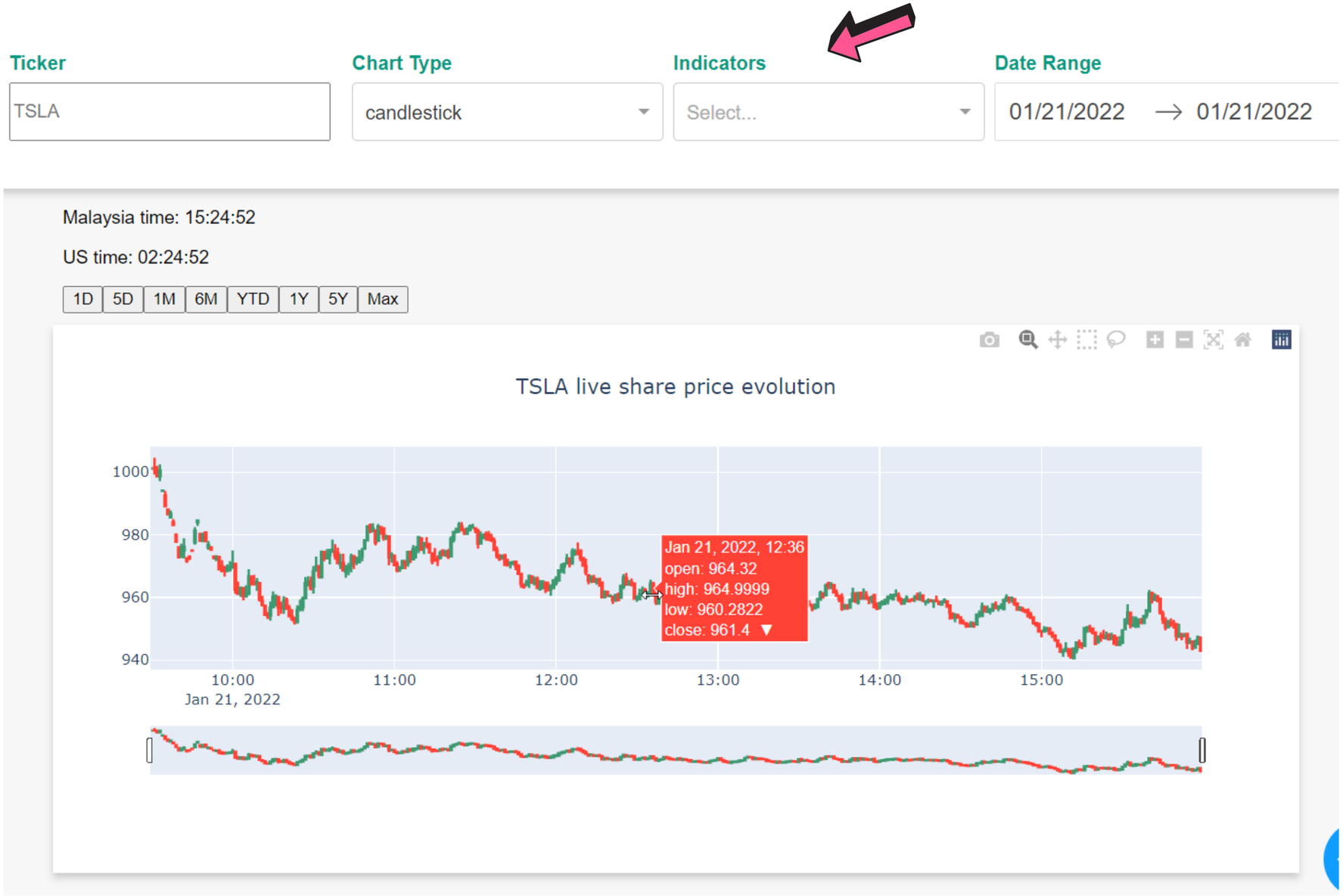
- Single dropdown choice for user to choose the chart type to be plotted on the graph

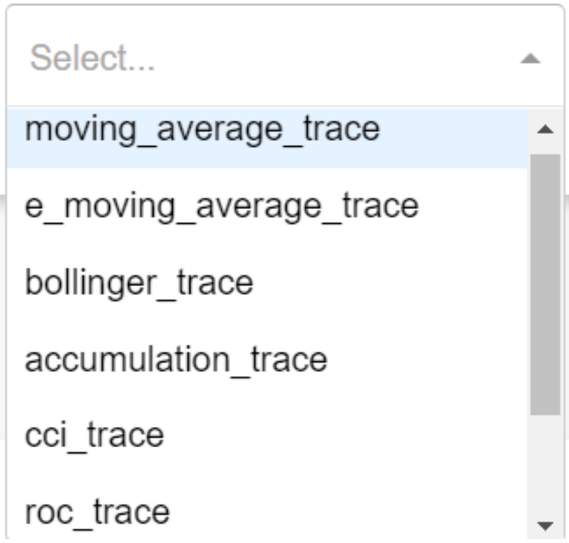




Home page

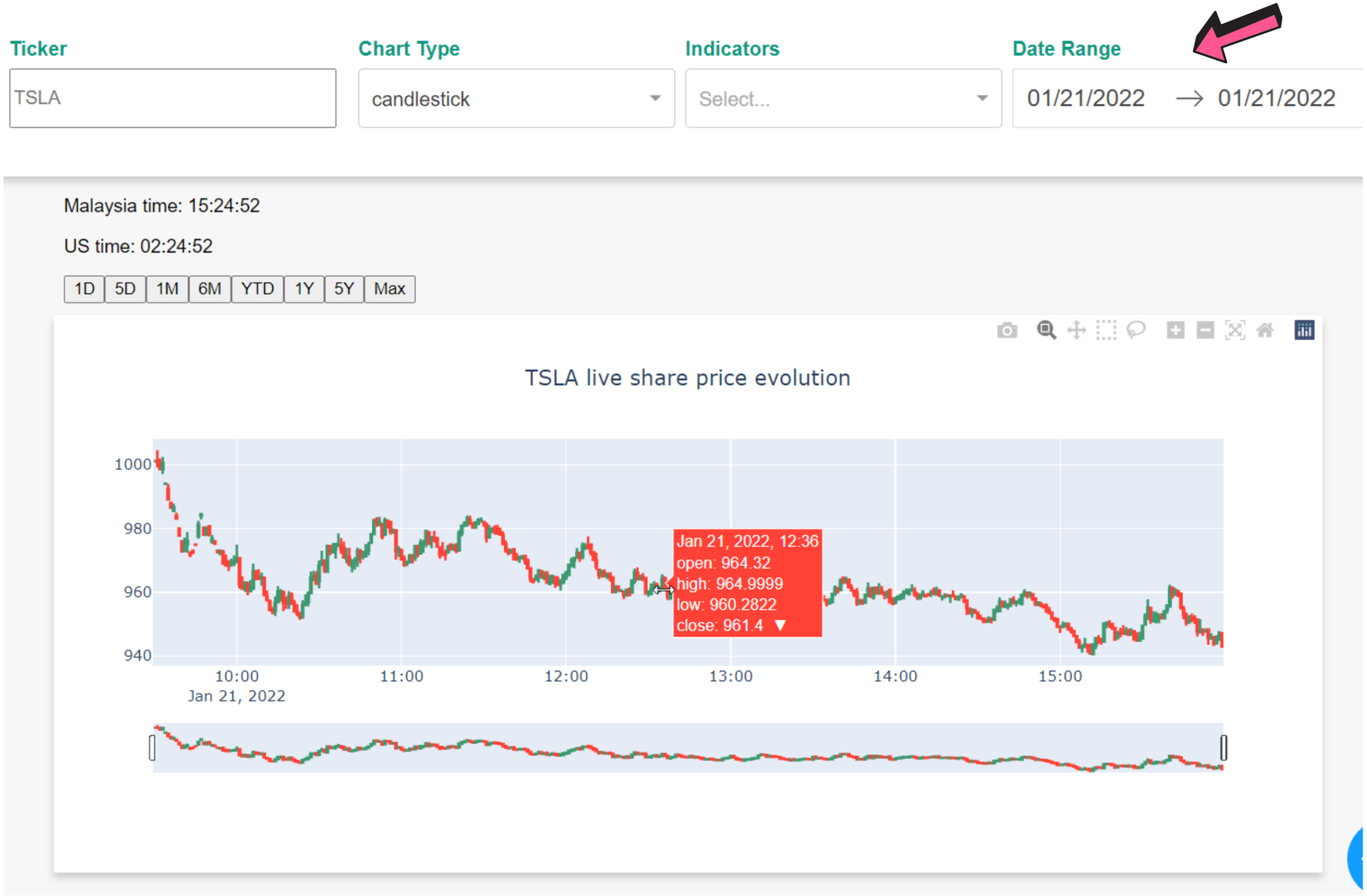
- Multi dropdown choice for user to choose the technical indicators they wish to add on the graph





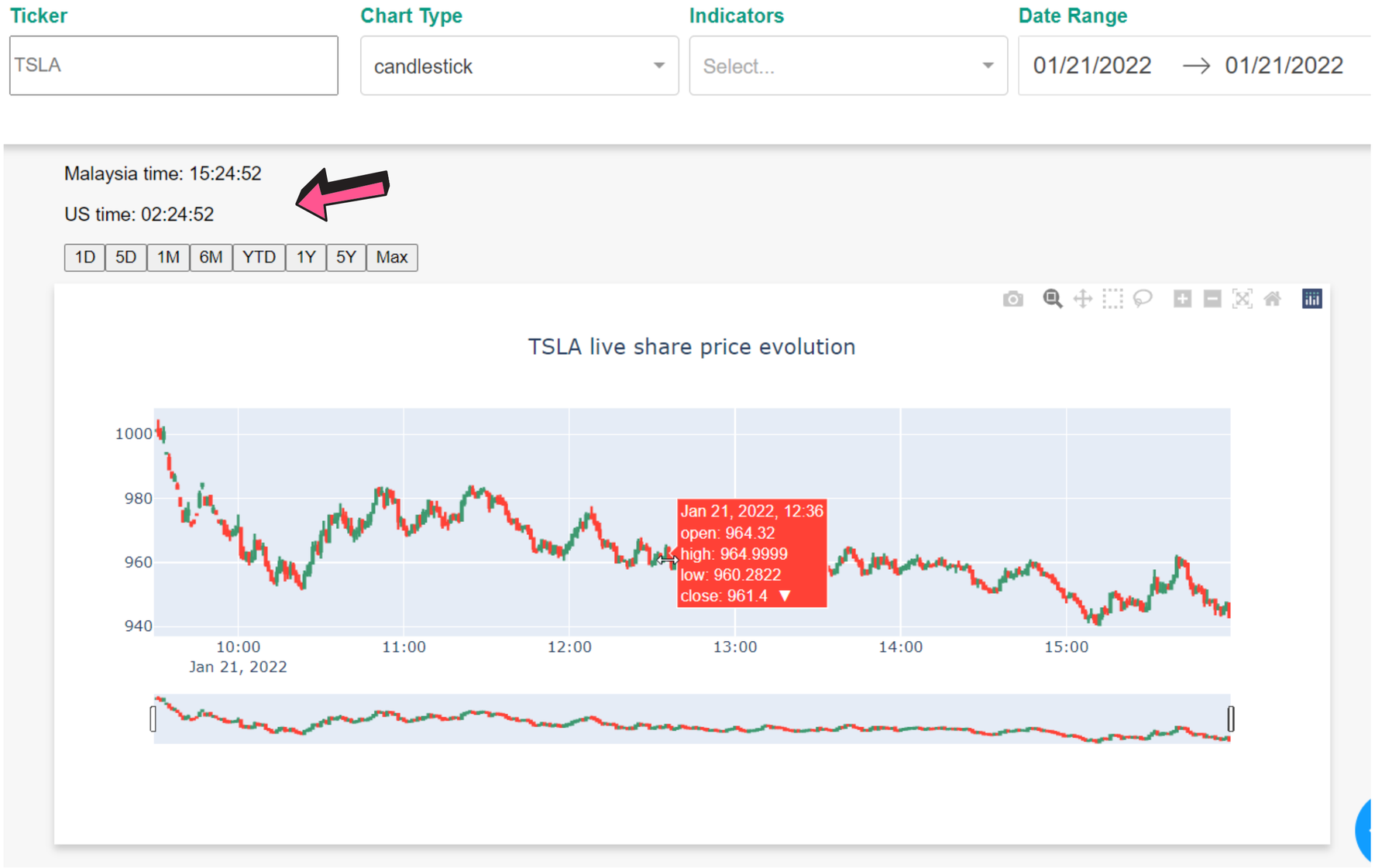
Home page

- Date Range Picker
- User could pick any day, month, year.



Home page

- Real time clock in Malaysia and US updated in second.
- Stock operate from US time 9.30am to 4.00pm.

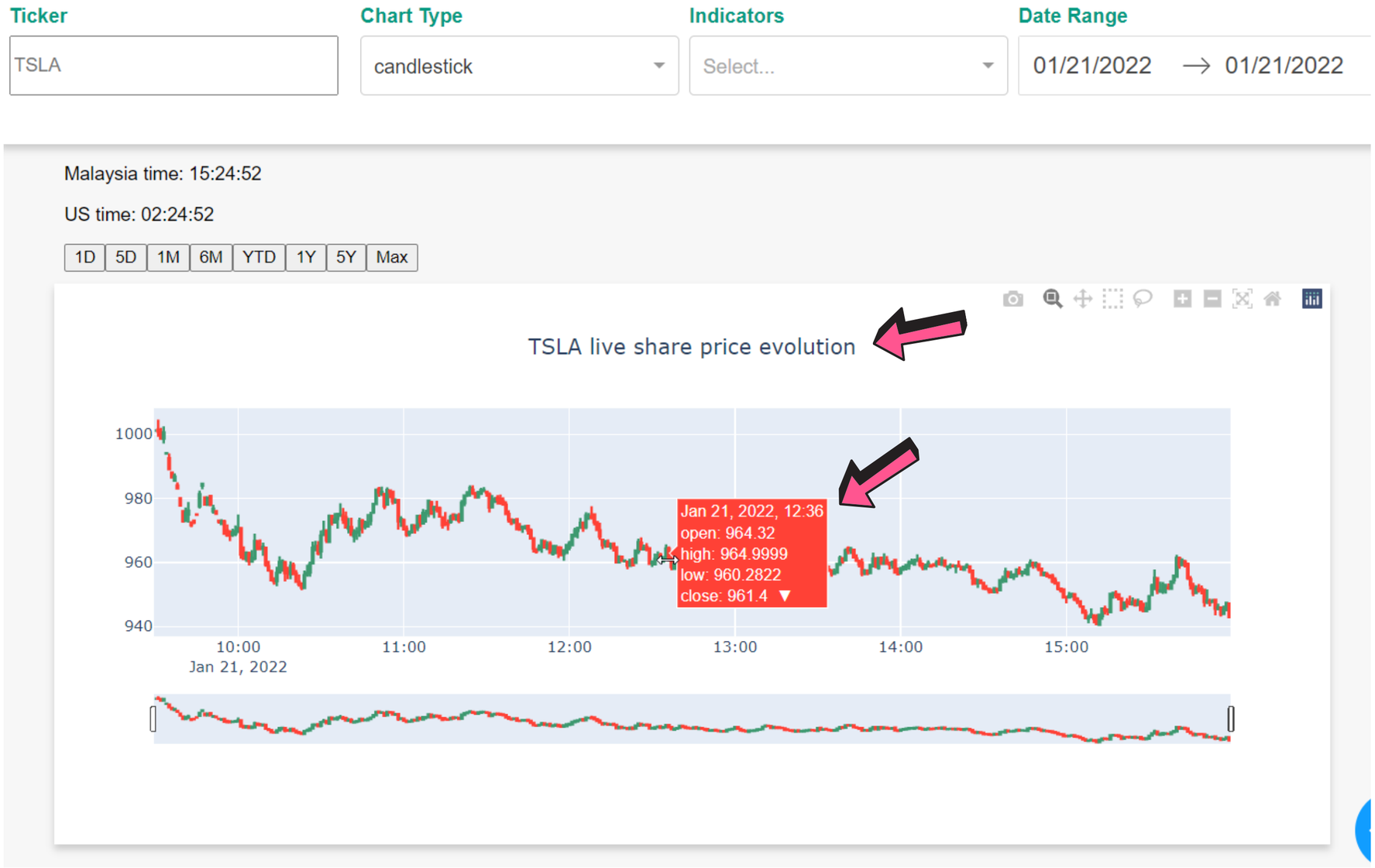


Home page

- Shortcut button to let user select the duration of the stock price to be shown on the graph
- In default the graph is plotted in daily mode which is '1D'.

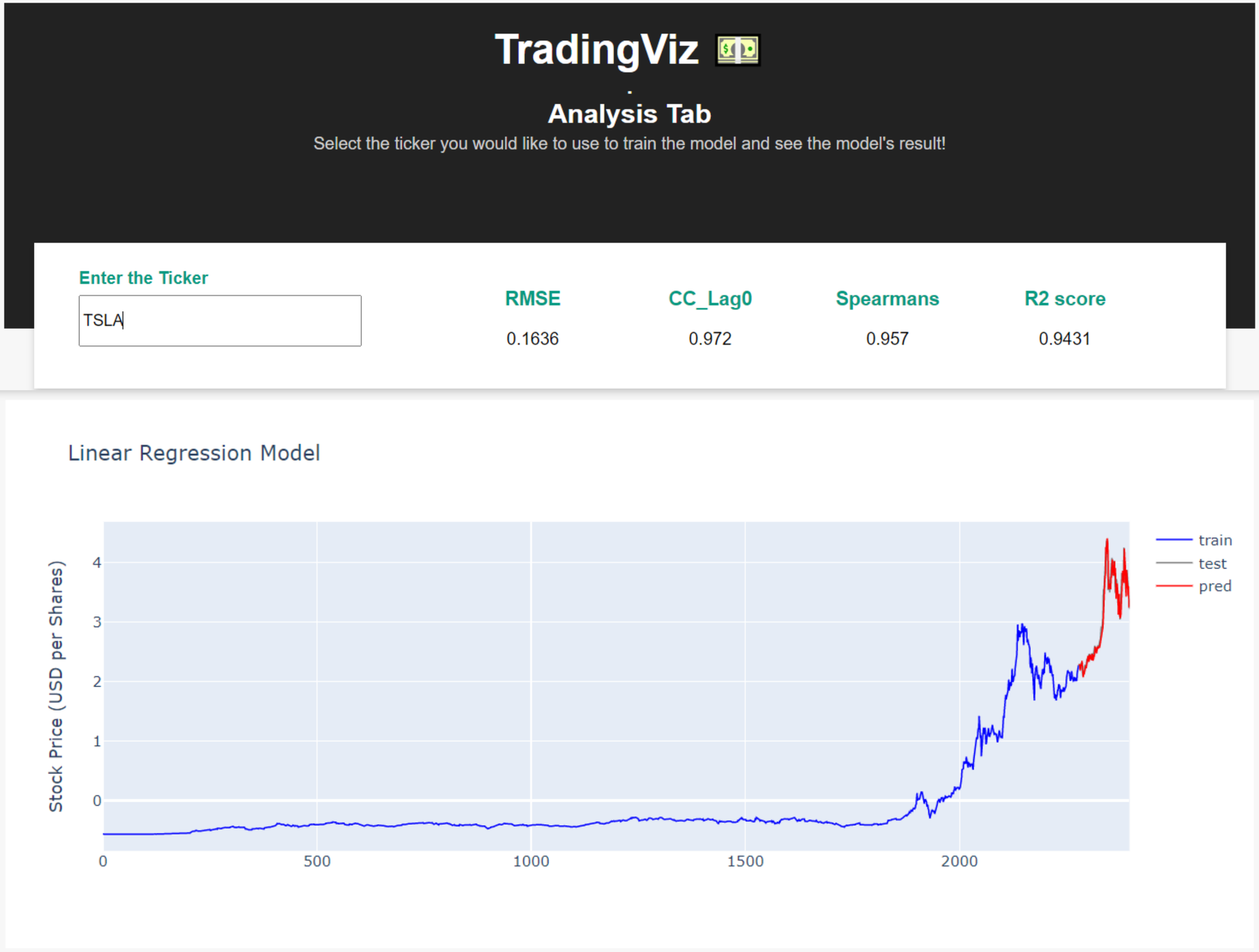
Home page

- Home page contains of interactive and real time charts that automatically update themselves every minutes (from US time 9.30am to 4.00pm)



Analysis page: TradingViz

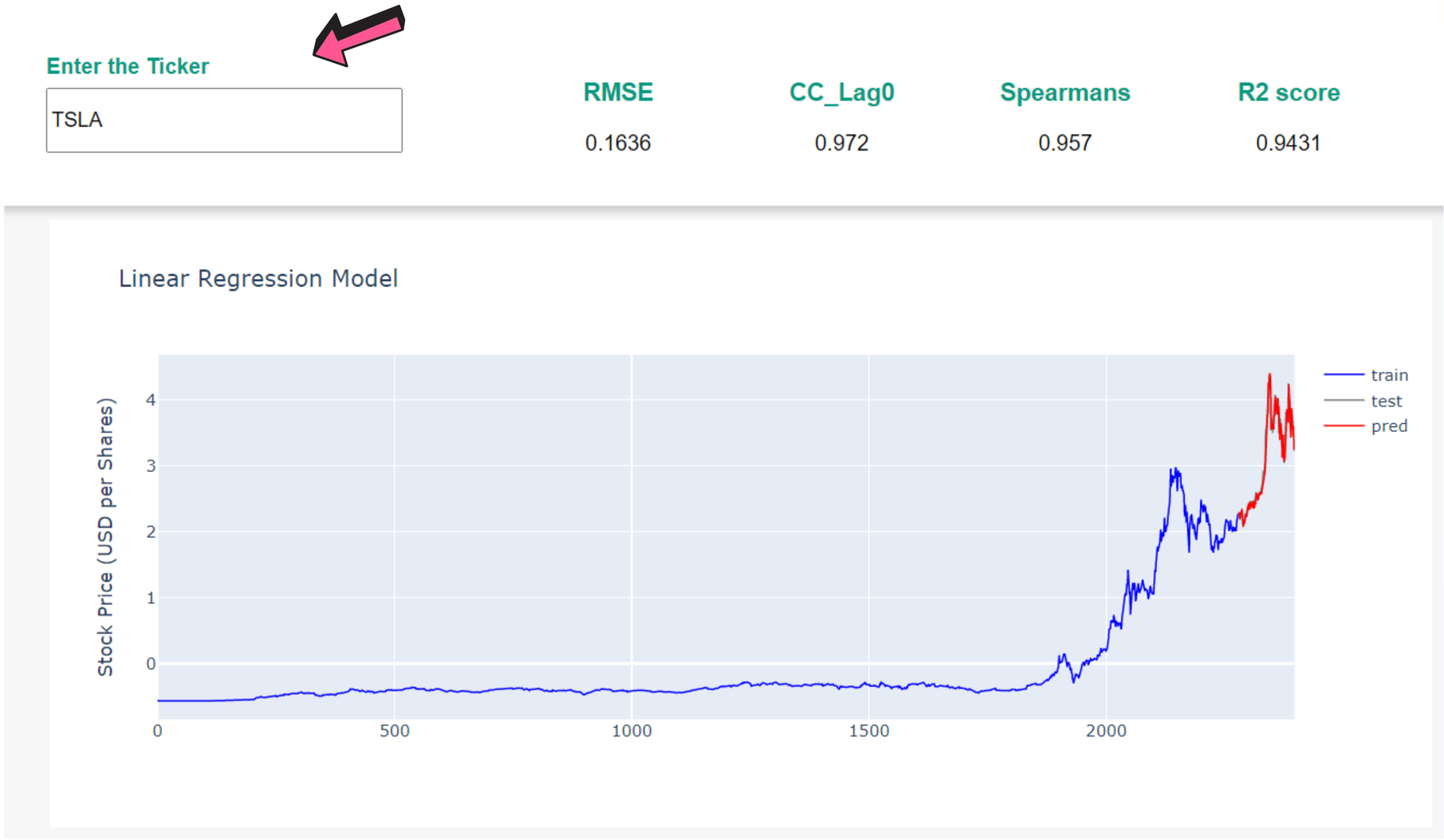
User manual



Analysis page

- Input box for user to key the symbol of the stock they wish to analyze, for example AAPL for Apple Inc.

Symbol	Name
AAPL	Apple Inc. Common Stock
MSFT	Microsoft Corporation Common Stock
GOOGL	Alphabet Inc. Class A Common Stock
GOOG	Alphabet Inc. Class C Capital Stock
AMZN	Amazon.com, Inc. Common Stock
TSLA	Tesla, Inc. Common Stock
FB	Meta Platforms, Inc. Class A Common Stock



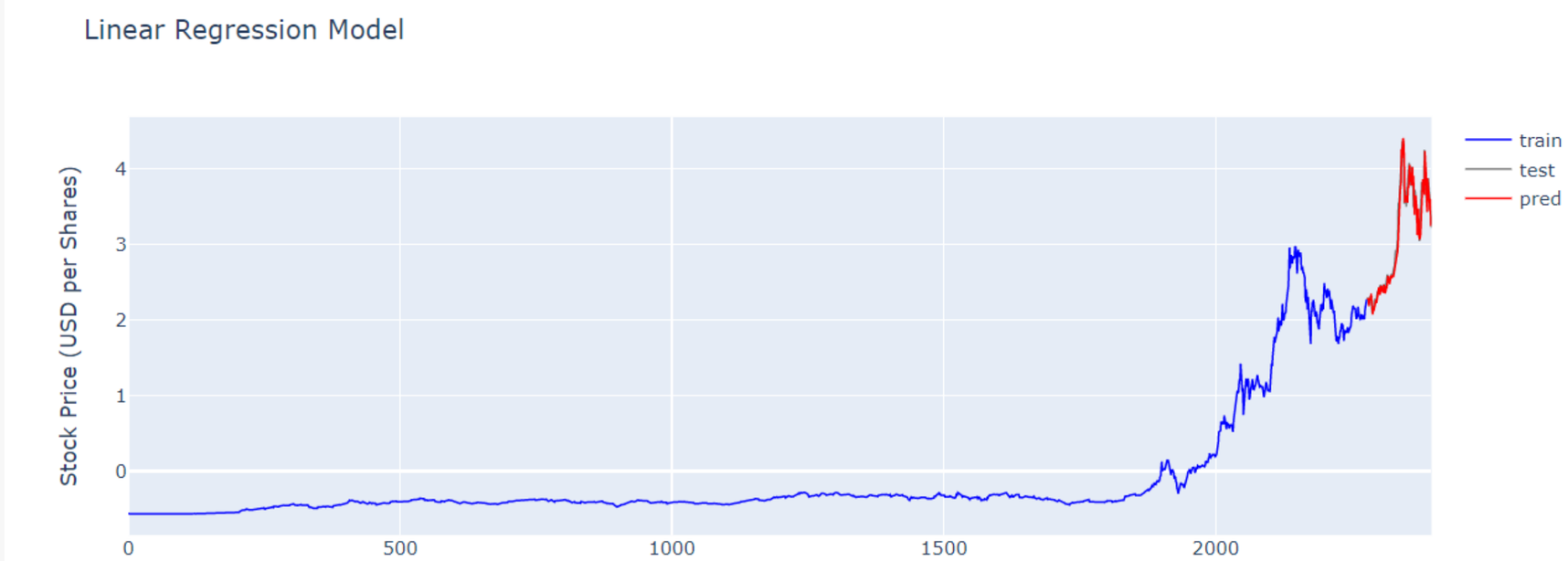
Analysis page

- RMSE value

Enter the Ticker

TSLA

RMSE	CC_Lag0	Spearmans	R2 score
0.1636	0.972	0.957	0.9431



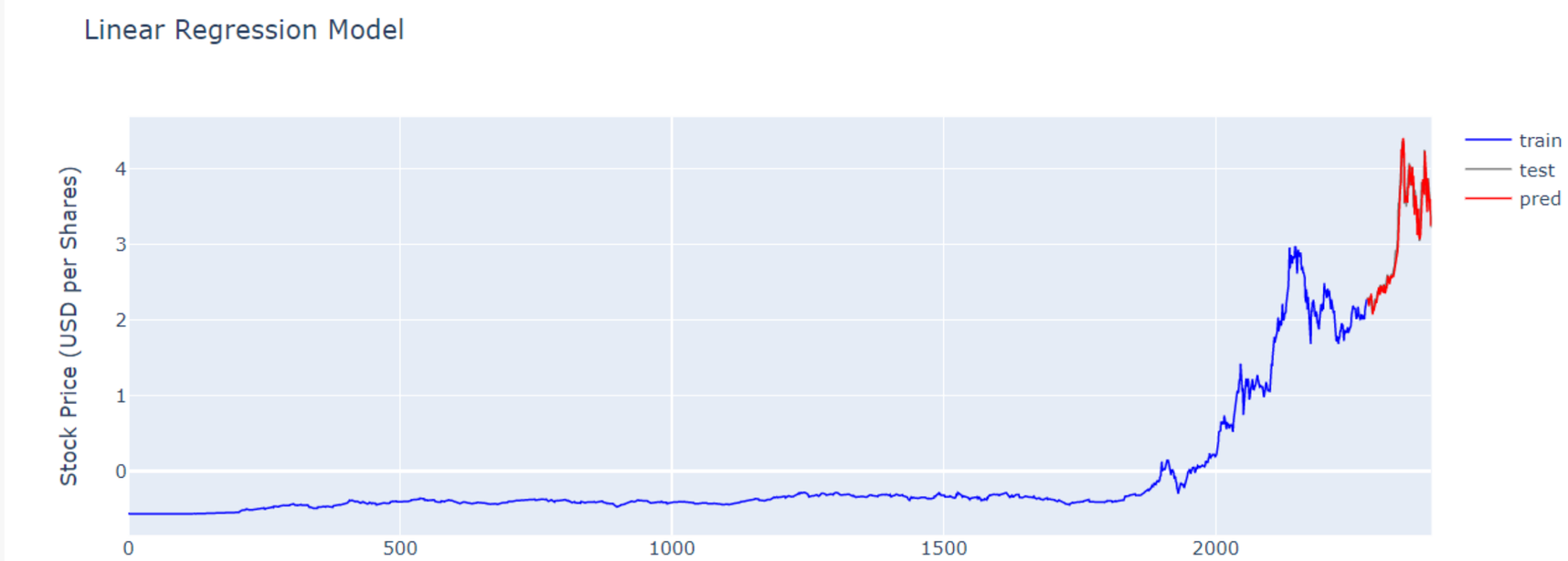
Analysis page

- CC_Lag0 value

Enter the Ticker

TSLA

RMSE	CC_Lag0	Spearmans	R2 score
0.1636	0.972	0.957	0.9431

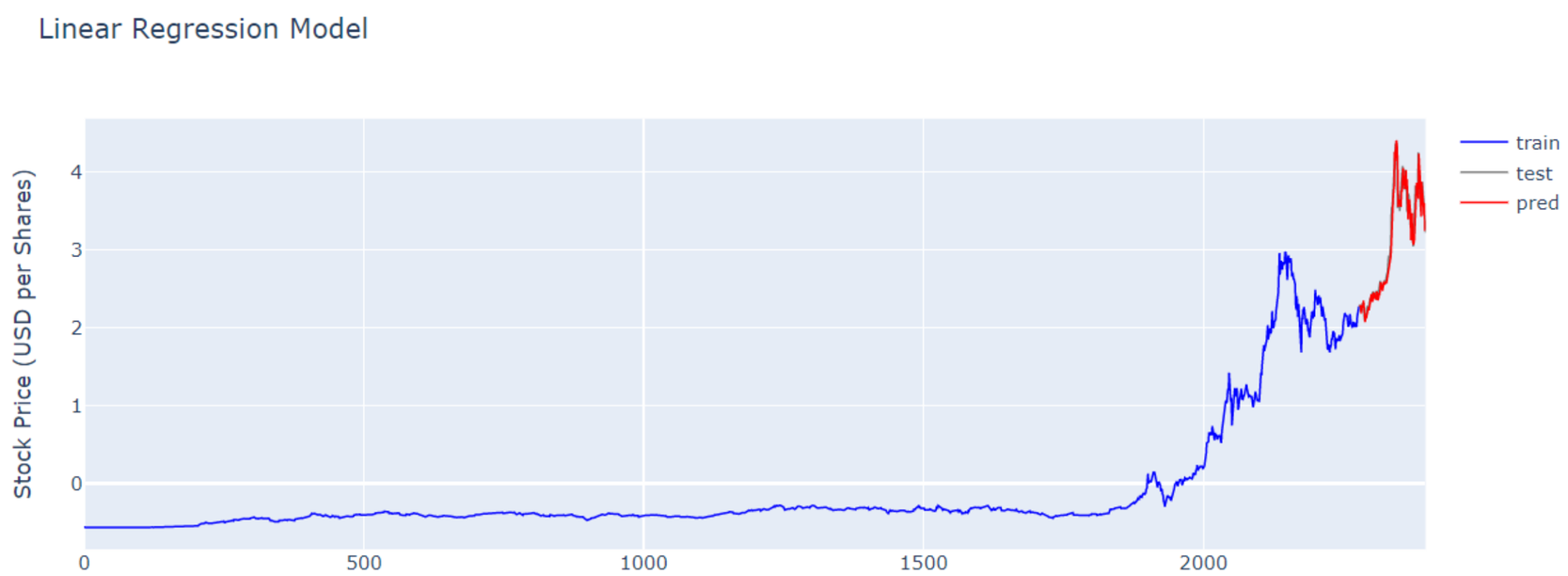


Analysis page

Enter the Ticker

TSLA

RMSE	CC_Lag0	Spearmans	R2 score
0.1636	0.972	0.957	0.9431



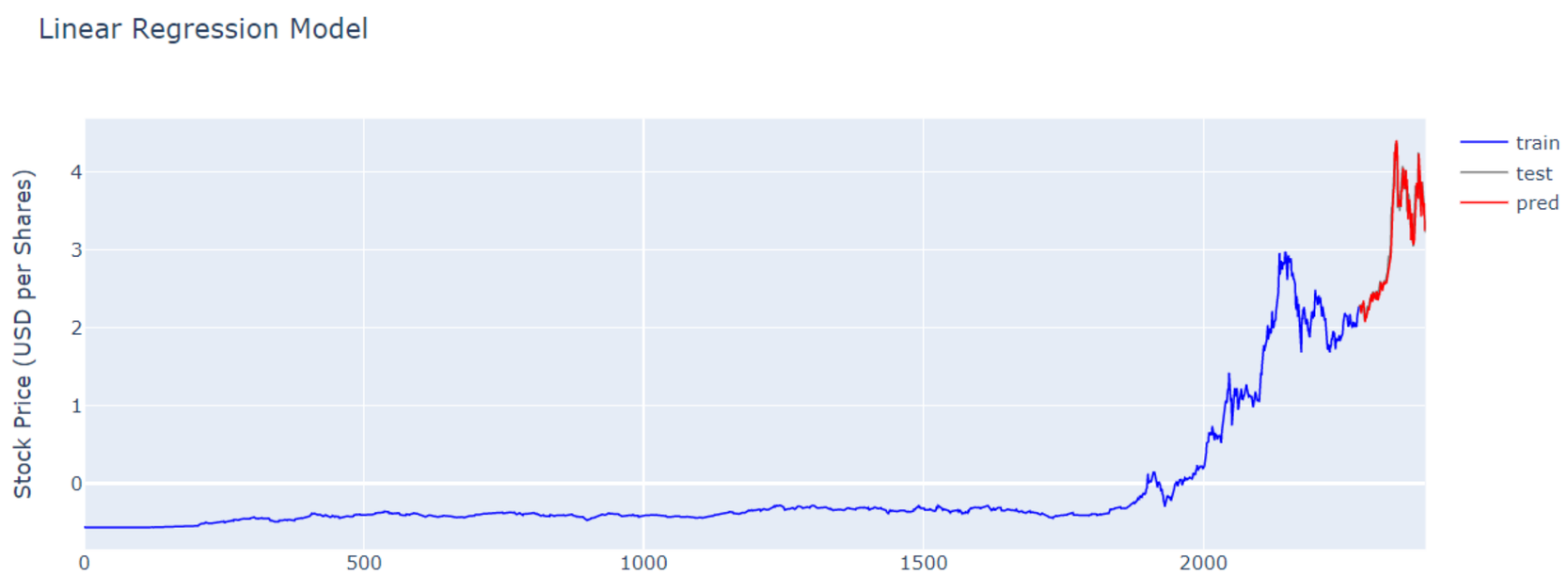
- Spearmans value

Analysis page

Enter the Ticker

TSLA

RMSE	CC_Lag0	Spearmans	R2 score
0.1636	0.972	0.957	0.9431



- R2 score value

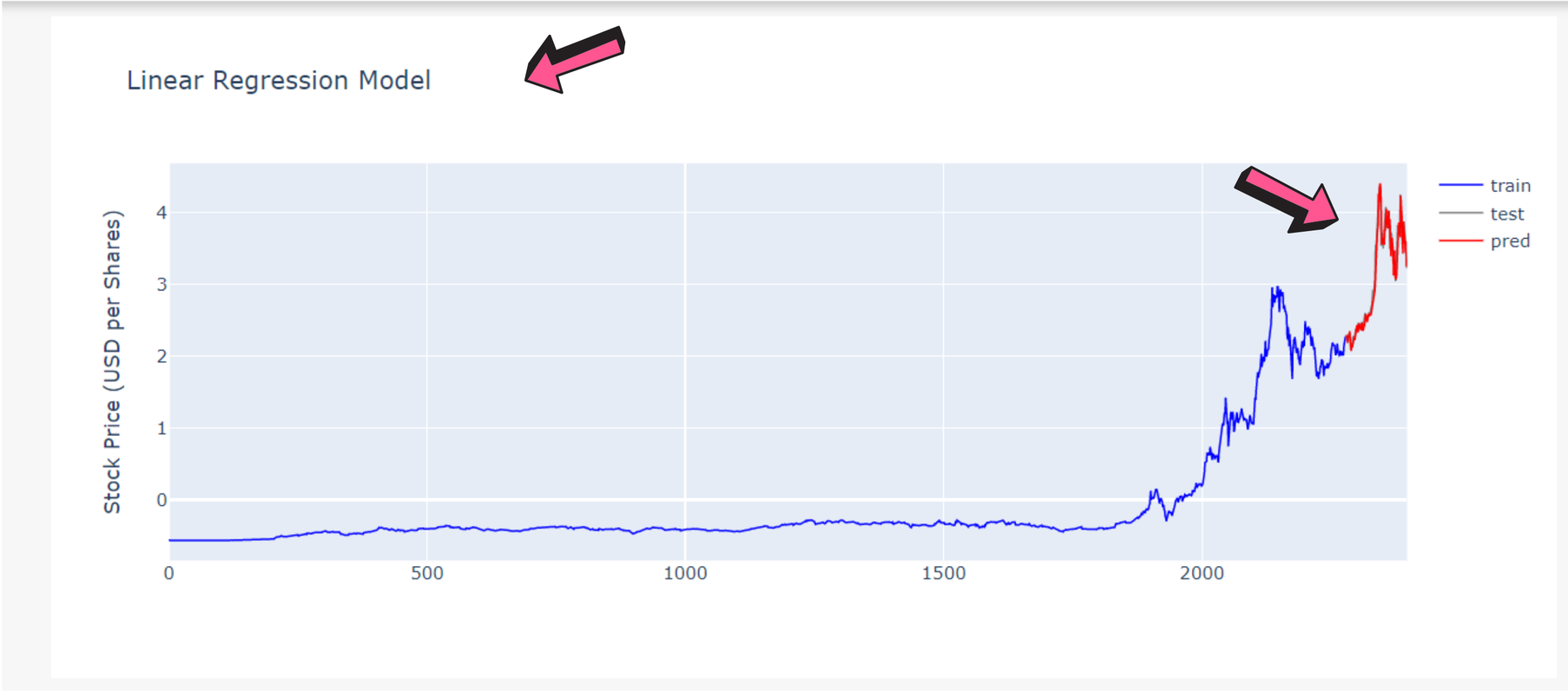
Analysis page

- Interactive chart plotted with train, test and predicted values

Enter the Ticker

TSLA

RMSE	CC_Lag0	Spearmans	R2 score
0.1636	0.972	0.957	0.9431



Forecast page: TradingViz

User manual

TradingViz

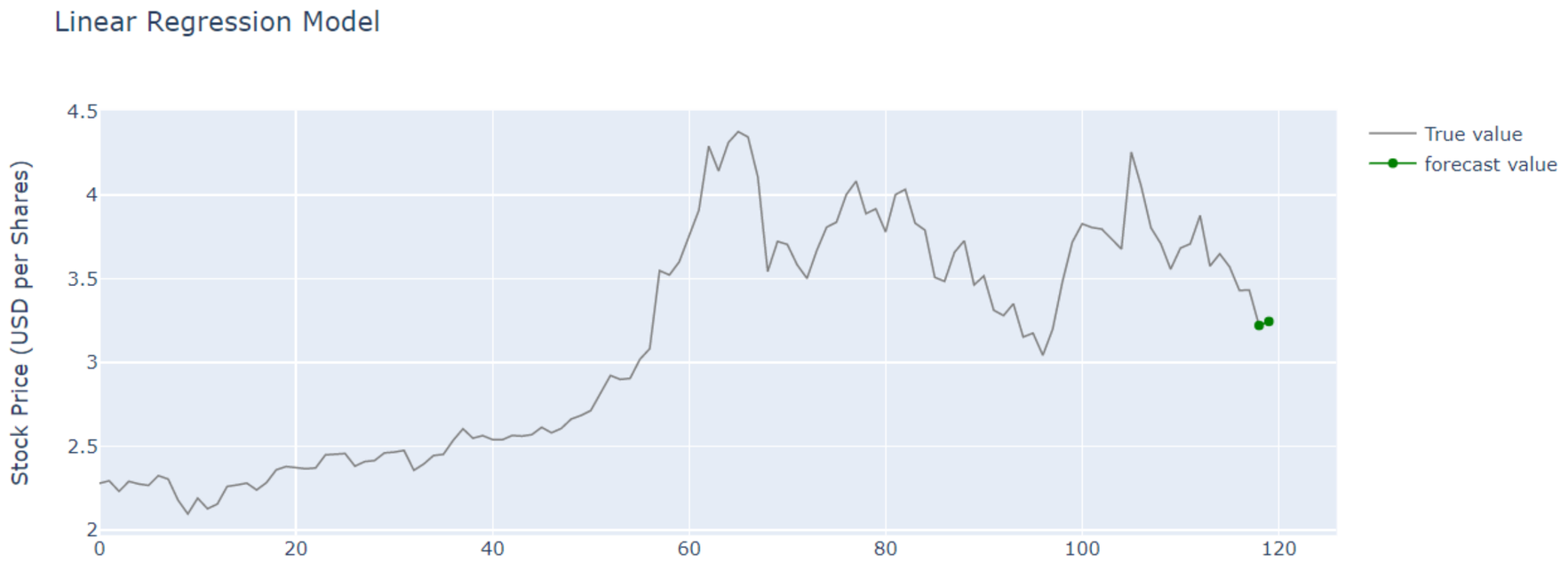
Forecasting Tab

Select the ticker you would like to forecast and see the result!

Enter the Ticker

TSLA

RMSE	R2 score	Next Move	Return
0.1636	0.9431	UP	0.0075



Forecast page

- Input box for user to key the symbol of the stock they wish to analyze, for example AAPL for Apple Inc.

Symbol	Name
AAPL	Apple Inc. Common Stock
MSFT	Microsoft Corporation Common Stock
GOOGL	Alphabet Inc. Class A Common Stock
GOOG	Alphabet Inc. Class C Capital Stock
AMZN	Amazon.com, Inc. Common Stock
TSLA	Tesla, Inc. Common Stock
FB	Meta Platforms, Inc. Class A Common Stock

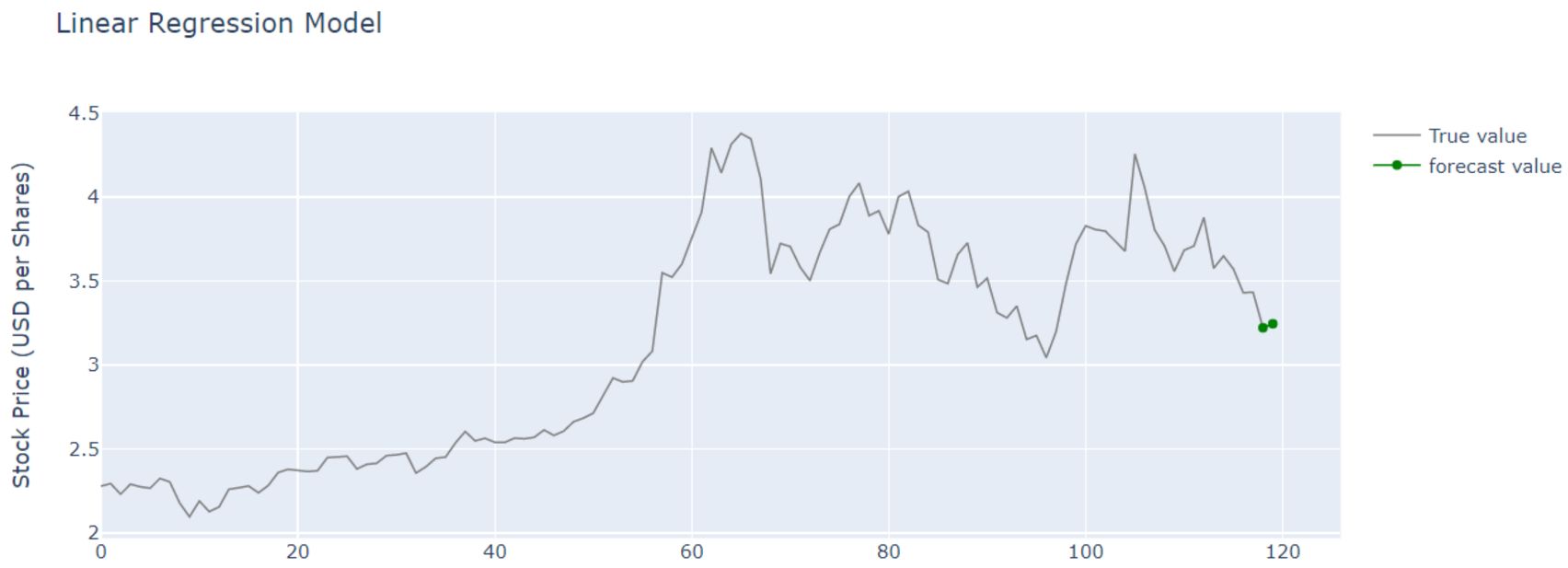
Forecast page

- RMSE value

Enter the Ticker

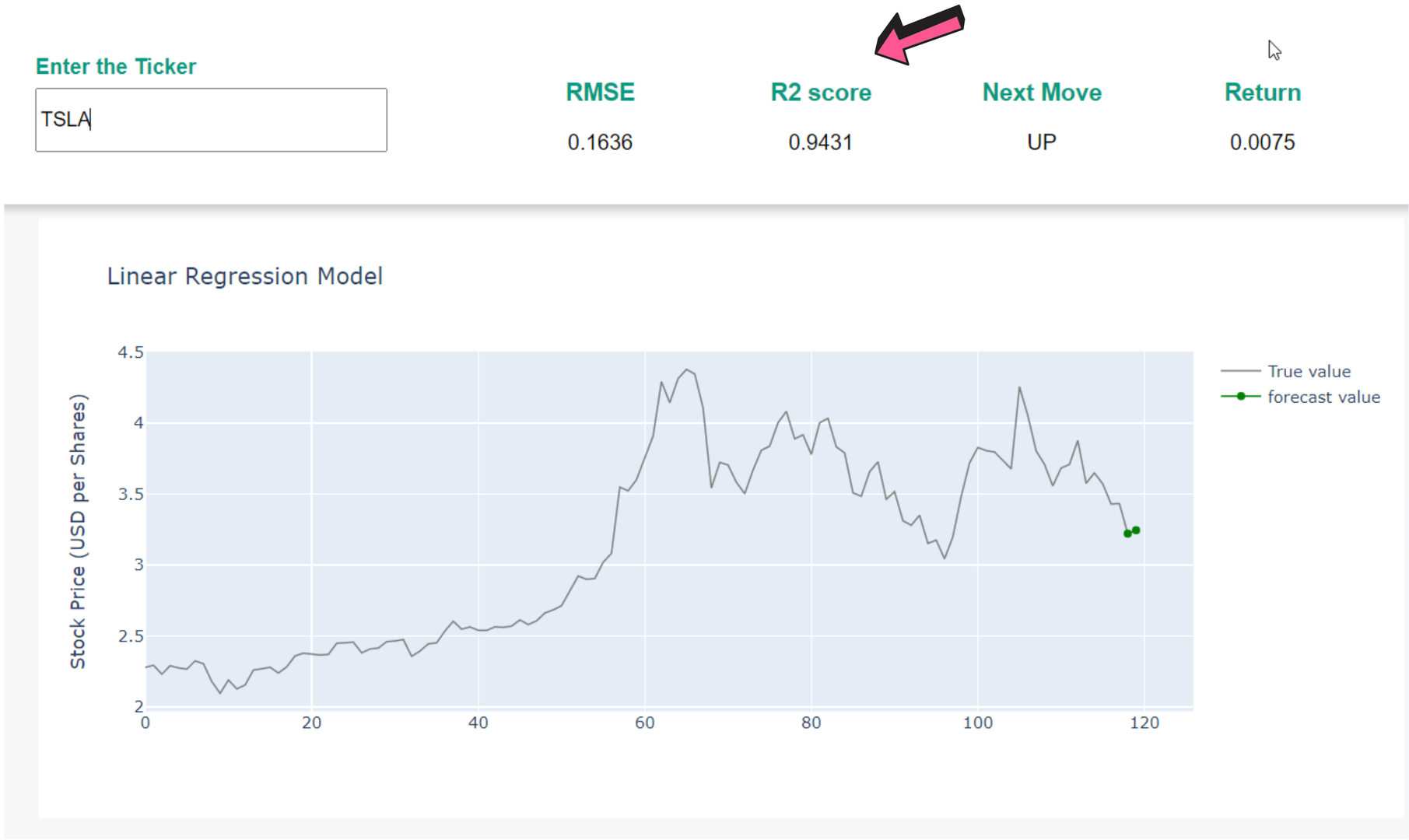
TSLA

RMSE	R2 score	Next Move	Return
0.1636	0.9431	UP	0.0075



Forecast page

- R2 score value



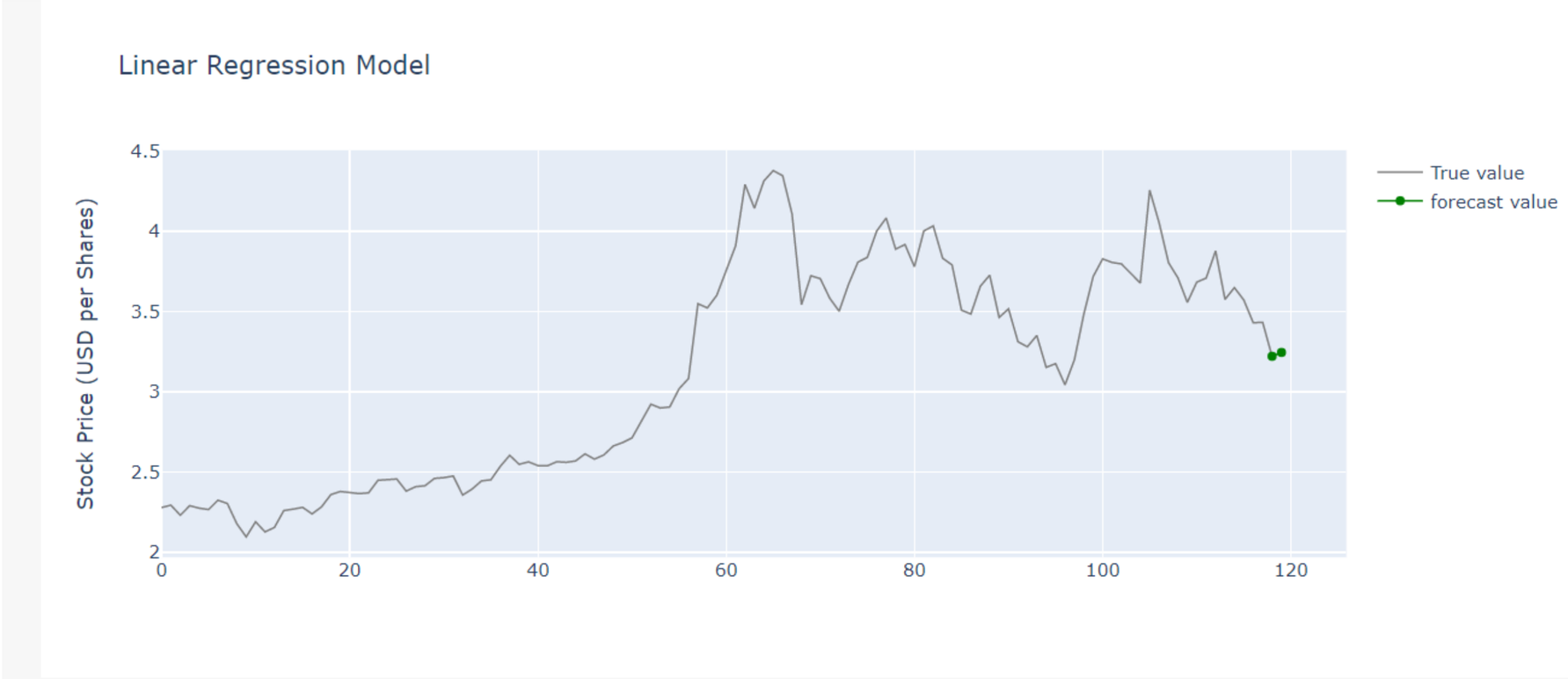
Forecast page

- Next Move of the stock price forecasted (UP/DOWN)

Enter the Ticker

TSLA

RMSE	R2 score	Next Move	Return
0.1636	0.9431	UP	0.0075



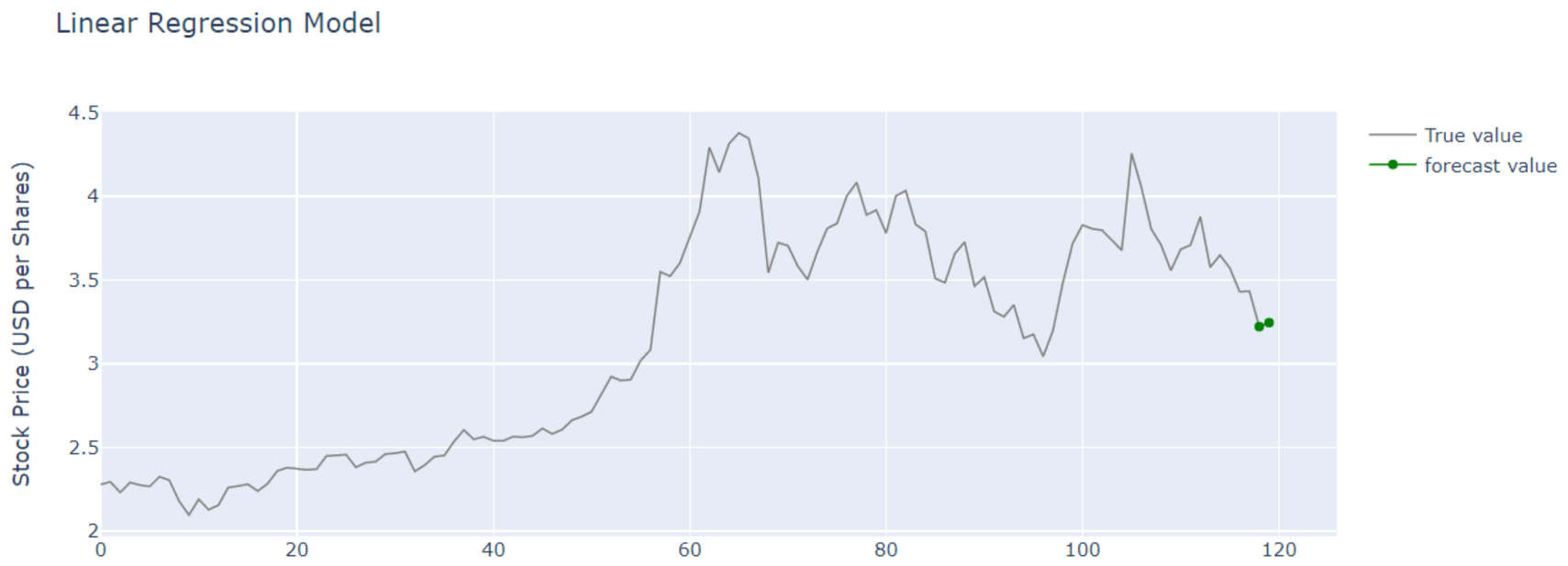
Forecast page

- Calculated by "formula"

Enter the Ticker

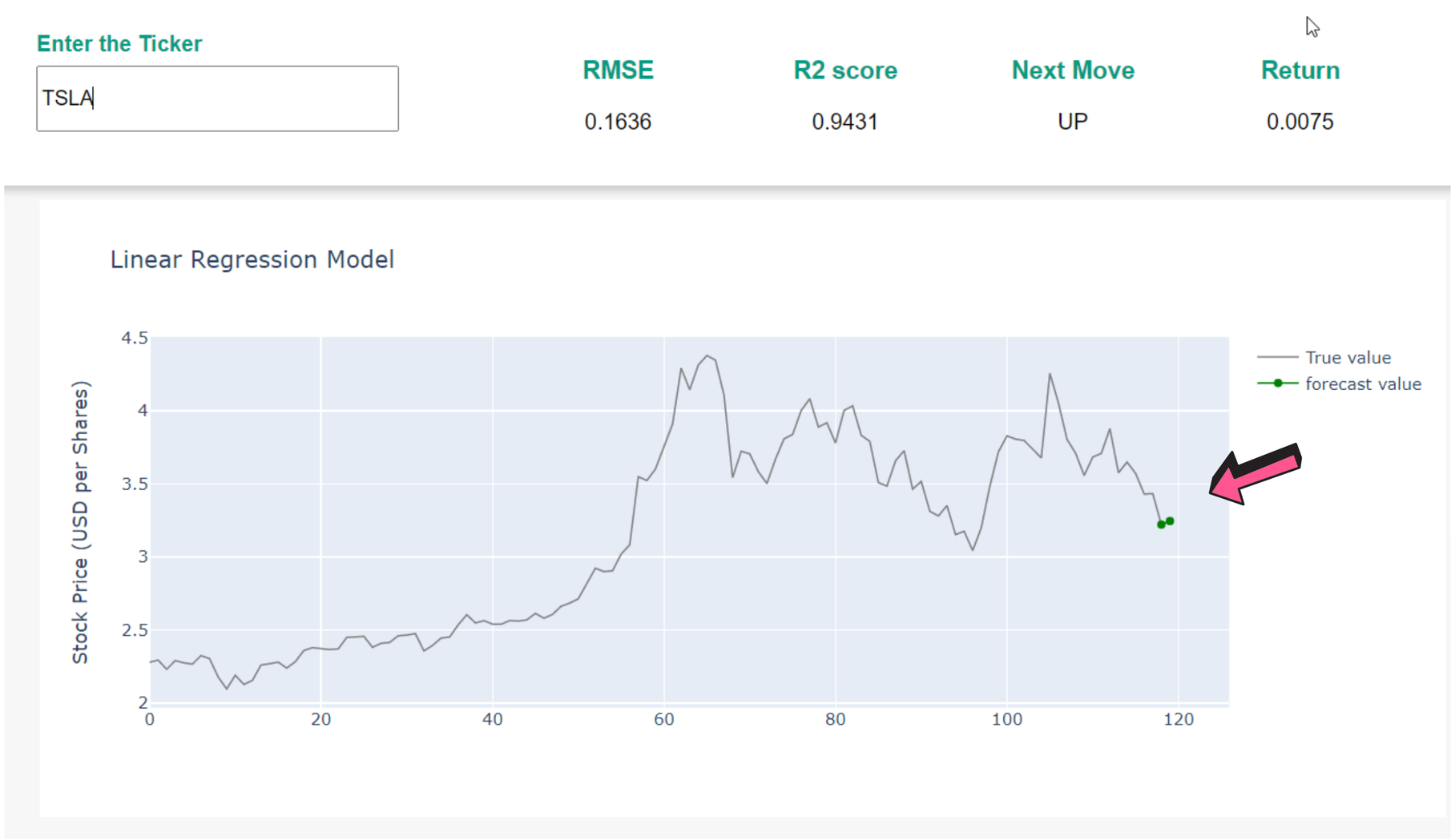
TSLA

RMSE	R2 score	Next Move	Return
0.1636	0.9431	UP	0.0075



Forecast page

- Next Move of the stock price forecasted plotted on the graph.
- If it is moving UP, the line plotted is in green color, and red in the opposite.



References

1. https://machinelearningmastery.com/linear-regression-for-machine-learning/
2. https://www.mathworks.com/discovery/neural-network.html#:~:text=A%20neural%20network%20(also%20called,data%2C%20and%20forecast%20future%20events.
3. https://machinelearningmastery.com/gentle-introduction-long-short-term-memory-networks-experts/
4. https://medium.com/wwblog/evaluating-regression-models-using-rmse-and-r%C2%B2-42f77400efee
5. https://towardsdatascience.com/preprocessing-time-series-data-for-supervised-learning-2e27493f44ae
6. https://support.minitab.com/en-us/minitab-express/1/help-and-how-to/modeling-statistics/regression/supporting-topics/basics/a-comparison-of-the-pearson-and-spearmancorrelation-methods/
7. https://research.google.com/colaboratory/faq.html
8. https://medium.com/plotly/introducing-dash-5ecf7191b503
9. https://www.heroku.com/what
10. https://www.investopedia.com/terms/d/daytrader.asp#:~:text=A%20day%20trader%20is%20a,which%20can%20also%20amplify%20losses.

Thank You !!!