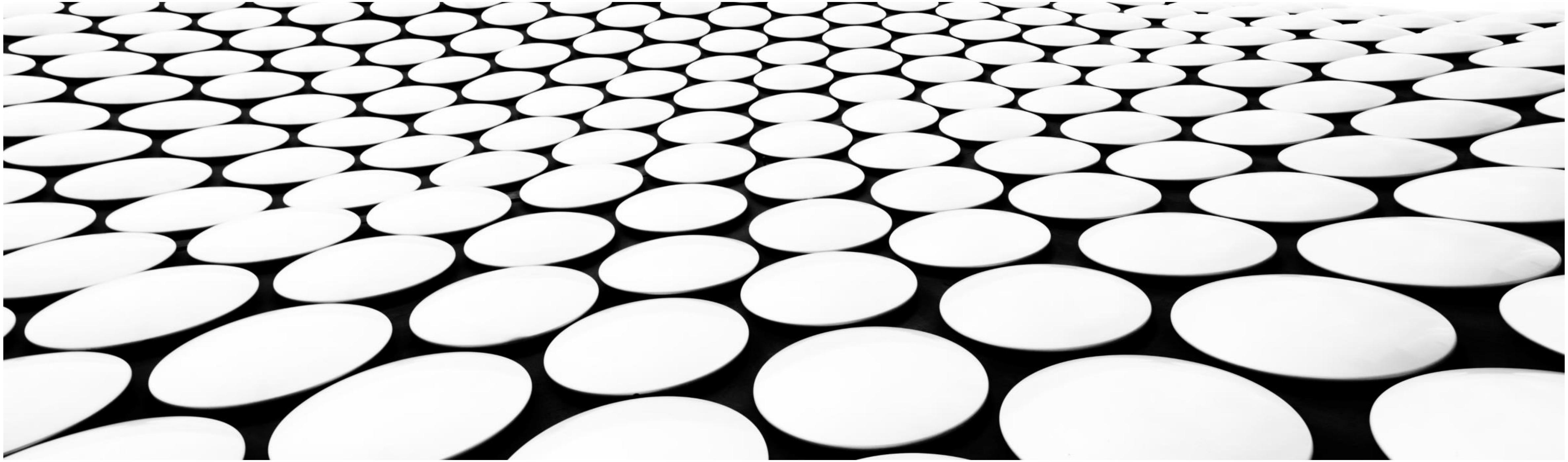
GROUP C - PRESENTATION

MODELS AND MACHINE LEARNING



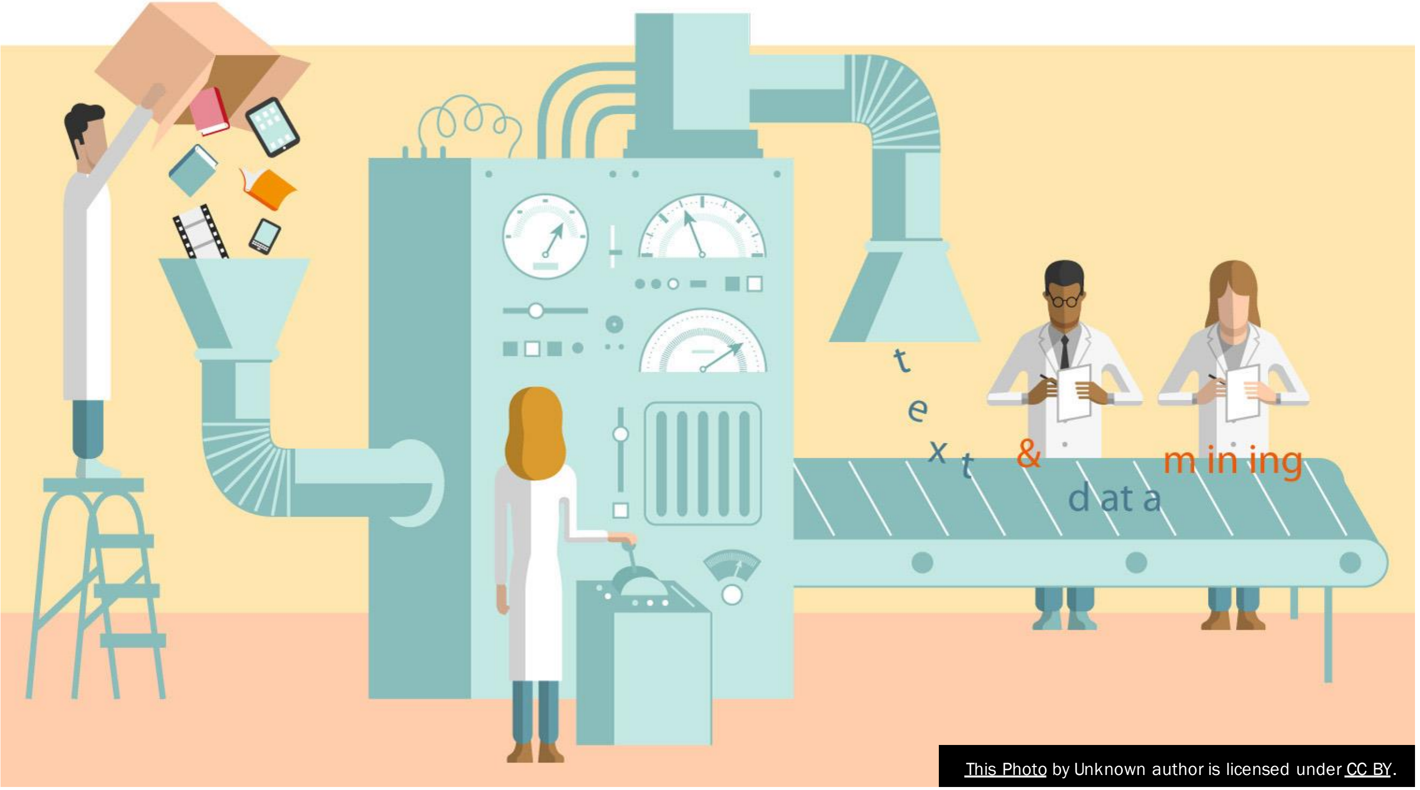
INTRODUCTION

The purpose of this presentation is to highlight the ability of machine learning to create 'buy' or 'sell' recommendations.

Our team uses multiple models and different types of data to produce varied results. Using evaluation models' comparative analysis is used with back testing to provide potential trading strategies.

DATA PREPARATION YFINANCE

- Yfinance is a python package that enables us to fetch historical market data from the Yahoo Finance API.
- Data is saved to a csv files to be used in the models.
- The Data is usually quite clean



TESTING DATA

Define Y (Target)

In this example we use a trading signal that throws a long when actual returns are positive (when the price is going up) and a short when actual returns are negative

Here we're using 1 and 0 so we can later diff the file to create actual trade signals. As is these signals just indicate whether the price is going up or down

```python
# Create a new column in the `trading_df` called "signal" setting its value to zero.
trading_df["signal"] = 0.0
```

```python
# Create the signal to buy
trading_df.loc[(trading_df["actual_returns"] >= 0), "signal"] = 1
```

```python
# Create the signal to sell
trading_df.loc[(trading_df["actual_returns"] < 0), "signal"] = 0
```

```python
# Copy the new "signal" column to a new Series called `y`.
y = trading_df["signal"].copy()
```

- Stock price movement indicators – created buy and sell signals utilizing the percentage change of the close price.
- How we created the trading signals – Diff function to identify changes in price movement, triggering trade signals
- Time Period (5 Ys for this iteration)
- Data splitting (1Y train, 4Y test) - 1:4 ratio split kept throughout versions of the model, initially train/test split assigned incorrectly to ratio.

FEATURES

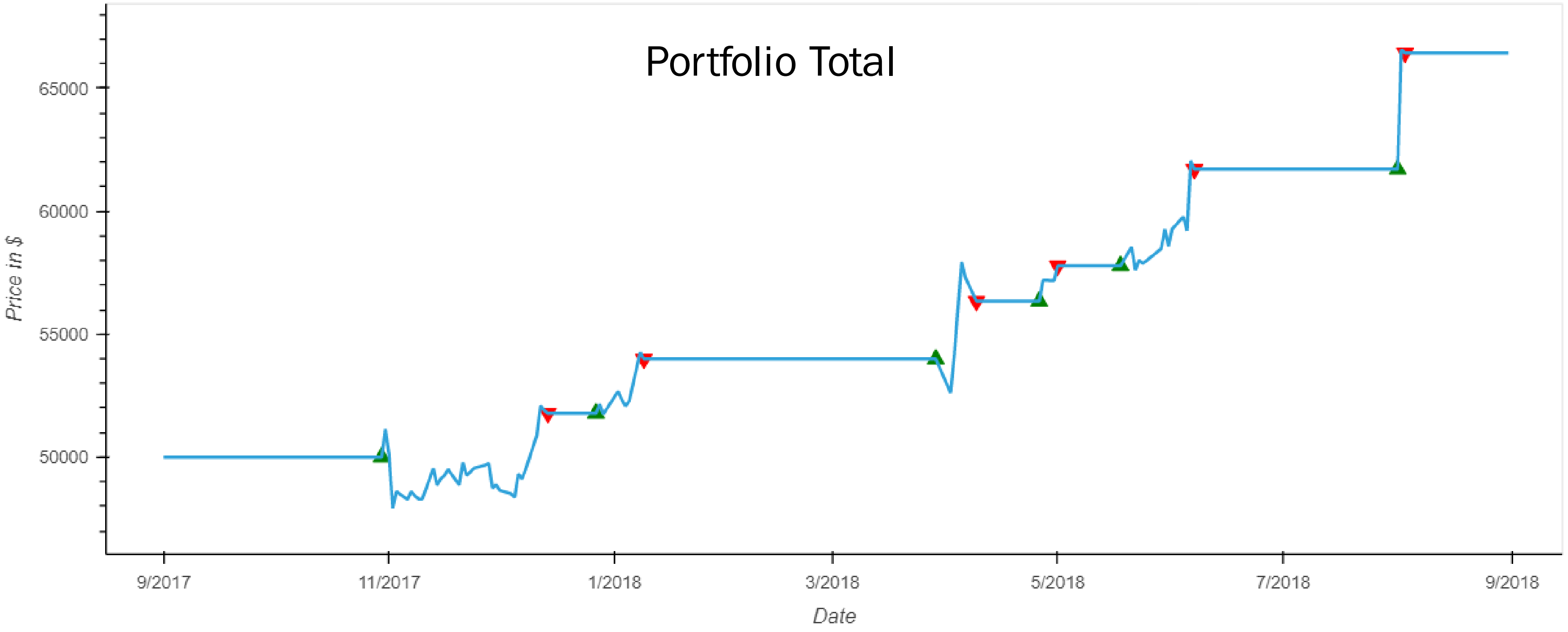
- **SMA (Simple Moving Averages)** - we use the SMA to reduce the daily price noise, thus we can observe more closely the longer-term price behaviour of the asset.
  - For our model we chose to use the "Fast SMA" with a short window of 4 days and a "Slow SMA" using a long window of 100 days

- **Bollinger bands** – A type of technical indicator that assists traders to understand the volatility of the asset and whether the price is high or low on a relative basis

FEATURES - FINTA PACKAGE

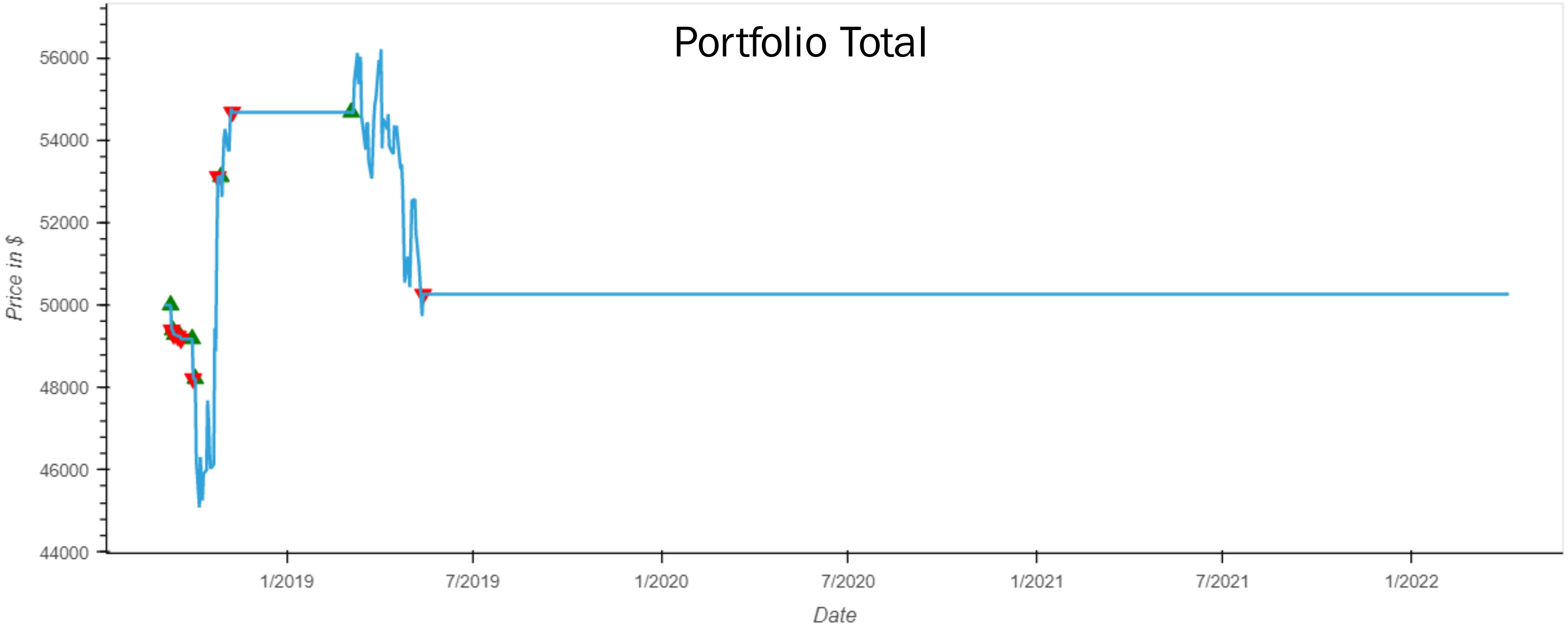
- We applied the information on from dataframe into the FINTA package to produce the technical indicators for assessment.
- All our features are shifted one day to predict future prices

TESTS

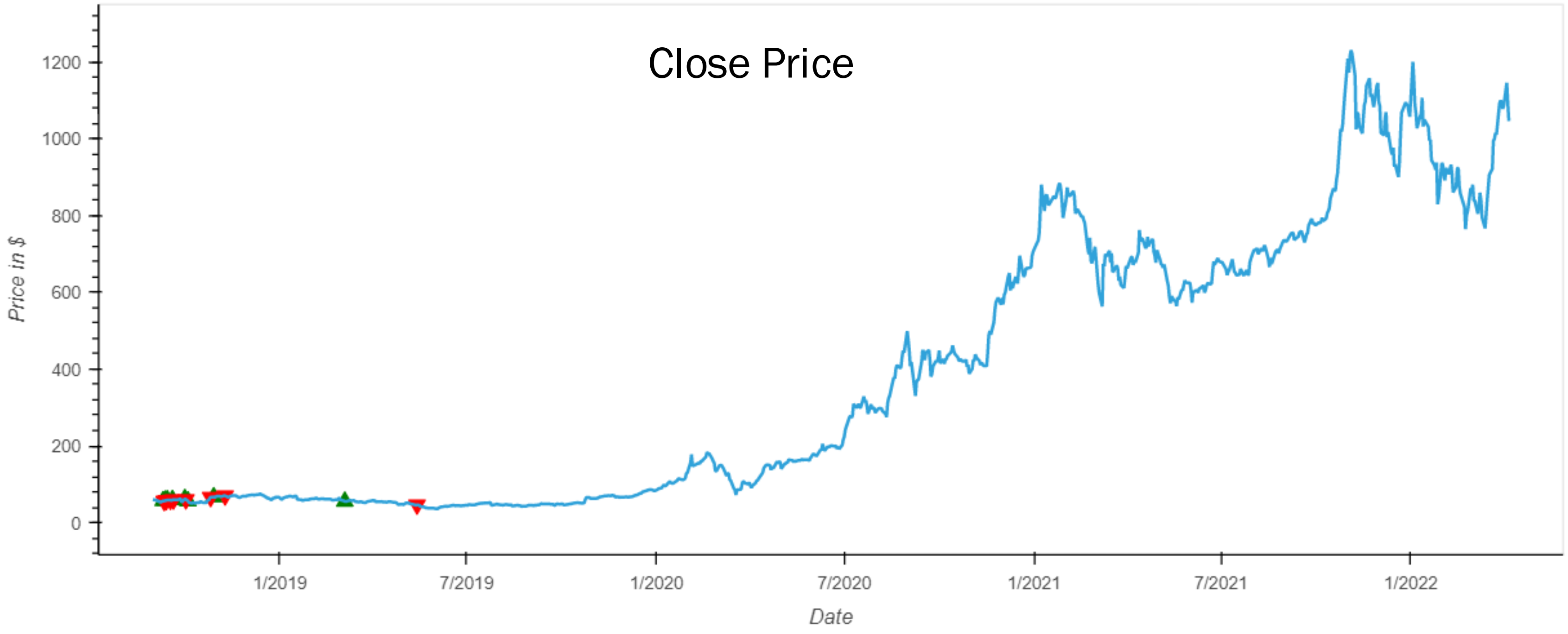


TRAINING
X = SMA FAST + SMA SLOW
MODEL = SVC
DATA = TSLA

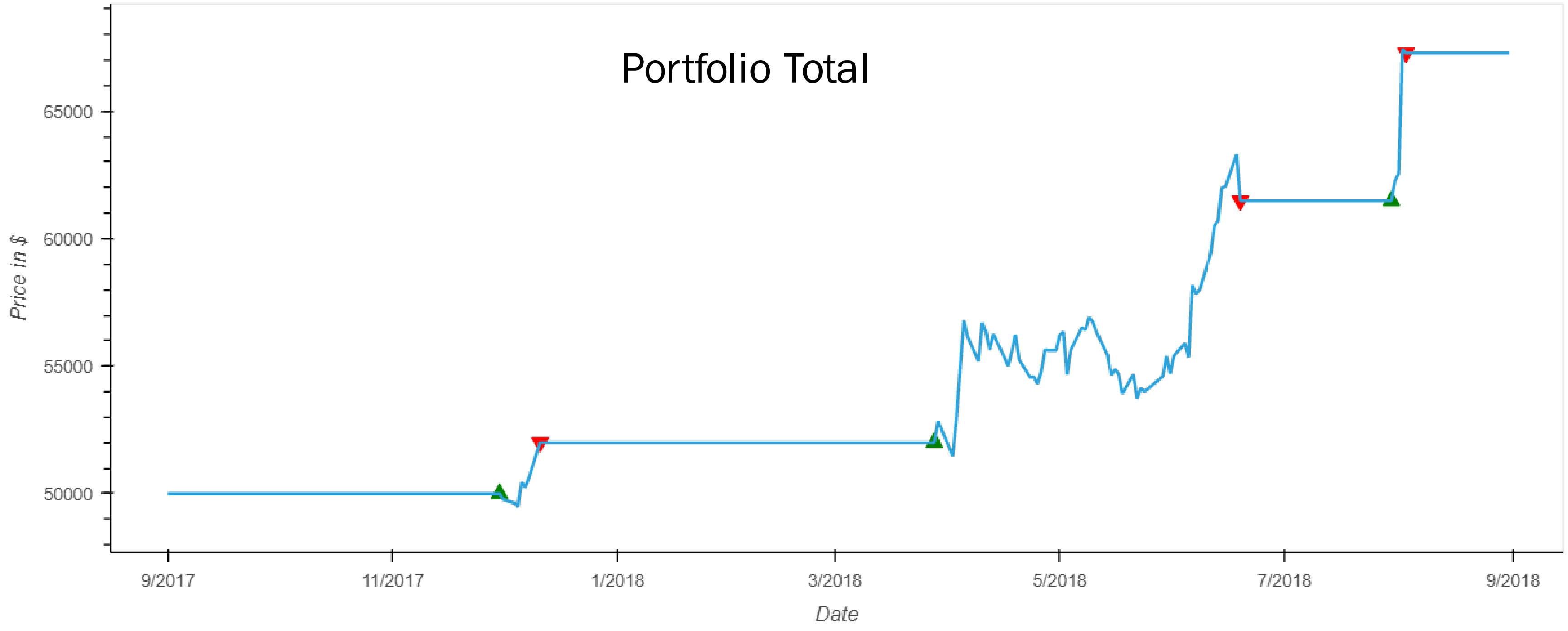
	Backtest
Annualized Return	0.6413517416162857
Cumulative Returns	7.163000030517555
Annual Volatility	0.47883554485147883
Sharpe Ratio	1.339398773779868
Sortino Ratio	1.7327893599882478



TESTING
X = SMA FAST + SMA SLOW
MODEL = SVC
DATA = TSLA



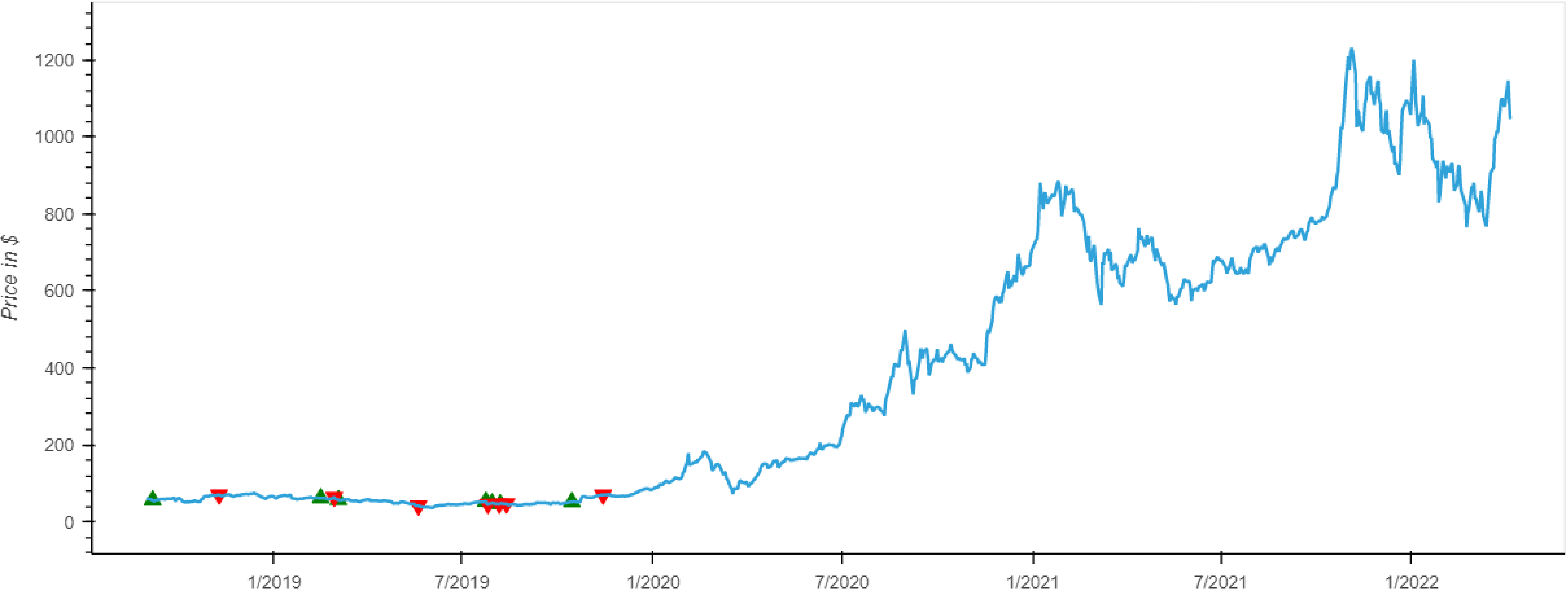
	Backtest
Annualized Return	-9.003507824857376
Cumulative Returns	-1.6196008300781248
Annual Volatility	5.551588843696507
Sharpe Ratio	-1.6217893792837907
Sortino Ratio	-inf



TRAINING
X = BOLLINGER BANDS + CLOSE
MODEL = SVC
DATA = TSLA

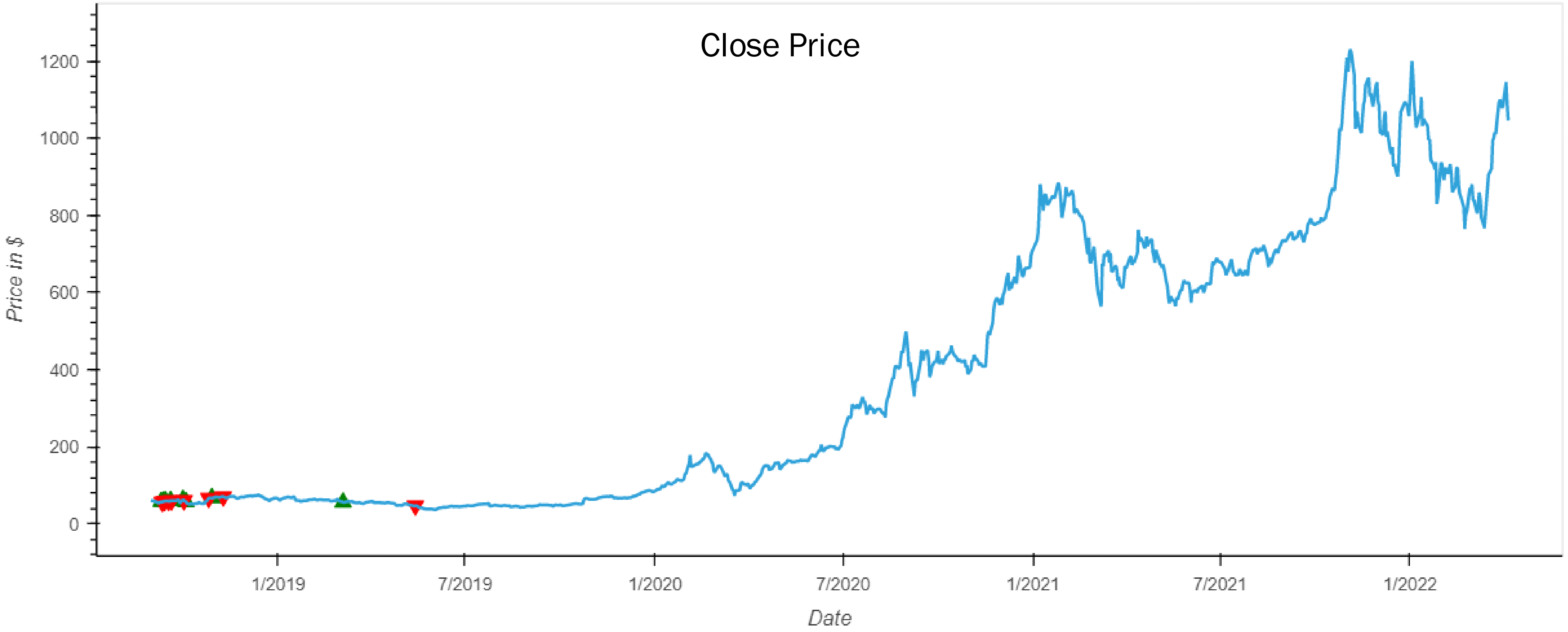
	Backtest
Annualized Return	0.30972977597730883
Cumulative Returns	0.34566001892089737
Annual Volatility	0.15331850722849982
Sharpe Ratio	2.0201721343118733
Sortino Ratio	2.280091850119308

TESTING
X = BOLLINGER BANDS + CLOSE
MODEL = SVC
DATA = TSLA



TRAINING
X = SMA + BOLLINGER BANDS + CLOSE
MODEL = SVC
DATA = TSLA

TESTING
X = SMA + BOLLINGER BANDS + CLOSE
MODEL = SVC
DATA = TSLA



WHAT WENT WRONG?

Bad Features?

SMA
BOLLINGER BAND
CLOSE PRICE

Not Enough Data?

2017-2022

CHANGES TO THE MODEL: DATA CHANGES

- Switched from TSLA to DIS.
  - DIS is a much more active stock over its trading life, unlike TSLA which experiences high growth over a short period.
- We made a mistake! Training split was wrong (it should be 80 Training and 20 Testing)
  - Generally you need more training data to help your ML model learn

CHANGES TO THE MODEL: NEW FEATURES

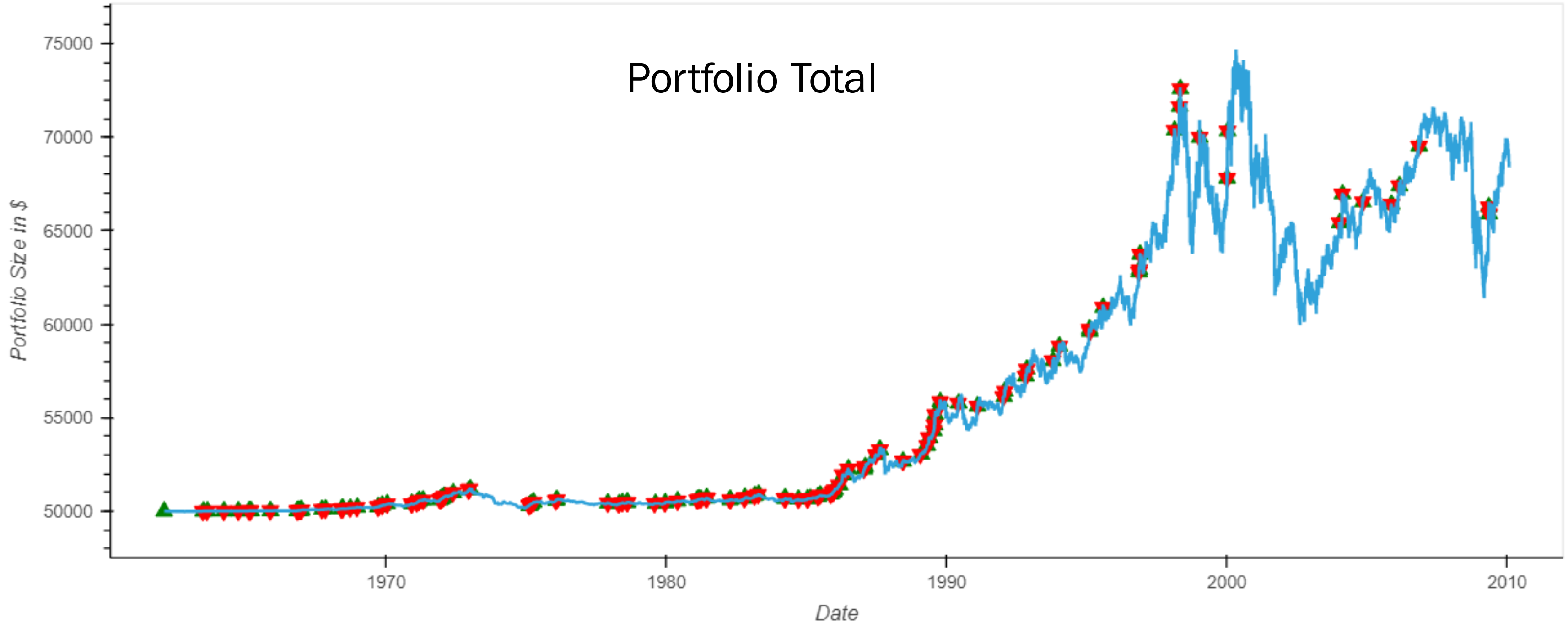
- New Features:
- RSI - Historical strength or weakness of a stock or market based on the closing prices of a recent trading period. It is primarily used to attempt to identify overbought or oversold conditions in the trading of an asset.
- Average Directional Movement Index (ADX) - The ADX indicator is an average of expanding price range values. ADX stands for Average Directional Movement Index and **can be used to help measure the overall strength of a trend.**
- Average True Range (ATR) - ATR provides an indication of the degree of price volatility. Strong moves, in either direction, are often accompanied by large ranges, or large True Ranges. (average chance in price over set period of time, showing volitility)

PROBLEMS WITH PREVIOUS FEATURES

- Previous Metrics are derivatives of the close price
- These aren't typically good metrics for predicting the market, they need to be interpreted first
  - Creating an indicator using fast and slow sma SMA crossover
  - You dont use close price to predict close price

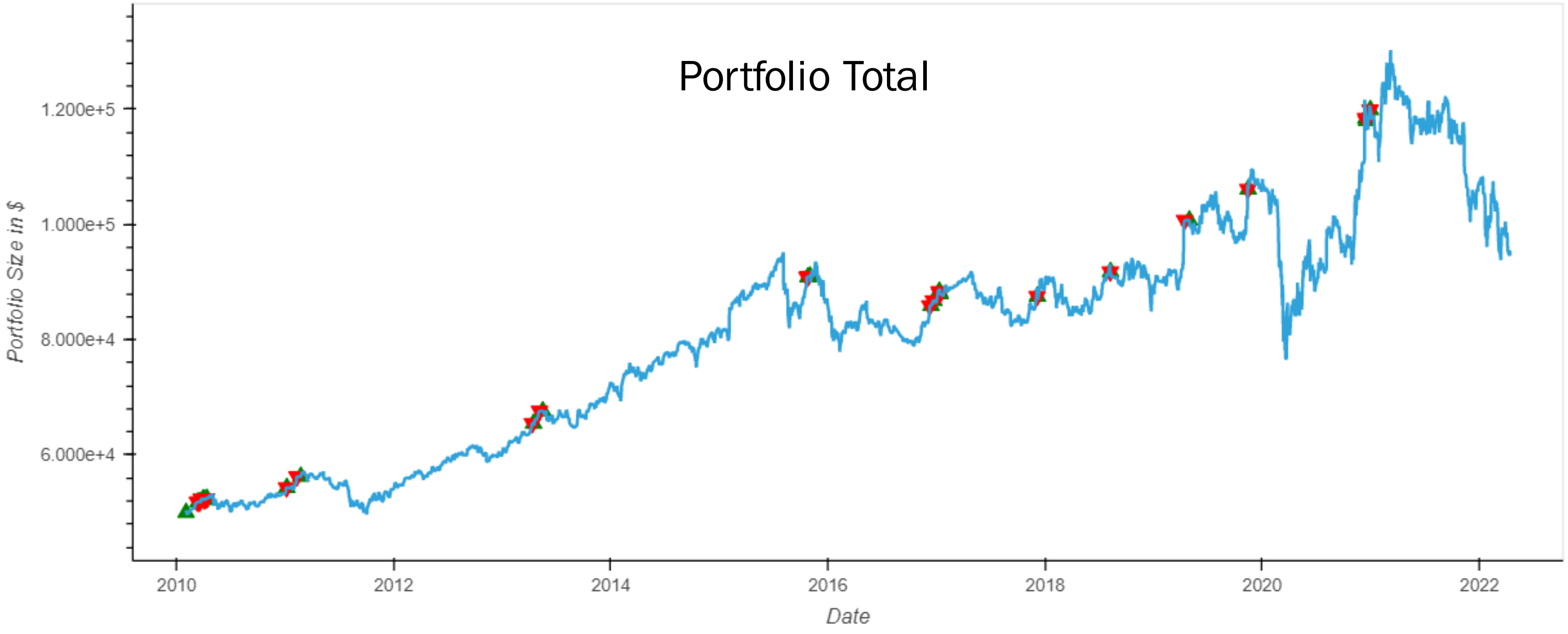
TESTS

(AGAIN)



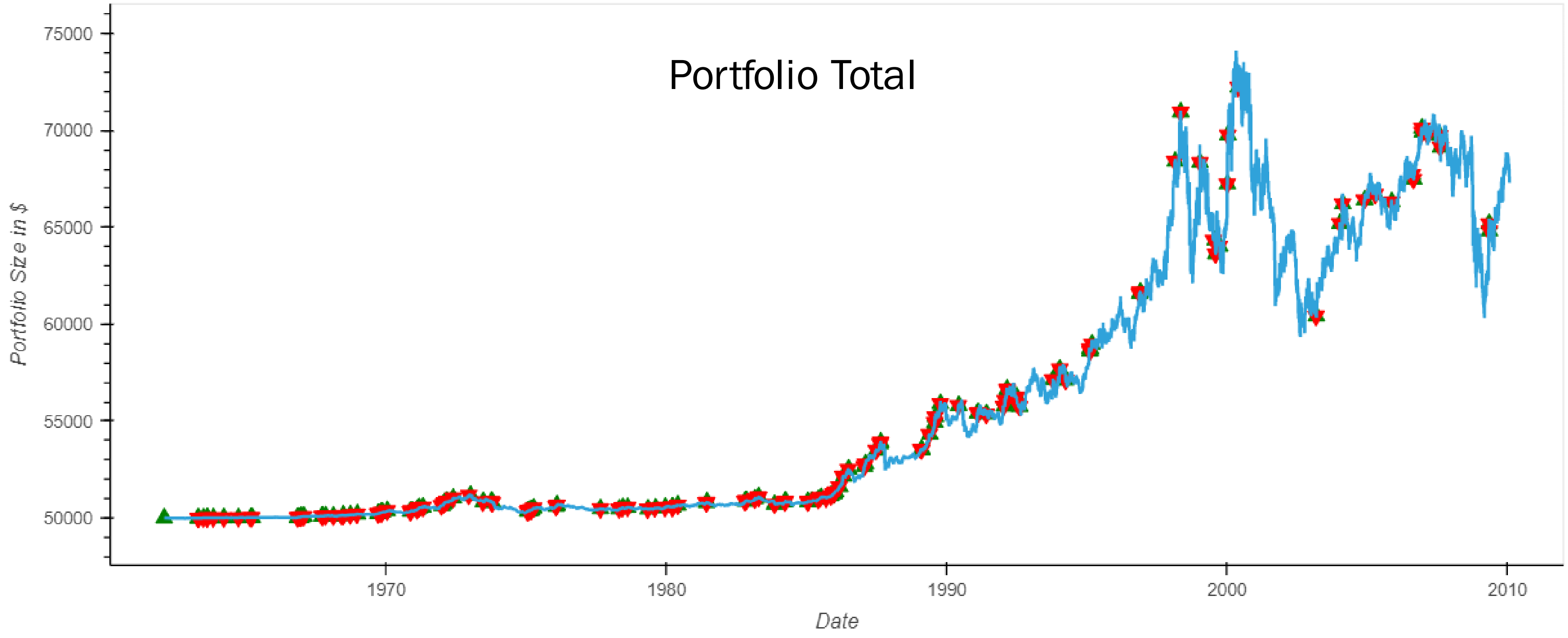
TRAINING
X = RSI
MODEL = SVC
DATA = DIS

	Backtest
Annualized Return	0.007337
Cumulative Returns	0.370367
Annual Volatility	0.039087
Sharpe Ratio	0.187699
Sortino Ratio	0.259984

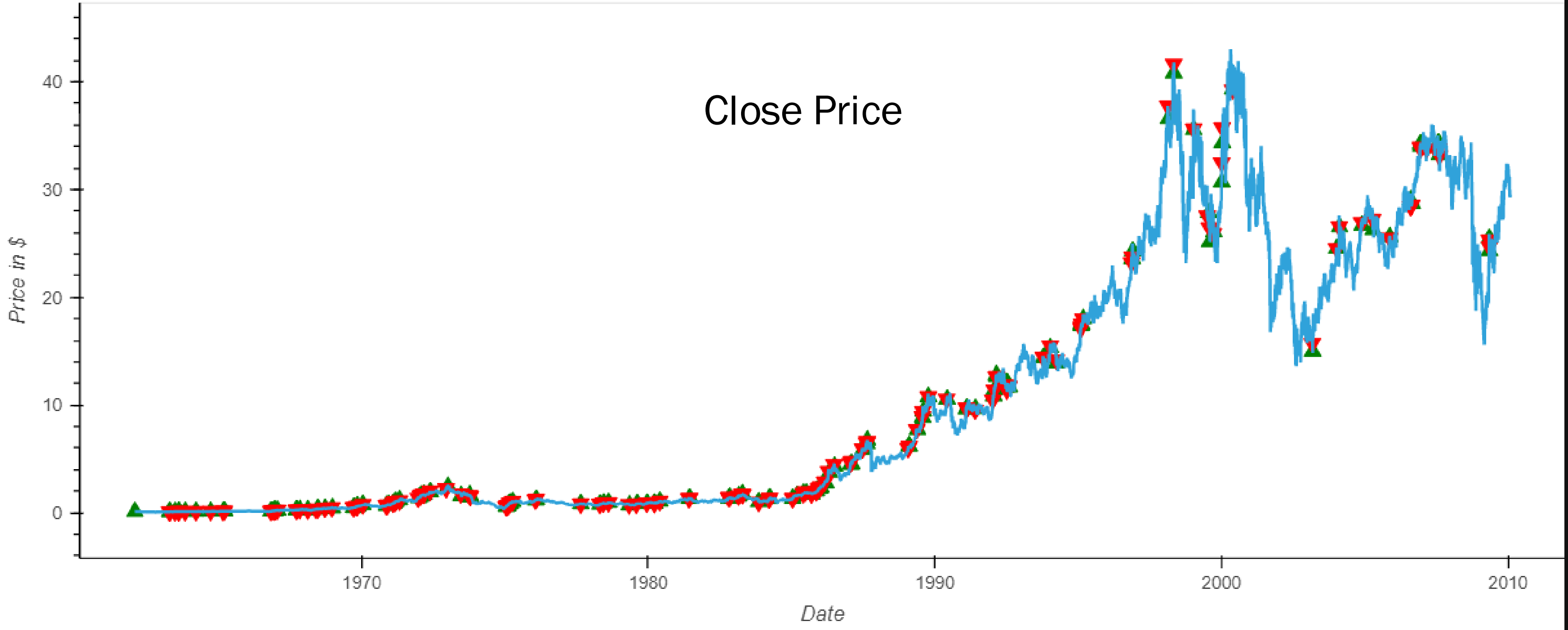


TESTING
X = RSI
MODEL = SVC
DATA = DIS

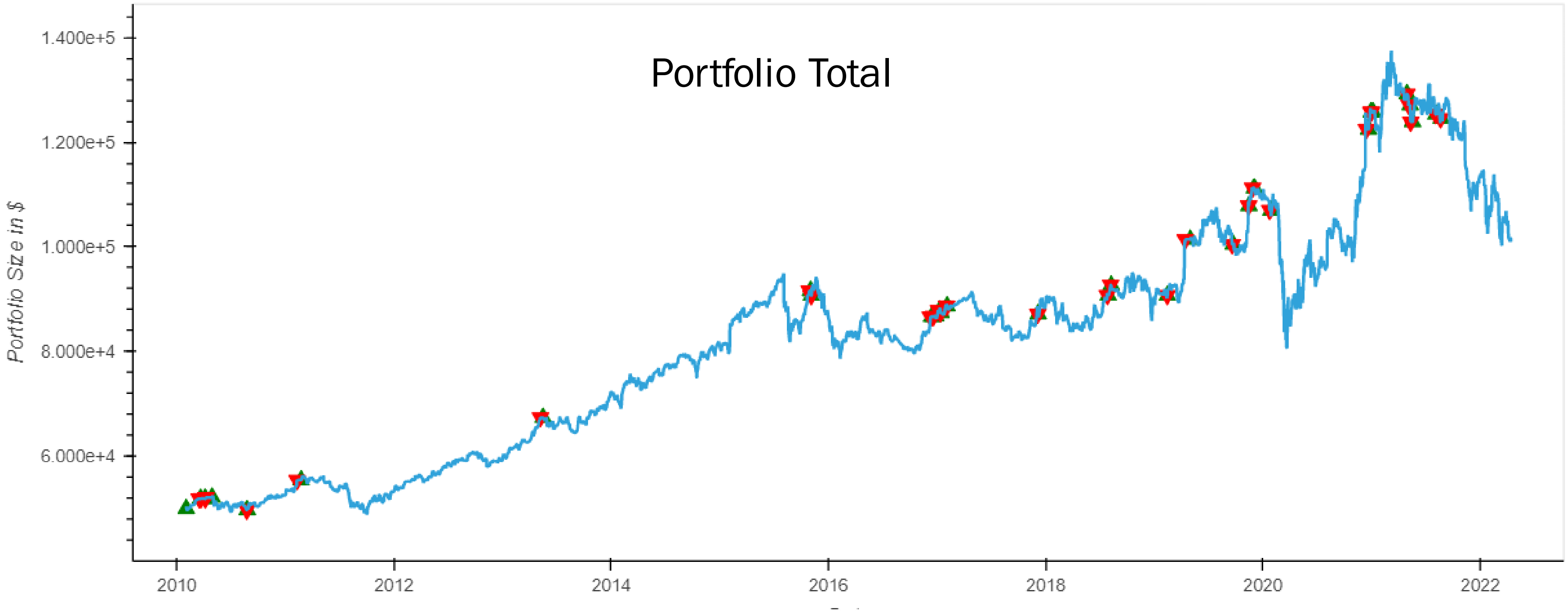
	Backtest
Annualized Return	0.062632
Cumulative Returns	0.891
Annual Volatility	0.144213
Sharpe Ratio	0.434301
Sortino Ratio	0.591691



TRAINING
X = RSI + ADX
MODEL = SVC
DATA = DIS

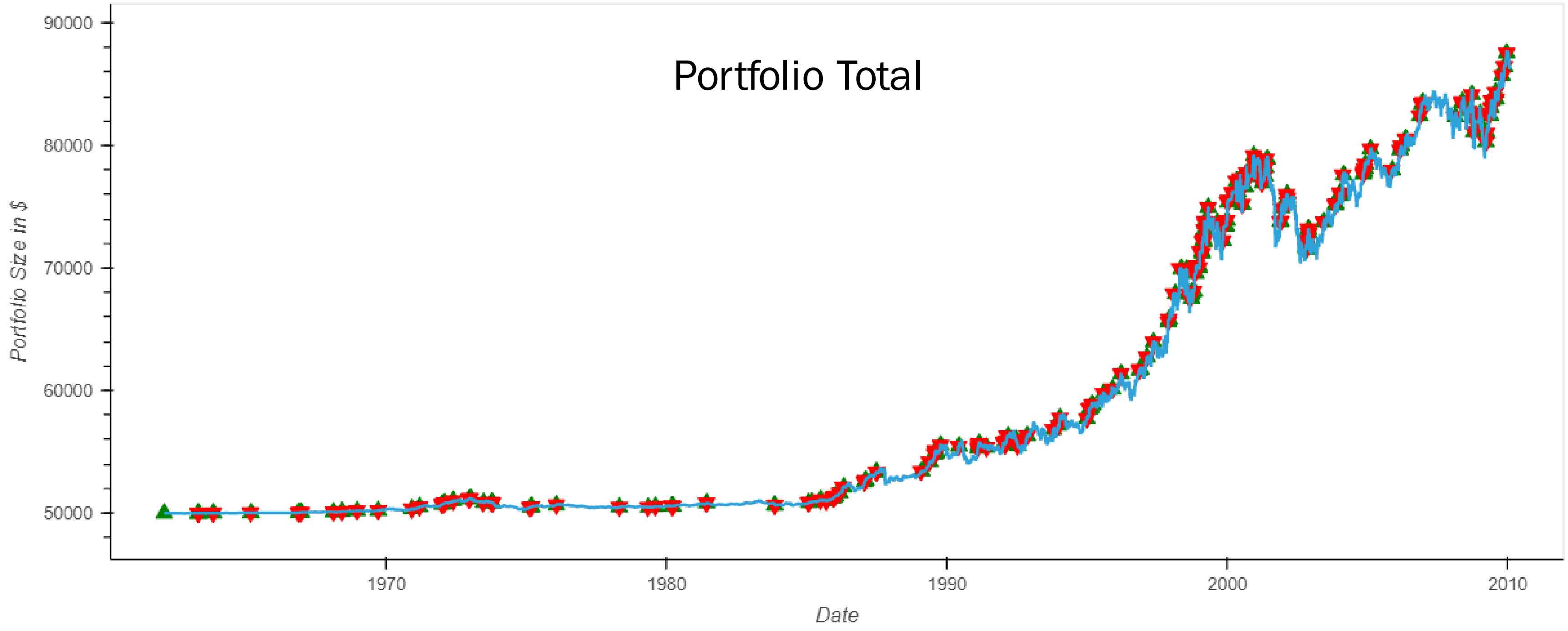


	Backtest
Annualized Return	0.007017
Cumulative Returns	0.348456
Annual Volatility	0.039516
Sharpe Ratio	0.177579
Sortino Ratio	0.246122



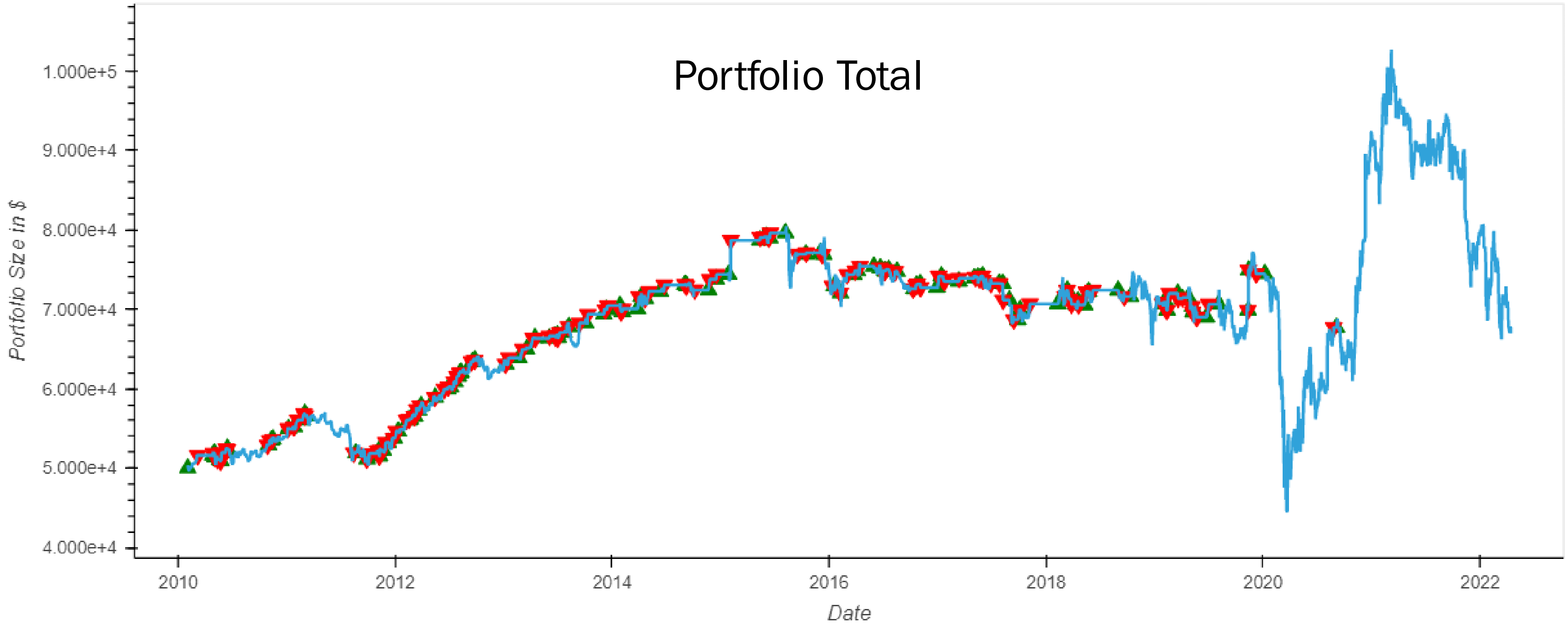
TESTING
X = RSI + ADX
MODEL = SVC
DATA = DIS

	Backtest
Annualized Return	0.067356
Cumulative Returns	1.0186
Annual Volatility	0.13973
Sharpe Ratio	0.482047
Sortino Ratio	0.655529



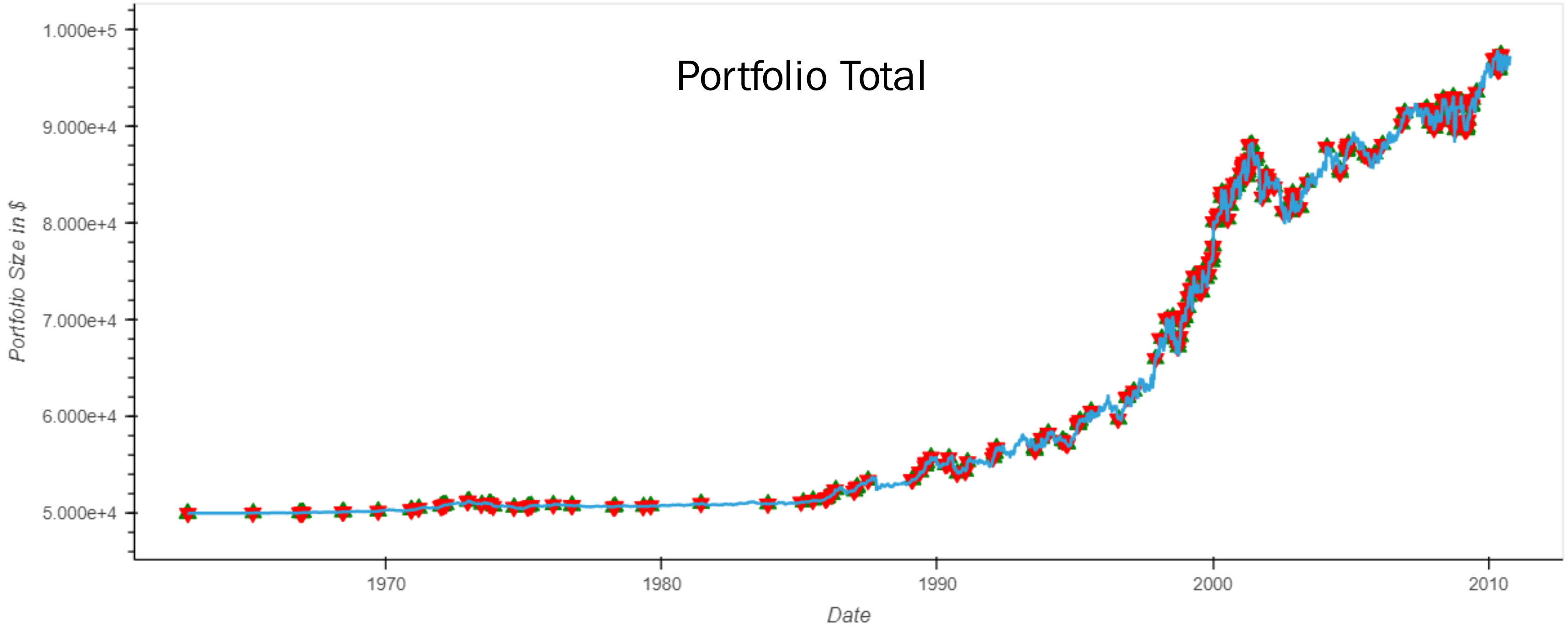
TRAINING
X = RSI + ADX + ATR
MODEL = SVC
DATA = DIS

	Backtest
Annualized Return	0.011823
Cumulative Returns	0.727229
Annual Volatility	0.029036
Sharpe Ratio	0.407168
Sortino Ratio	0.556874

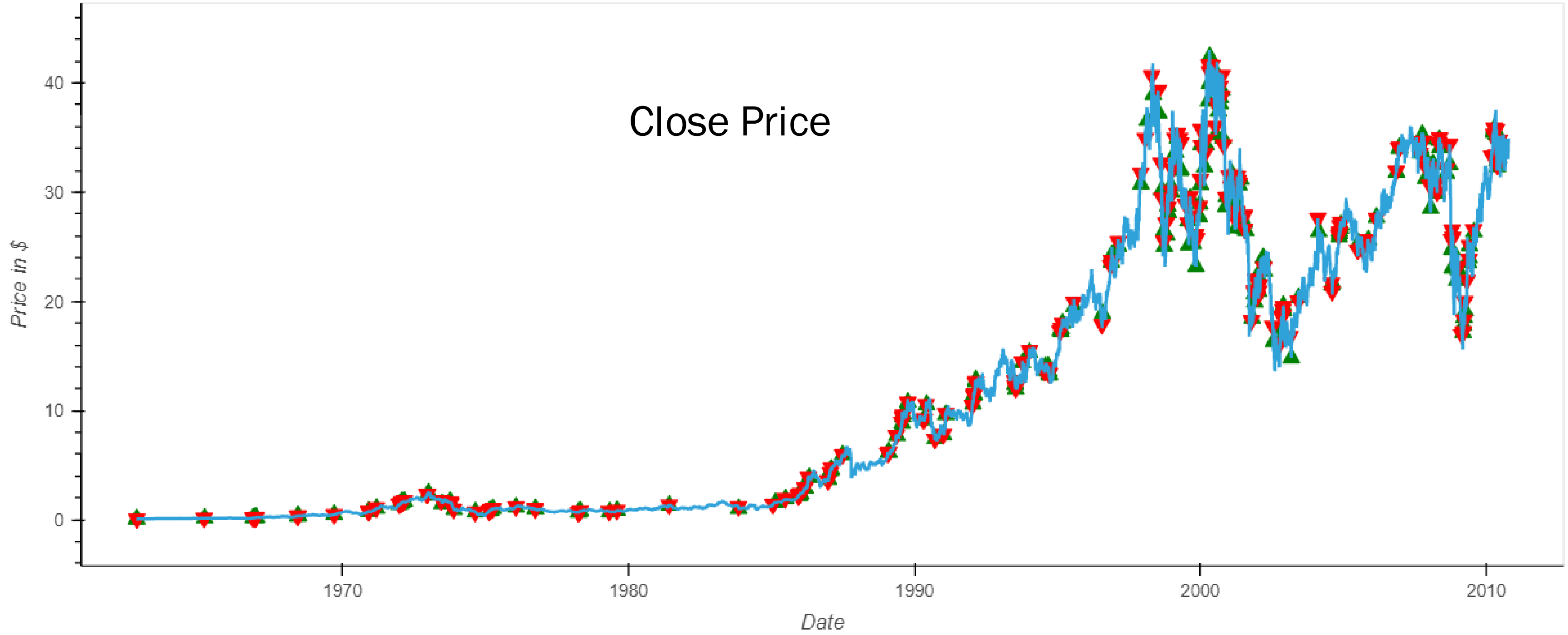


TESTING
X = RSI + ADX + ATR
MODEL = SVC
DATA = DIS

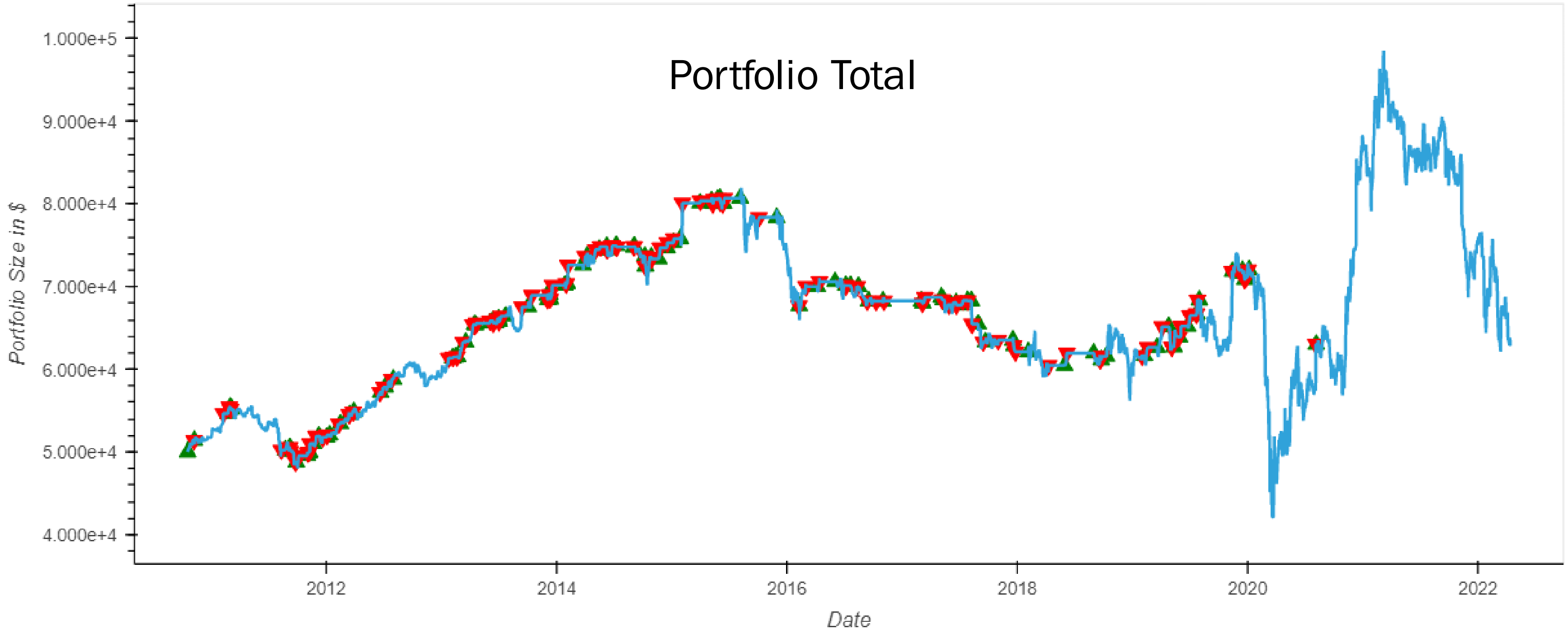
	Backtest
Annualized Return	0.039806
Cumulative Returns	0.3391
Annual Volatility	0.178612
Sharpe Ratio	0.222862
Sortino Ratio	0.30114



TRAINING
X = RSI + ADX + ATR + SMA SIG
MODEL = SVC
DATA = DIS

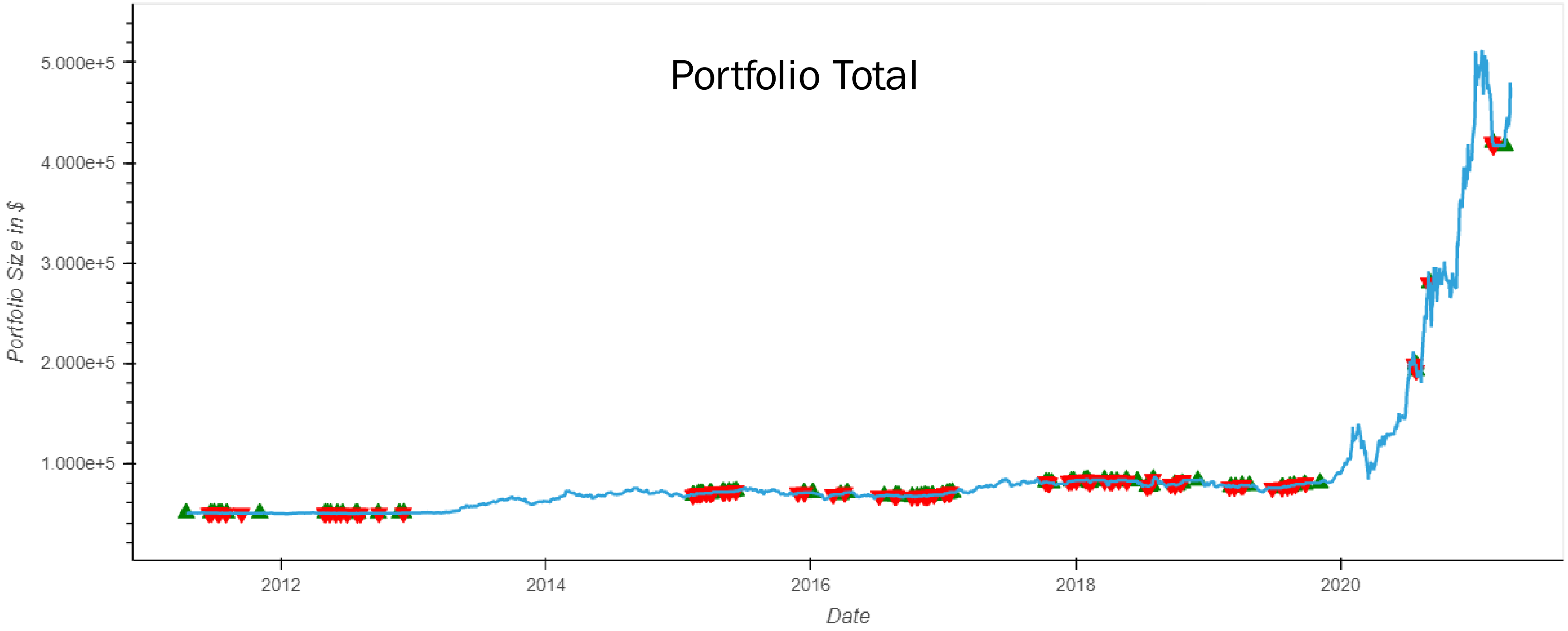


	Backtest
Annualized Return	0.014286
Cumulative Returns	0.945891
Annual Volatility	0.028228
Sharpe Ratio	0.506094
Sortino Ratio	0.676575



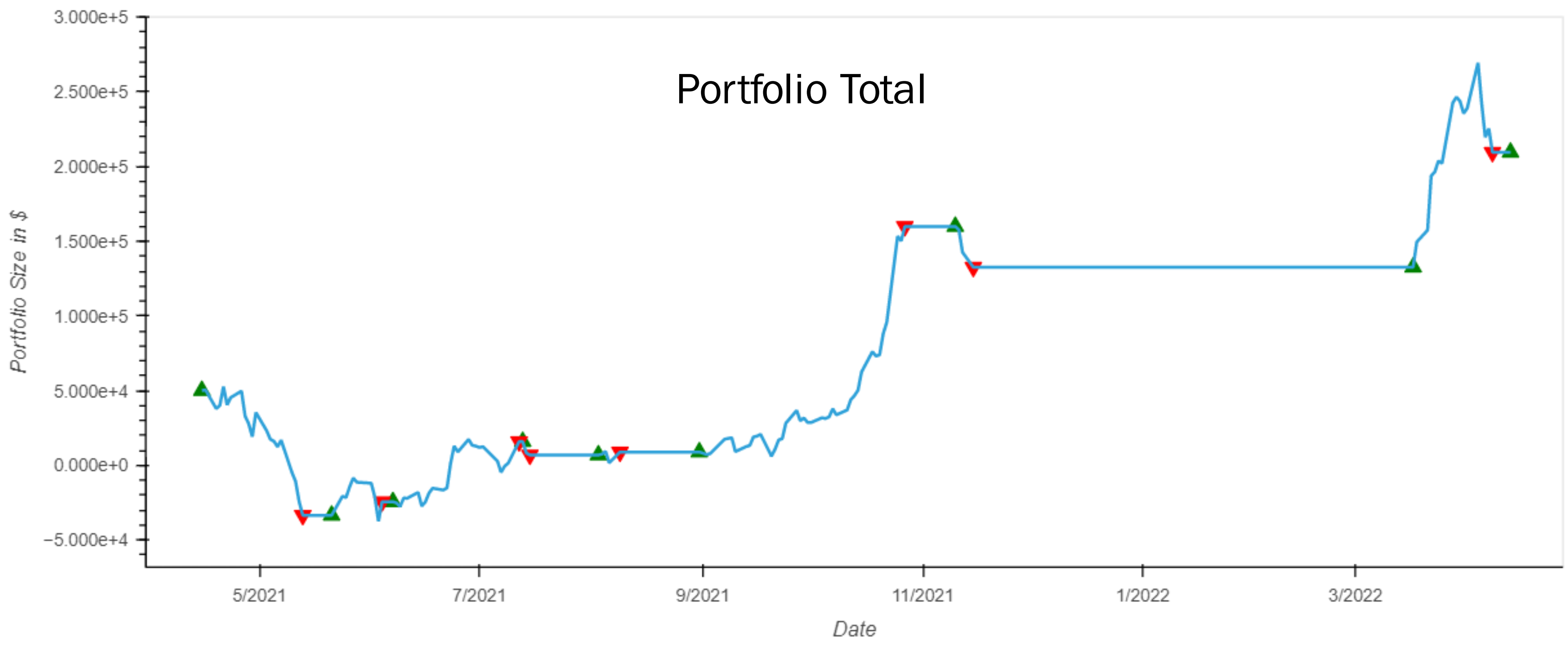
TESTING
X = RSI + ADX + ATR + SMA SIG
MODEL = SVC
DATA = DIS

	Backtest
Annualized Return	0.038987
Cumulative Returns	0.257
Annual Volatility	0.195855
Sharpe Ratio	0.19906
Sortino Ratio	0.269621

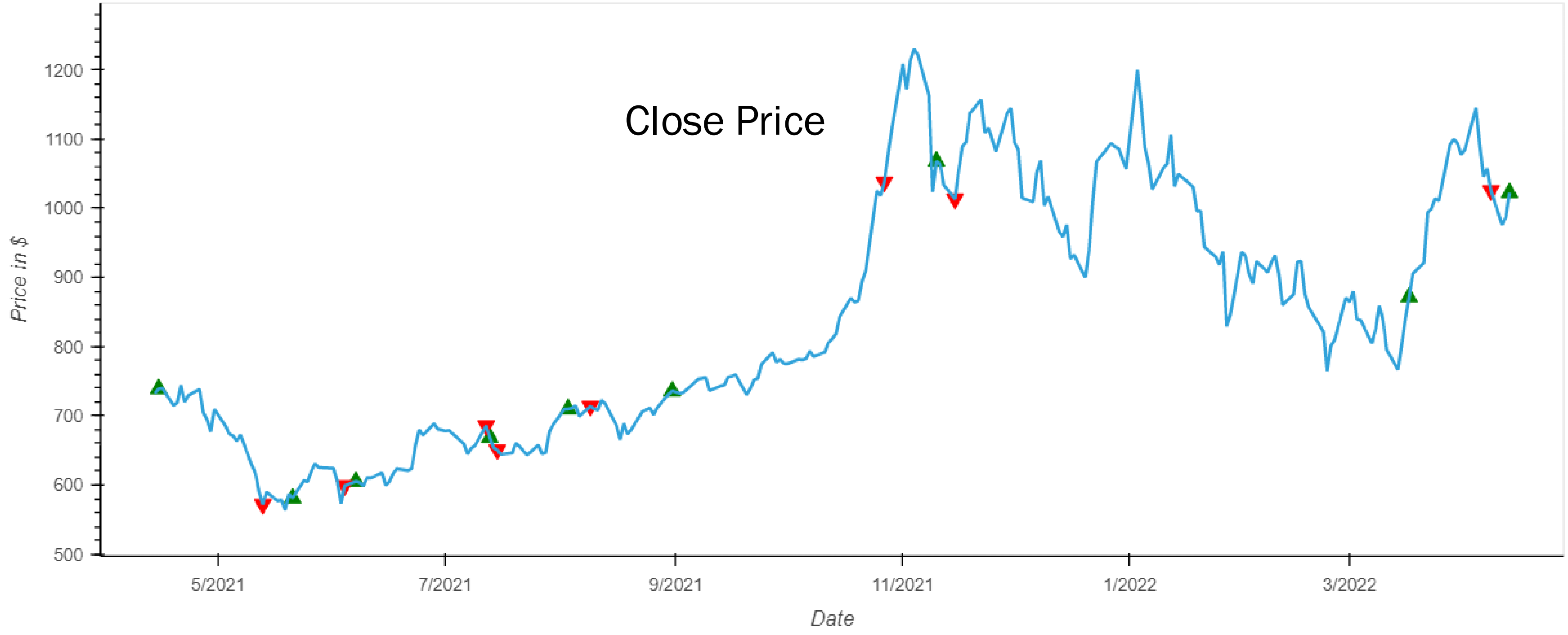


TRAINING
X = RSI + ADX + ATR + SMA SIG
MODEL = SVC
DATA = TSLA

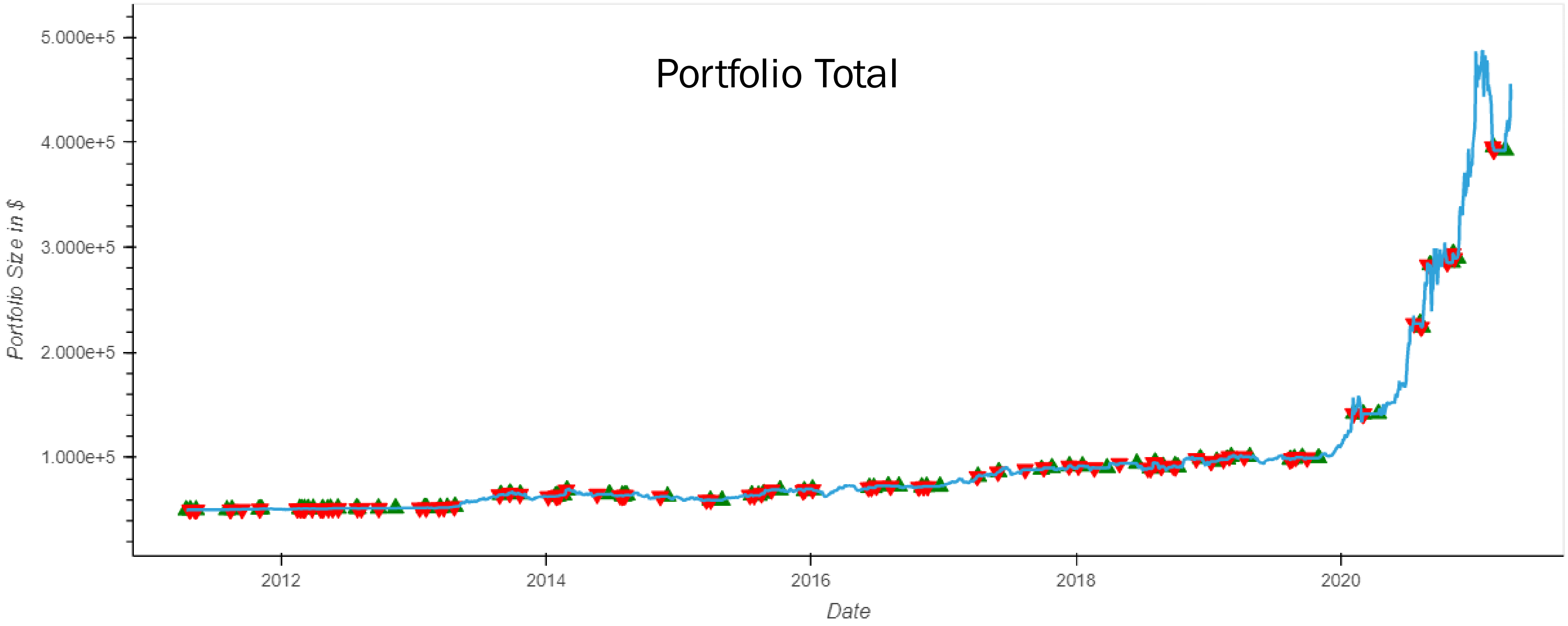
	Backtest
Annualized Return	0.249107
Cumulative Returns	8.30516
Annual Volatility	0.225819
Sharpe Ratio	1.103128
Sortino Ratio	1.420658



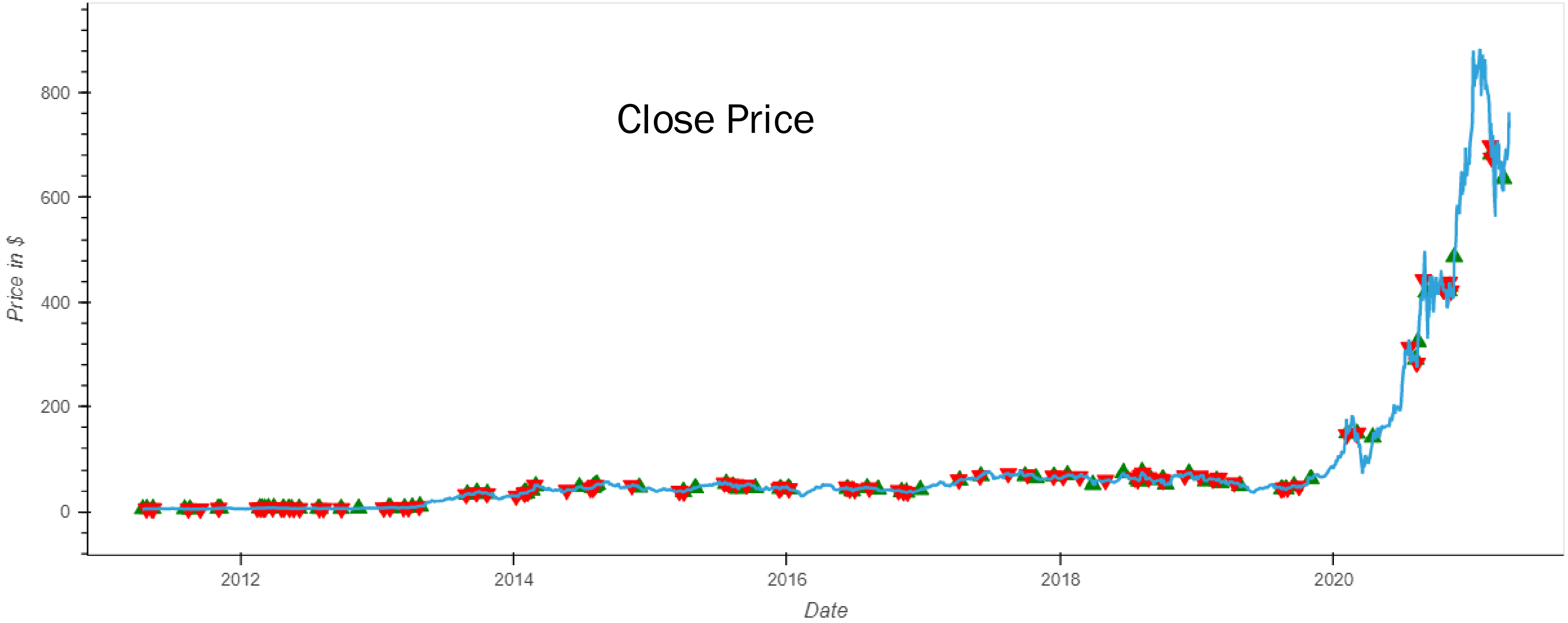
TESTING
X = RSI + ADX + ATR + SMA SIG
MODEL = SVC
DATA = TSLA



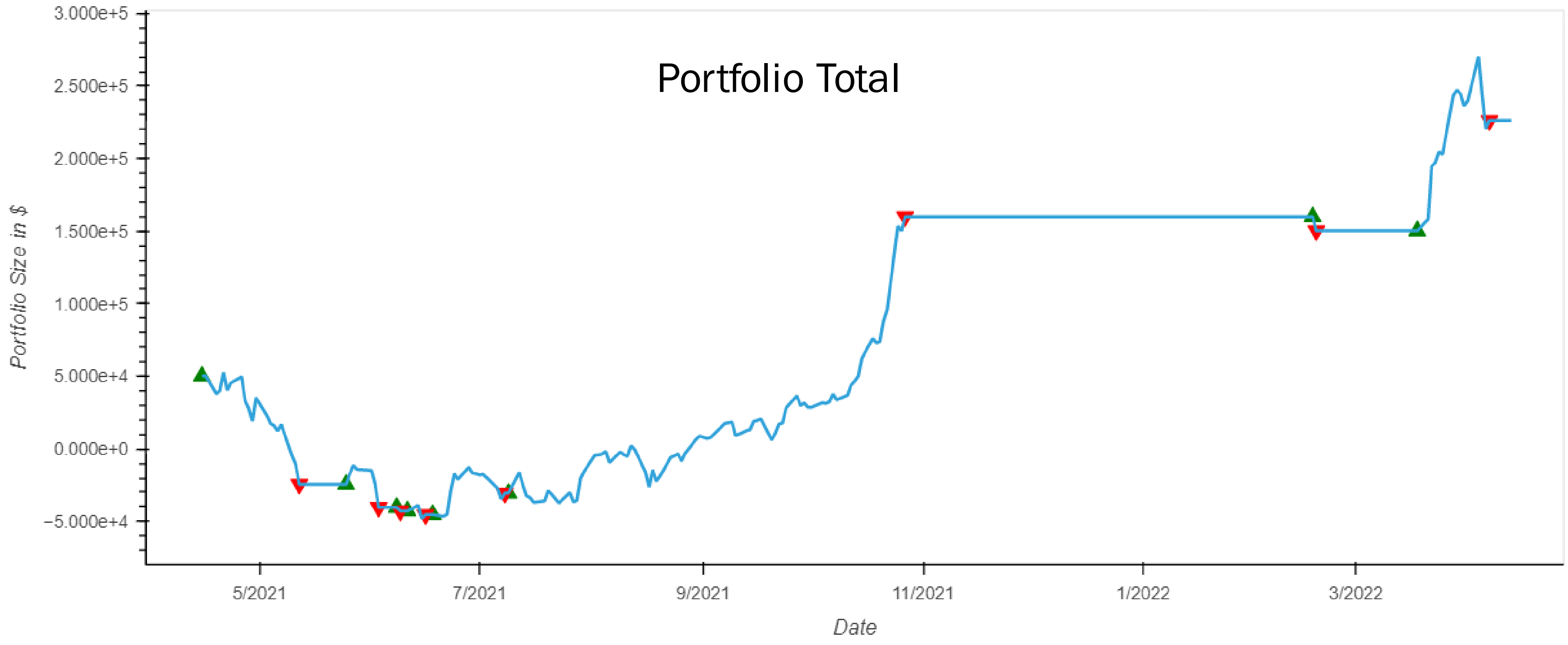
	Backtest
Annualized Return	21.805591
Cumulative Returns	3.191301
Annual Volatility	15.871312
Sharpe Ratio	1.3739
Sortino Ratio	1.450372



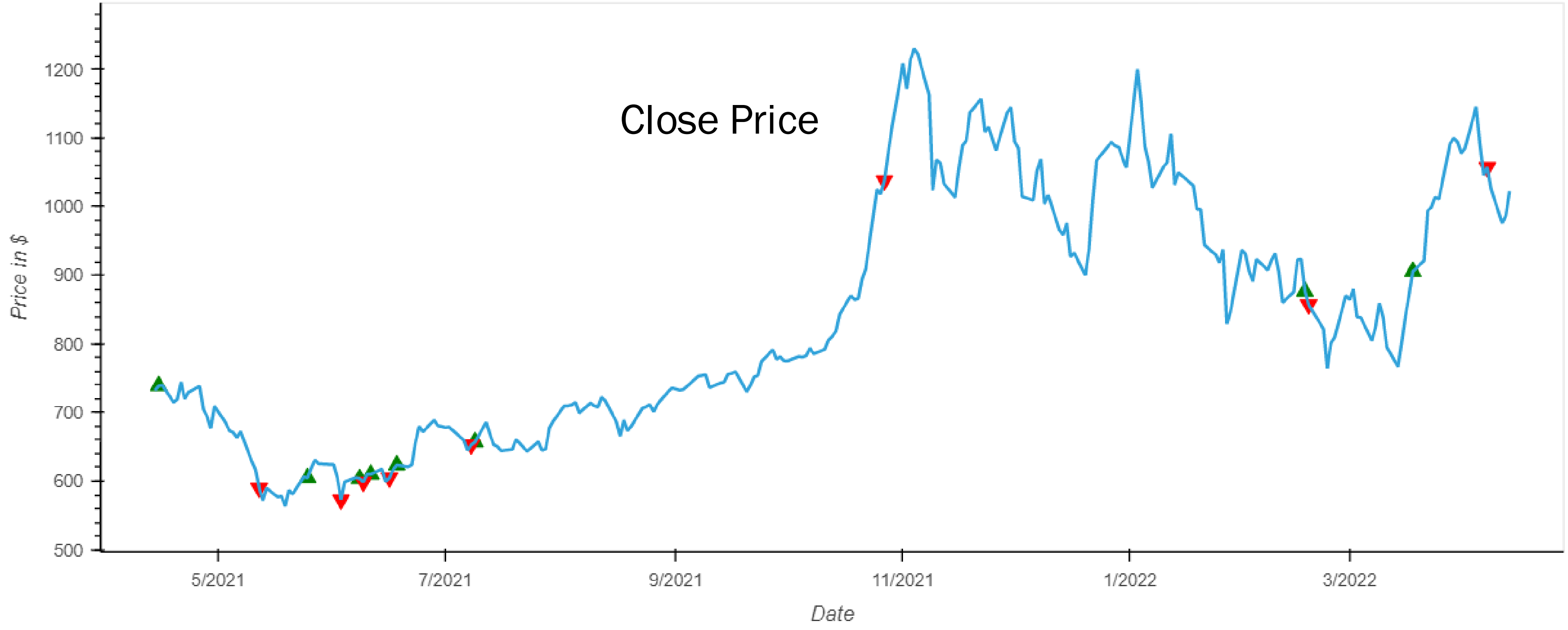
TRAINING
X = RSI + ADX + ATR
MODEL = SVC
DATA = TSLA



	Backtest
Annualized Return	0.236036
Cumulative Returns	7.81432
Annual Volatility	0.188583
Sharpe Ratio	1.251627
Sortino Ratio	1.614629



TESTING
X = RSI + ADX + ATR
MODEL = SVC
DATA = TSLA



	Backtest
Annualized Return	39.355464
Cumulative Returns	3.5227
Annual Volatility	34.464912
Sharpe Ratio	1.141899
Sortino Ratio	1.155445

SOME NOTES FROM OUR TESTING

There are a lot of tests that we didn't include, and even more that we didn't record!

- Adding close price ruins the performance of any model we create
  - Its derivatives are probably just as bad
- Backtesting metrics can be misleading, are we really earning more than the market?

OVERFITTING

What values did we set for the program?

- Timeframe
- Training timex
- TA metrics
- SMA window
- TA metric periods

CONCLUSION