Stock Prediction

TEAM A

Antony Alexos

Anton Evmorfopoulos

Tilemachos Tsiapras

HERMES TEAM

Contents

- Introduction to Stock Prediction

Contents

- Introduction to Stock Prediction
- Data Engineering

HERMES TEAM

Contents

- Introduction to Stock Prediction
- Data Engineering
- Simple Models

HERMES TEAM

Contents

HERMES TEAM

Contents

- Introduction to Stock Prediction
- Data Engineering
- Simple Models
- More Data Engineering
- Simple Models but with more data

HERMES TEAM

Contents

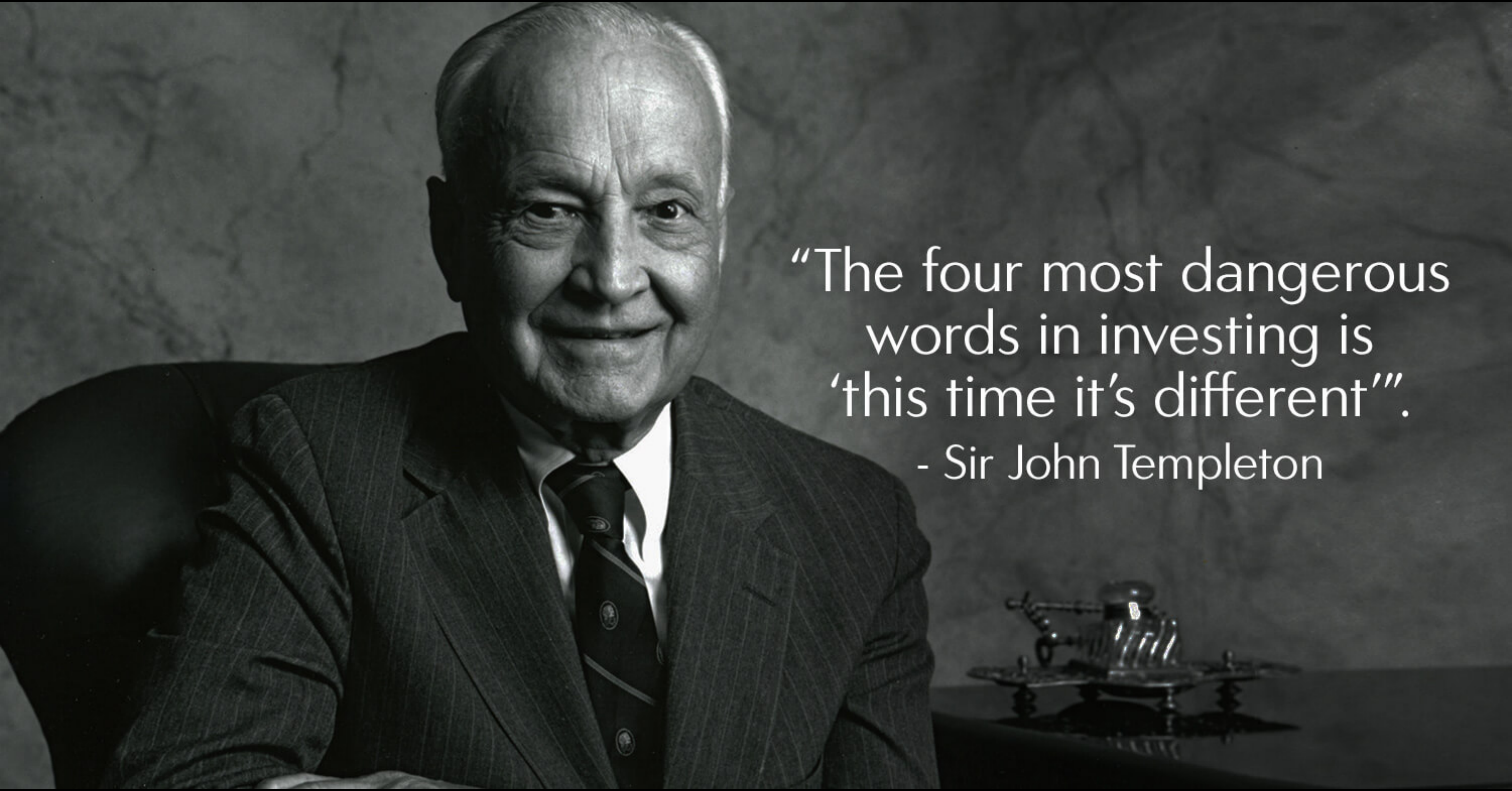
HERMES TEAM

Contents

HERMES
TEAM

Contents

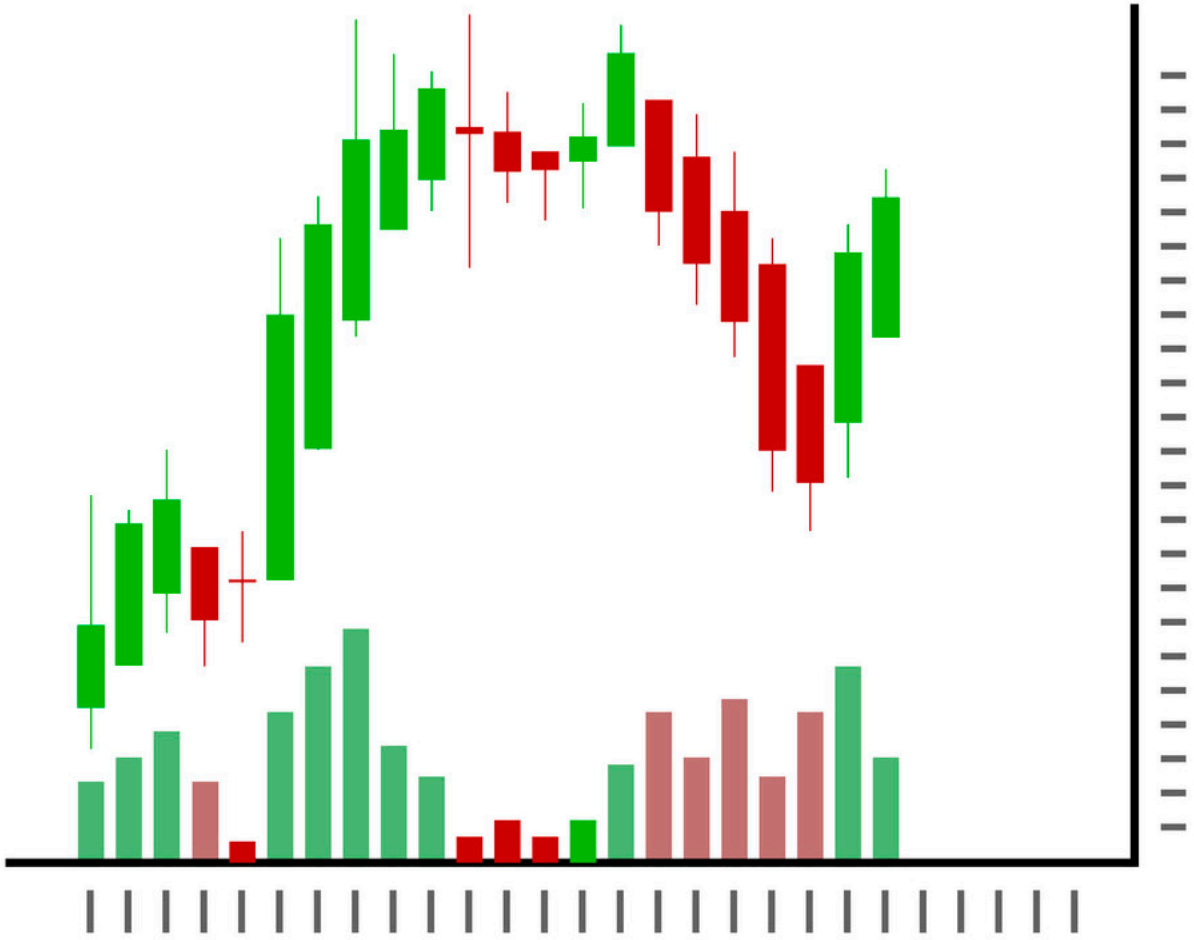
HERMES TEAM


"The four most dangerous
words in investing is
'this time it's different'".
- Sir John Templeton

Fundamental Analysis

Technical Analysis

HERMES TEAM



HERMES
TEAM

Machine Learning

Data Engineering

Getting the Data

- Yahoo API

```
df=pdr.get_data_yahoo(ticker,start,end)
```

HERMES TEAM

Getting the Data

- Yahoo API

```
df=pdr.get_data_yahoo(ticker,start,end)
```

- Provides values for Open, High, Low, Close, Volume and Date for the stock symbol.

Getting the Data

- Yahoo API

```
df=pdr.get_data_yahoo(ticker,start,end)
```

- Provides values for Open, High, Low, Close, Volume and Date for the stock symbol.
- We chose to predict the price of Goldman Sachs.

HERMES TEAM

Transforming the Data

- First we transform the data to a Time Series format

	GS_Open(t-1)	GS_High(t-1)	GS_Low(t-1)	GS_Close(t-1)	GS_Volume(t-1)	GS_Close(t)	Date
1	196.649994	196.830002	193.770004	193.830002	1566800.0	194.410004	2015-01-02
2	195.300003	195.729996	192.699997	194.410004	1877700.0	188.339996	2015-01-05
3	193.059998	194.039993	187.479996	188.339996	3413200.0	184.529999	2015-01-06
4	188.300003	188.660004	183.929993	184.529999	3429200.0	187.279999	2015-01-07
5	186.850006	187.990005	185.770004	187.279999	1896800.0	190.270004	2015-01-08

HERMES TEAM

Custom Gain Metric

- We assume that you trade daily without stop-losses or targets

Custom Gain Metric

- We assume that you trade daily without stop-losses or targets
- At Open you buy or sell according to the prediction we made for this day.

HERMES TEAM

Custom Gain Metric

- We assume that you trade daily without stop-losses or targets
- At Open you buy or sell according to the prediction we made for this day.
- If you predict the same direction with the real one -> $$$$!

HERMES TEAM

Custom Gain Metric

- We assume that you trade daily without stop-losses or targets
- At Open you buy or sell according to the prediction we made for this day.
- If you predict the same direction with the real one -> $$$$!
- If you predict the opposite -> you just lost money.. :(

HERMES TEAM

Custom Gain Metric

- We assume that you trade daily without stop-losses or targets
- At Open you buy or sell according to the prediction we made for this day.
- If you predict the same direction with the real one -> $$$$!
- If you predict the opposite -> you just lost money.. :(
- The gain is the |Open-Close|

HERMES TEAM

Simple Models

The simplicity of the model is based on the simplicity of the data

HERMES TEAM

LSTM - Long Short Term Memory

Types of Models

- Our models are based on three things
  a) passing the data through time-steps dimension
  b) passing the data through feature dimension
  c) making the model with memory between batches

HERMES TEAM

Types of Models

- Our models are based on three things
  a) passing the data through time-steps dimension
  b) passing the data through feature dimension
  c) making the model with memory between batches
- We implement dropout to the model that has produced the best results.

HERMES TEAM

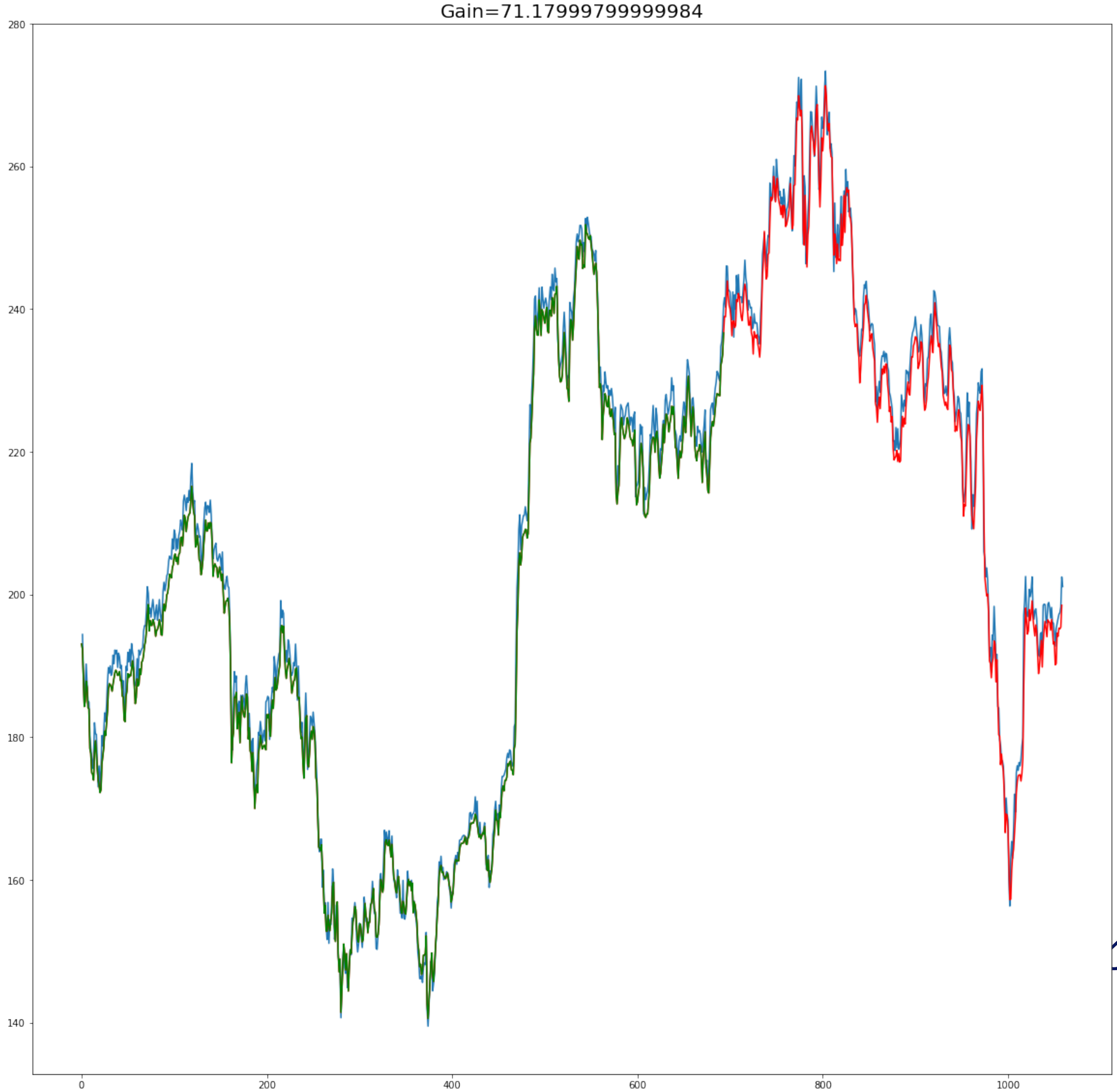
Types of Models

- Our models are based on three things
  a) passing the data through time-steps dimension
  b) passing the data through feature dimension
  c) making the model with memory between batches
- We implement dropout to the model that has produced the best results.
- We predict the prices of 1.5 year(365 trading days)

LSTM - Long Short Term Memory

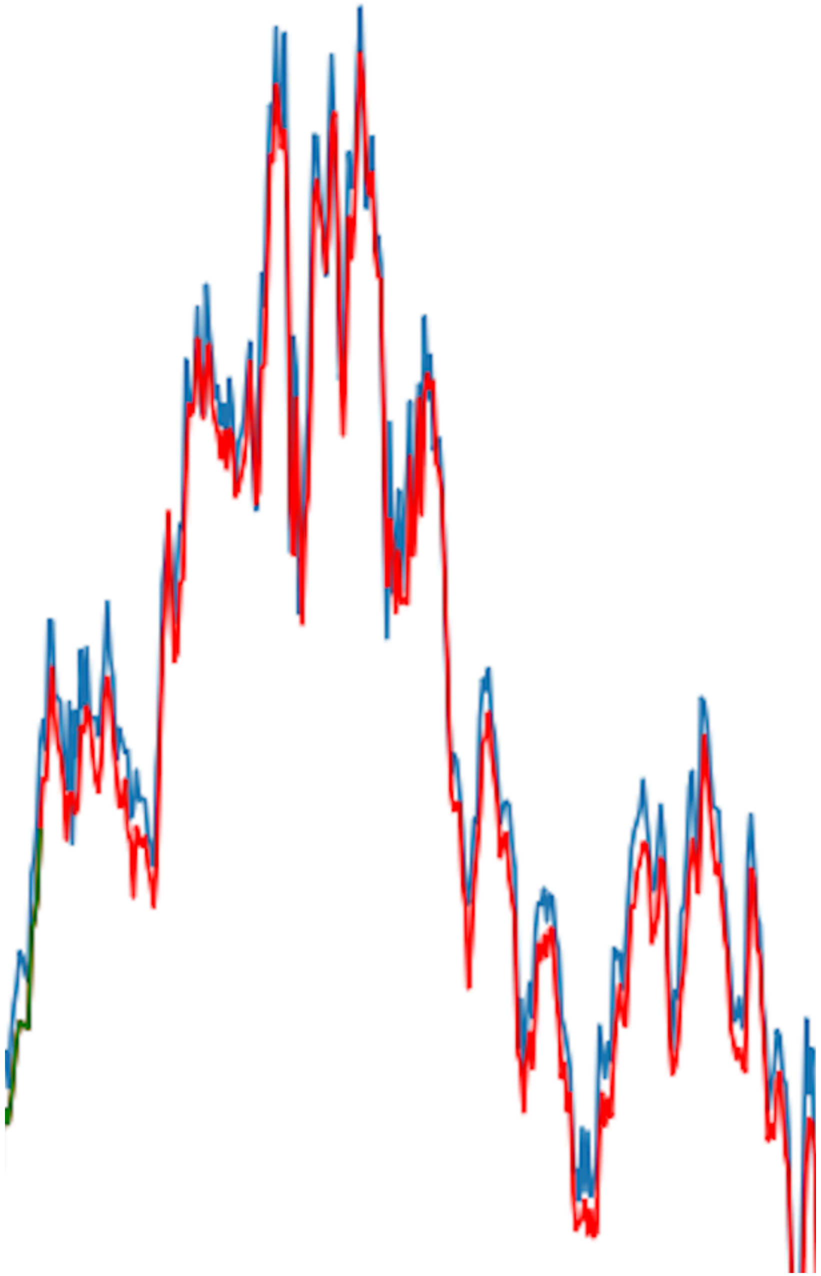
- [samples, time steps, features]
- X = (dataX, (len(dataX), seq_length, 1))

HERMES TEAM



Gain=71.17999799999984
280
260
240
220
200
180
160
140
0
200
400
600
800
1000

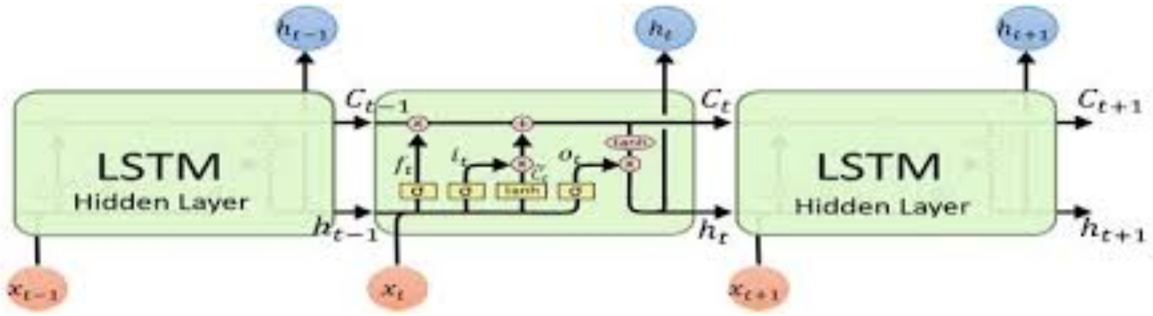
HERMES TEAM

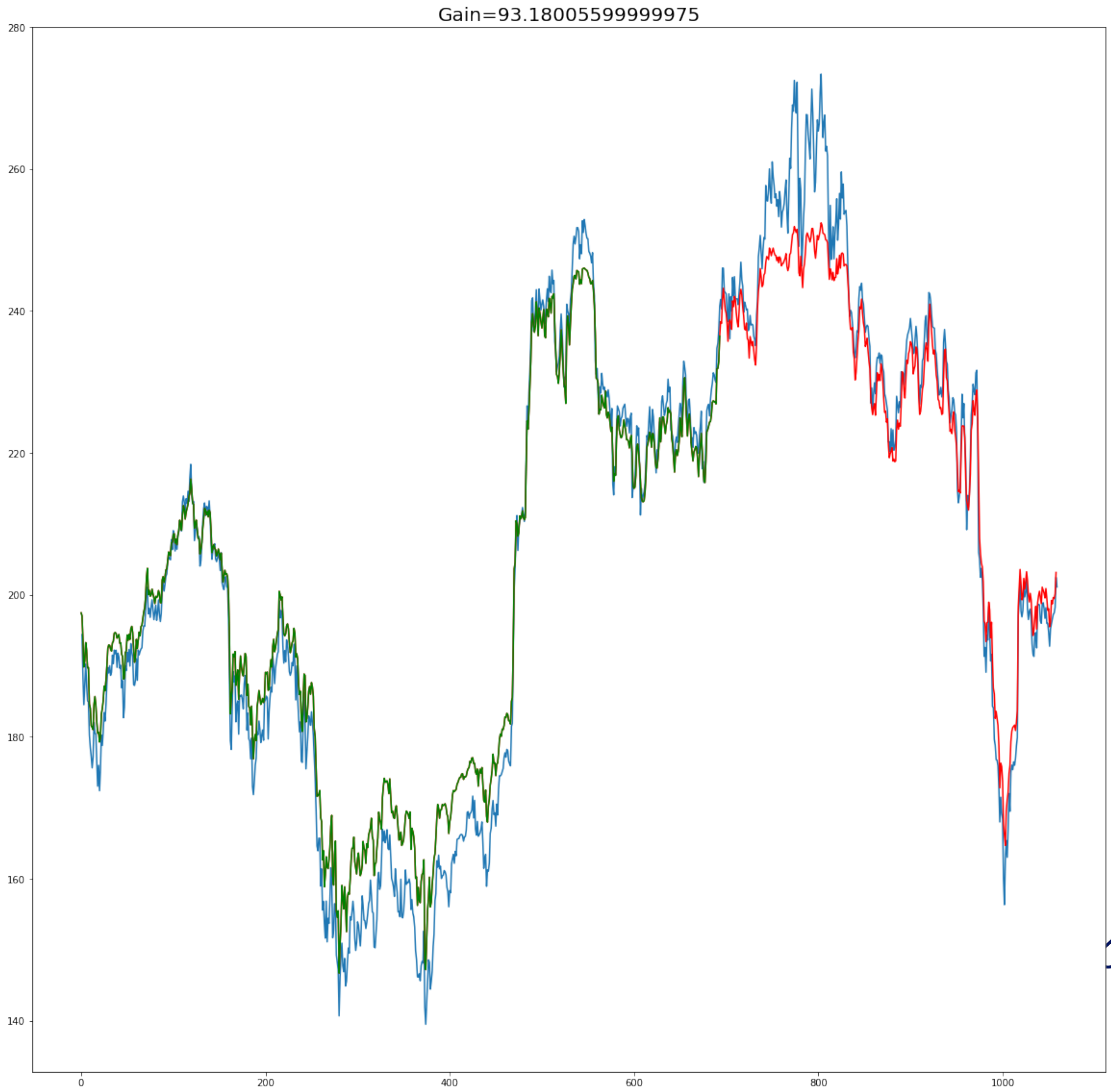


HERMES TEAM

Stacked LSTM - Stacked Long Short Term Memory

- Stacked LSTM with passing data as features
- Stacked LSTM with passing data as time steps
- Stacked LSTM with memory batches




Gain=93.18005599999975
280
260
240
220
200
180
160
140
0
200
400
600
800
1000

HERMES TEAM

More Data Engineering

Technical Indicators, more symbols and other features

More Symbols

HERMES TEAM

More Symbols

- NASDAQ, Hang Seng Index, NYSE, Nikkei 225

More Symbols

- NASDAQ, Hang Seng Index, NYSE, Nikkei 225
- Bank of America, Barclays, Credit Suisse, JPMorgan, Morgan Stanley

HERMES TEAM

More Symbols

- NASDAQ, Hang Seng Index, NYSE, Nikkei 225
- Bank of America, Barclays, Credit Suisse, JPMorgan, Morgan Stanley
- VIX

More Symbols

- NASDAQ, Hang Seng Index, NYSE, Nikkei 225
- Bank of America, Barclays, Credit Suisse, JPMorgan, Morgan Stanley
- VIX
- We keep only Close

Technical Indicators

- Moving Average 7 and 21

Technical Indicators

- Moving Average 7 and 21
- Exponential Moving Average(EMA)

Technical Indicators

- Moving Average 7 and 21
- Exponential Moving Average(EMA)
- Moving Average Convergence Divergence(MACD)

HERMES TEAM

Technical Indicators

- Moving Average 7 and 21
- Exponential Moving Average(EMA)
- Moving Average Convergence Divergence(MACD)
- Bollinger Bands

HERMES
TEAM

Technical Indicators

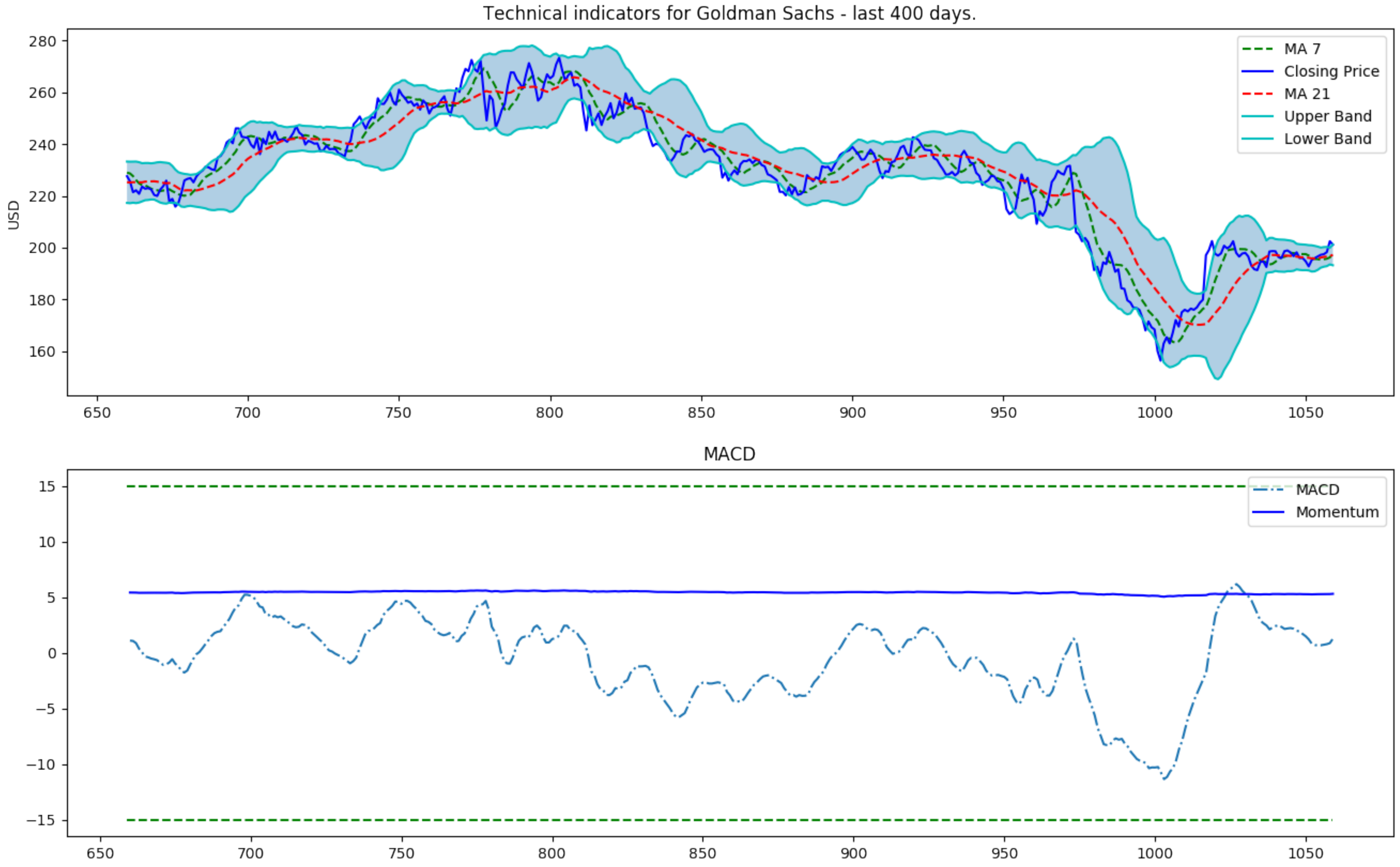
- Moving Average 7 and 21
- Exponential Moving Average(EMA)
- Moving Average Convergence Divergence(MACD)
- Bollinger Bands
- Momentum

HERMES TEAM

Technical Indicators

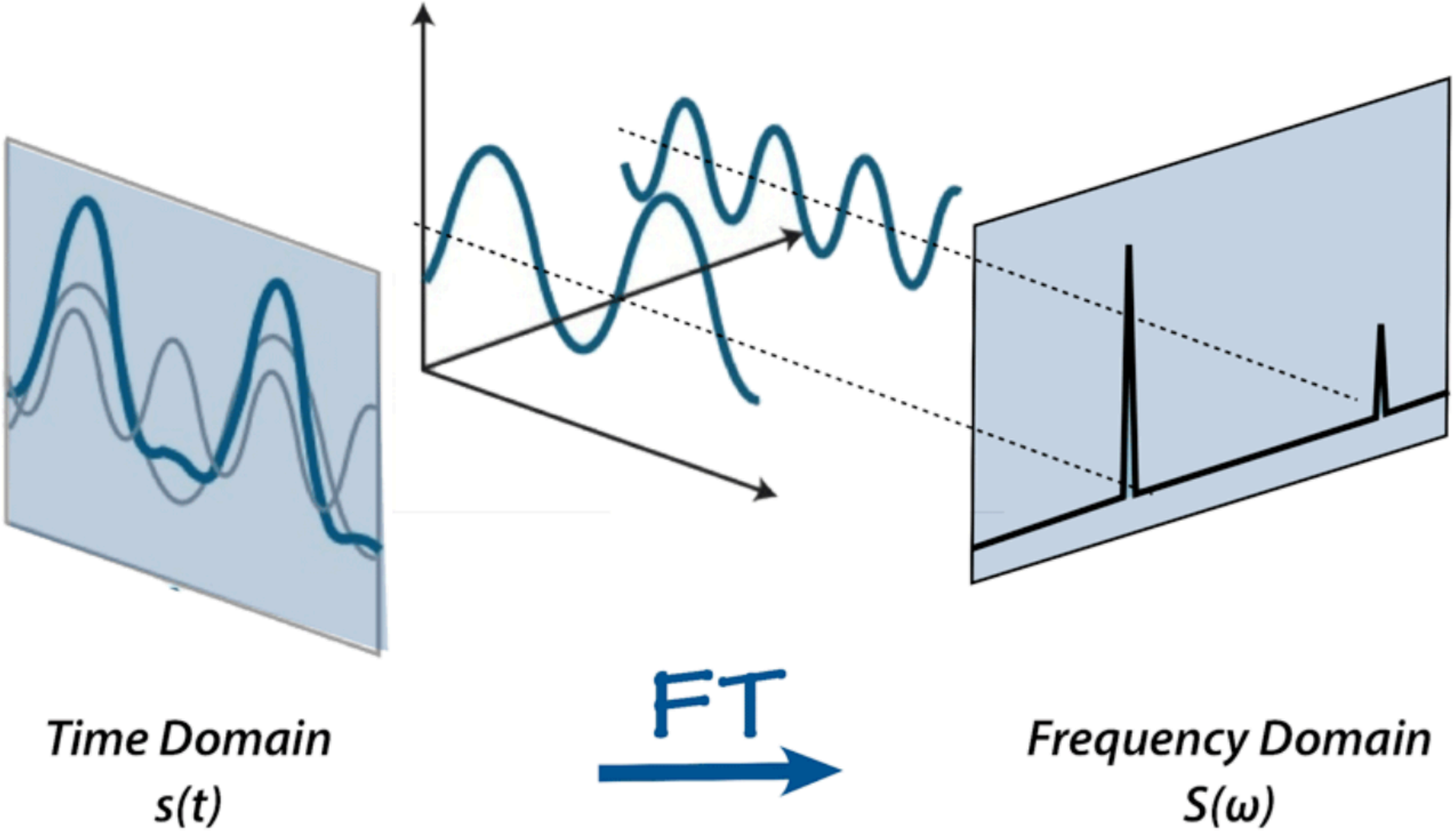
- Moving Average 7 and 21
- Exponential Moving Average(EMA)
- Moving Average Convergence Divergence(MACD)
- Bollinger Bands
- Momentum
- Log Momentum

HERMES TEAM



Technical indicators for Goldman Sachs - last 400 days.
USD
MA 7
Closing Price
MA 21
Upper Band
Lower Band
MACD
MACD
Momentum

Fourier



$$S(\omega) = \int_{-\infty}^{\infty} s(t) e^{-i\omega t} dt$$

HERMES TEAM

Fourier

- We use fourier to smooth the time series.

Fourier

- We use fourier to smooth the time series.
- Find patterns and trends.

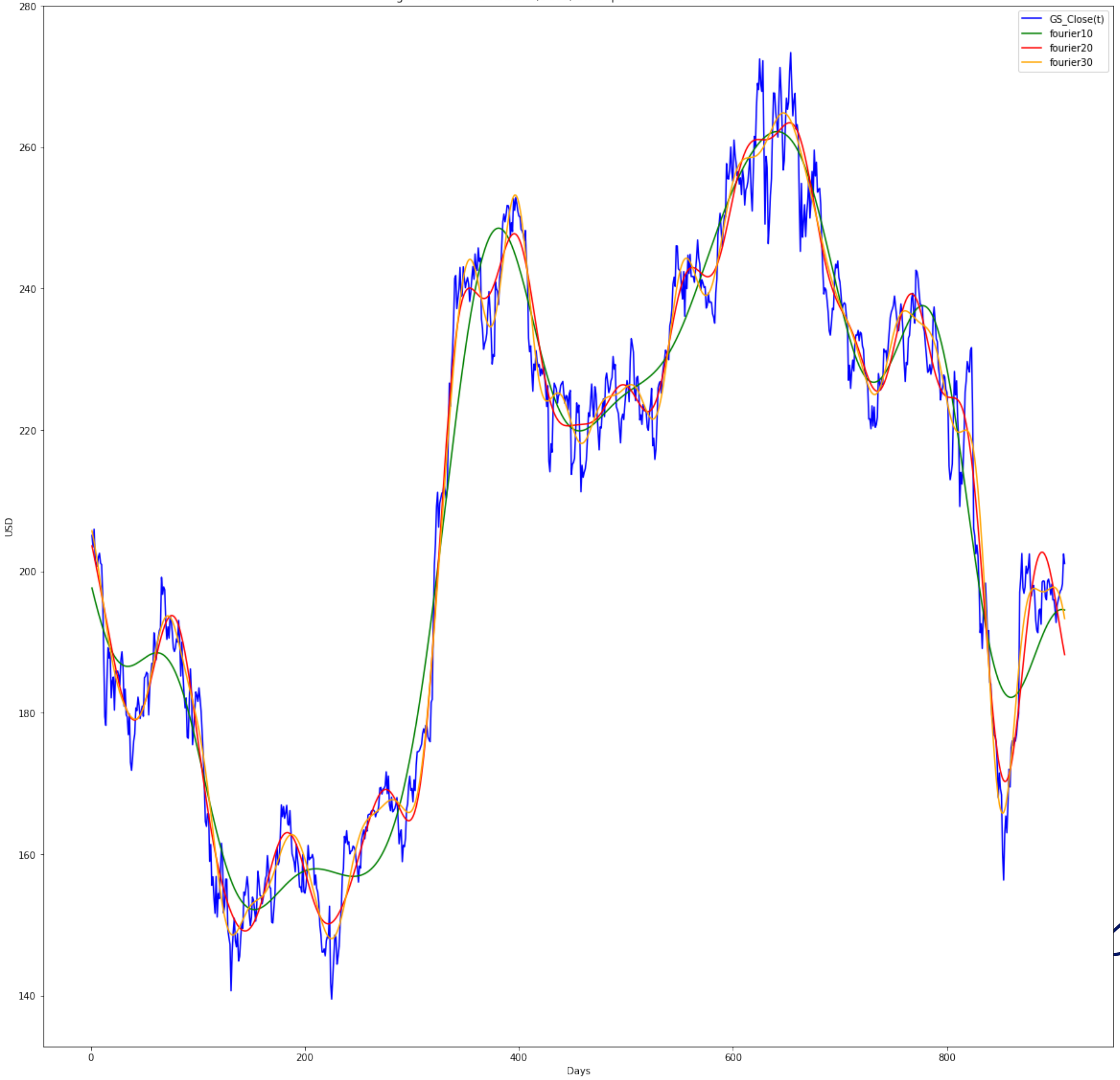
Fourier

- We use fourier to smooth the time series.
- Find patterns and trends.
- We basically denoise the data

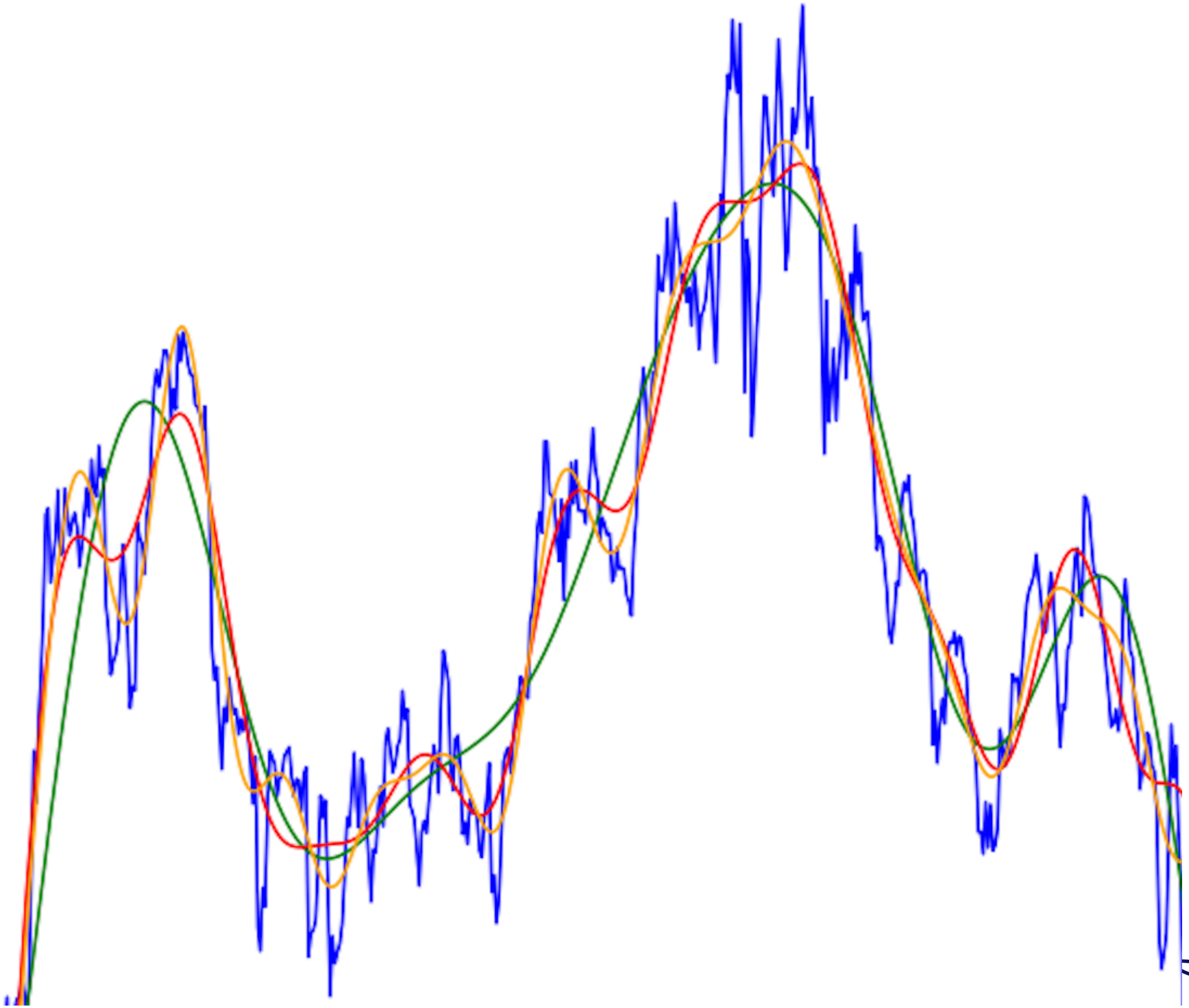
```
def filter_signal10(signal, threshold=1e3):
    fourier = rfft(signal)
    frequencies = rfftfreq(signal.size, d=10e-3/signal.size)  #change the number to change the plot
    fourier[frequencies > threshold] = 0
    return irfft(fourier)
```

HERMES TEAM

Figure 3: Goldman Sachs (close) stock prices & Fourier transforms



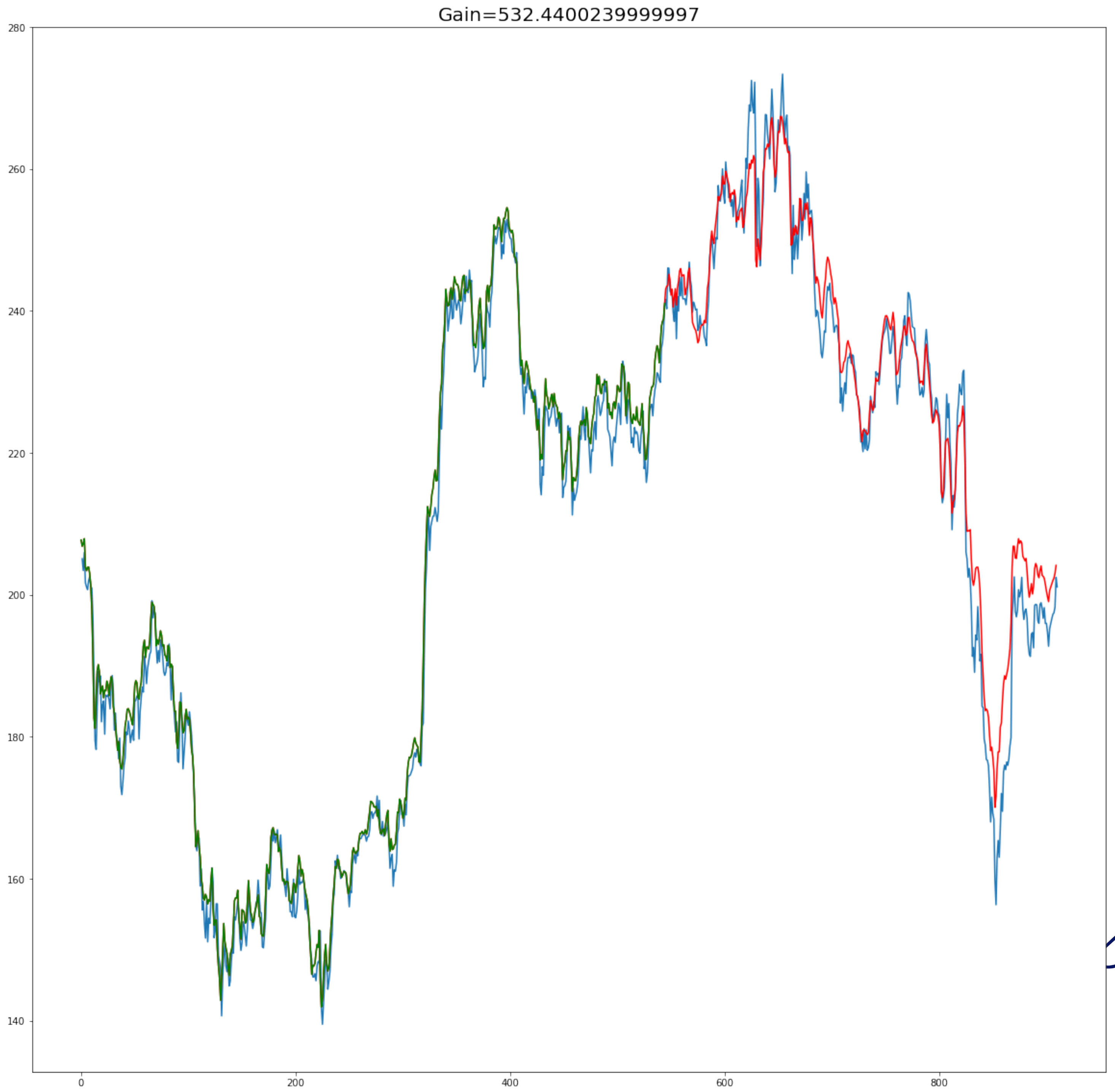
HERMES TEAM



HERMES TEAM

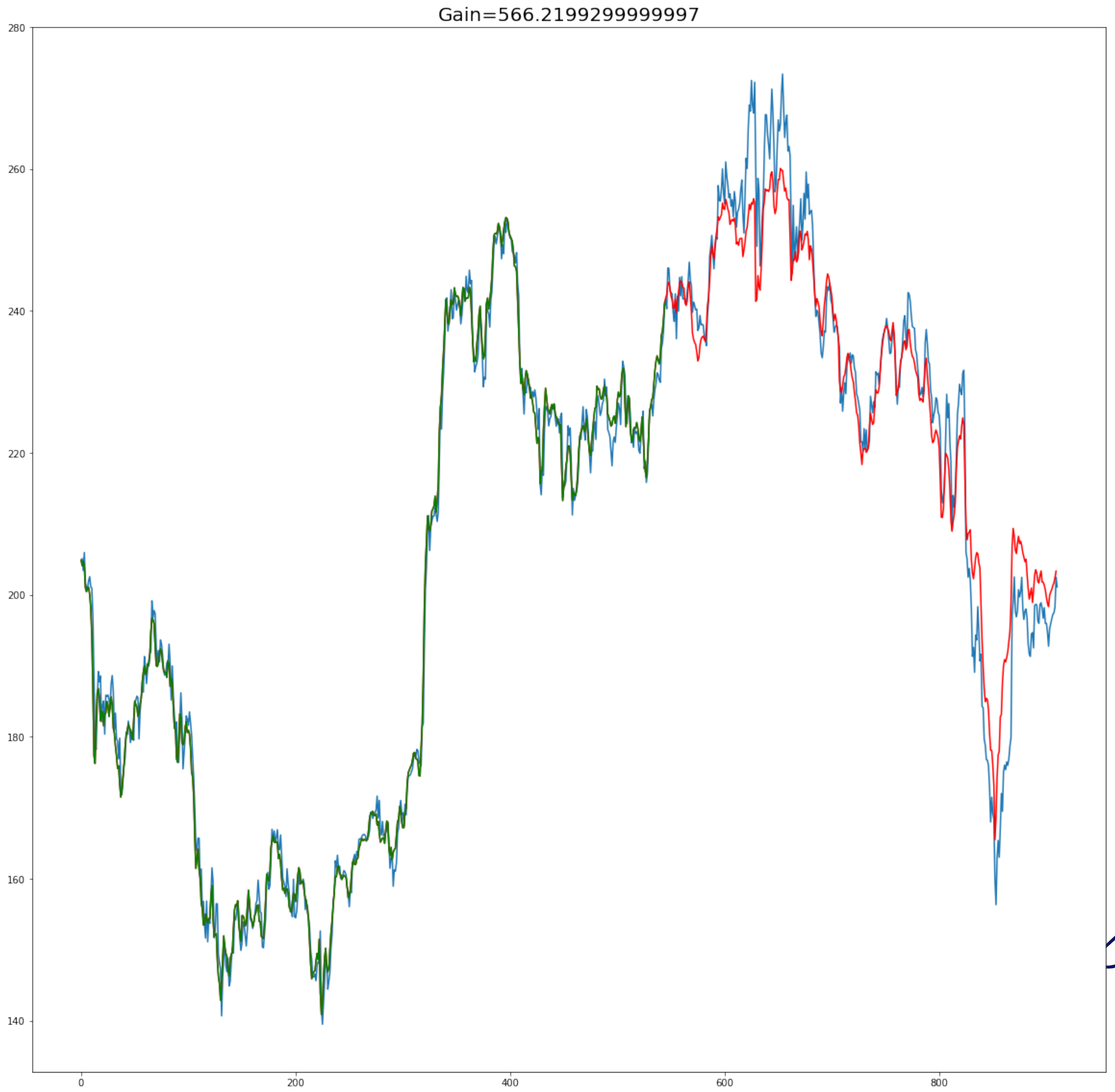
We try the same models but with more complicated Data

HERMES TEAM



Gain=532.4400239999997
280
260
240
220
200
180
160
140
0
200
400
600
800

HERMES TEAM


Gain=566.2199299999997
280
260
240
220
200
180
160
140
0
200
400
600
800

Feature Importance

XGBoost

- XGBoost(e**X**treme **G**radient **B**oosting) is an implementation of gradient boosted decision trees designed for speed and performance.

HERMES TEAM

XGBoost

- XGBoost(e**X**treme **G**radient **Boost**ing) is an implementation of gradient boosted decision trees designed for speed and performance.
- Some key algorithm implementation features include:
  a) **Sparse Aware** implementation with automatic handling of missing data values.
  b) **Block Structure** to support the parallelization of tree construction.
  c) **Continued Training** so that you can further boost an already fitted model on new data.

HERMES TEAM

XGBoost

- Subsequent trees, they learn from their predecessors.

HERMES TEAM

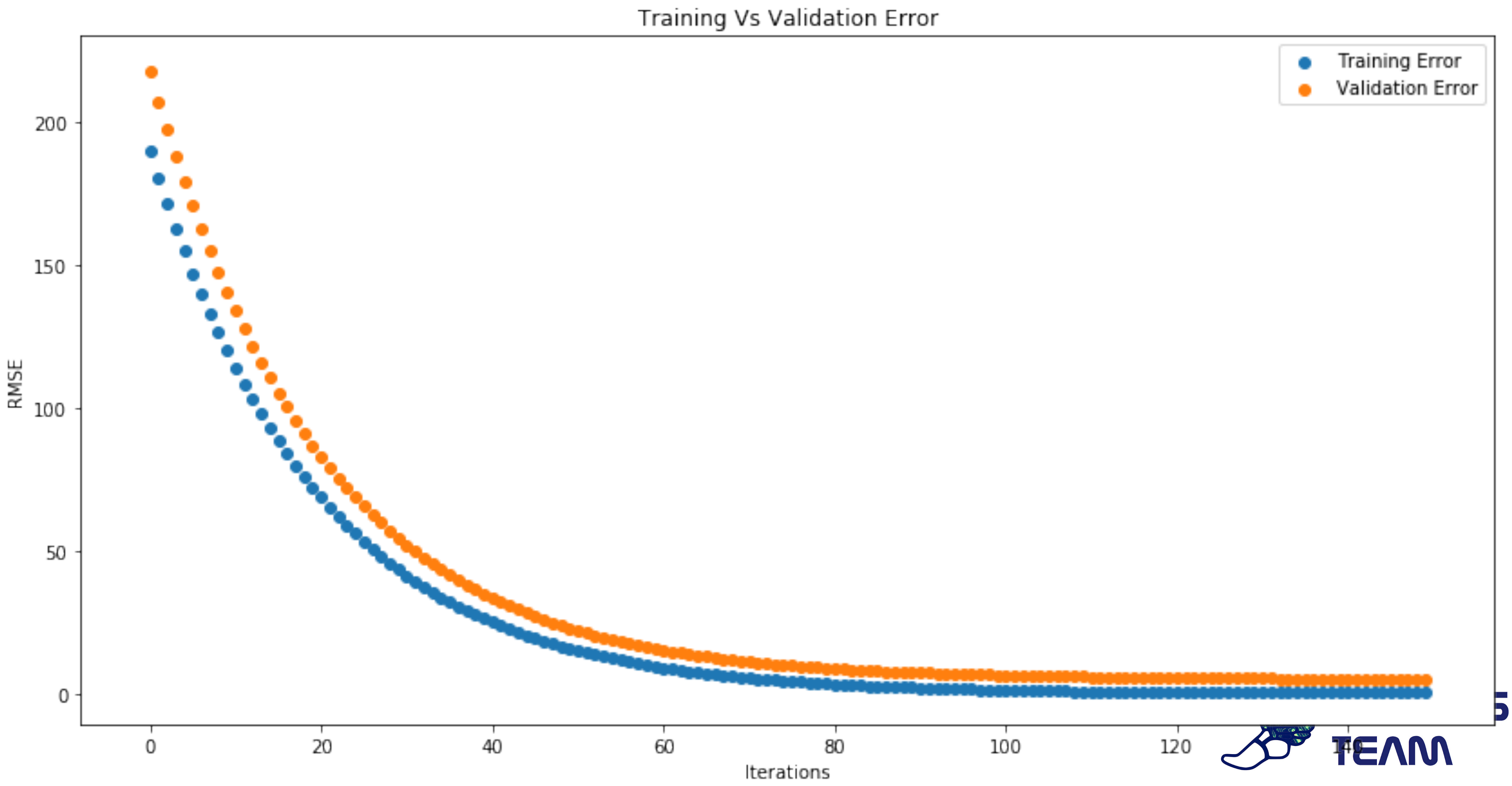
XGBoost

- Subsequent trees, they learn from their predecessors.
- Gradient Descent because we have a minimization problem.

XGBoost

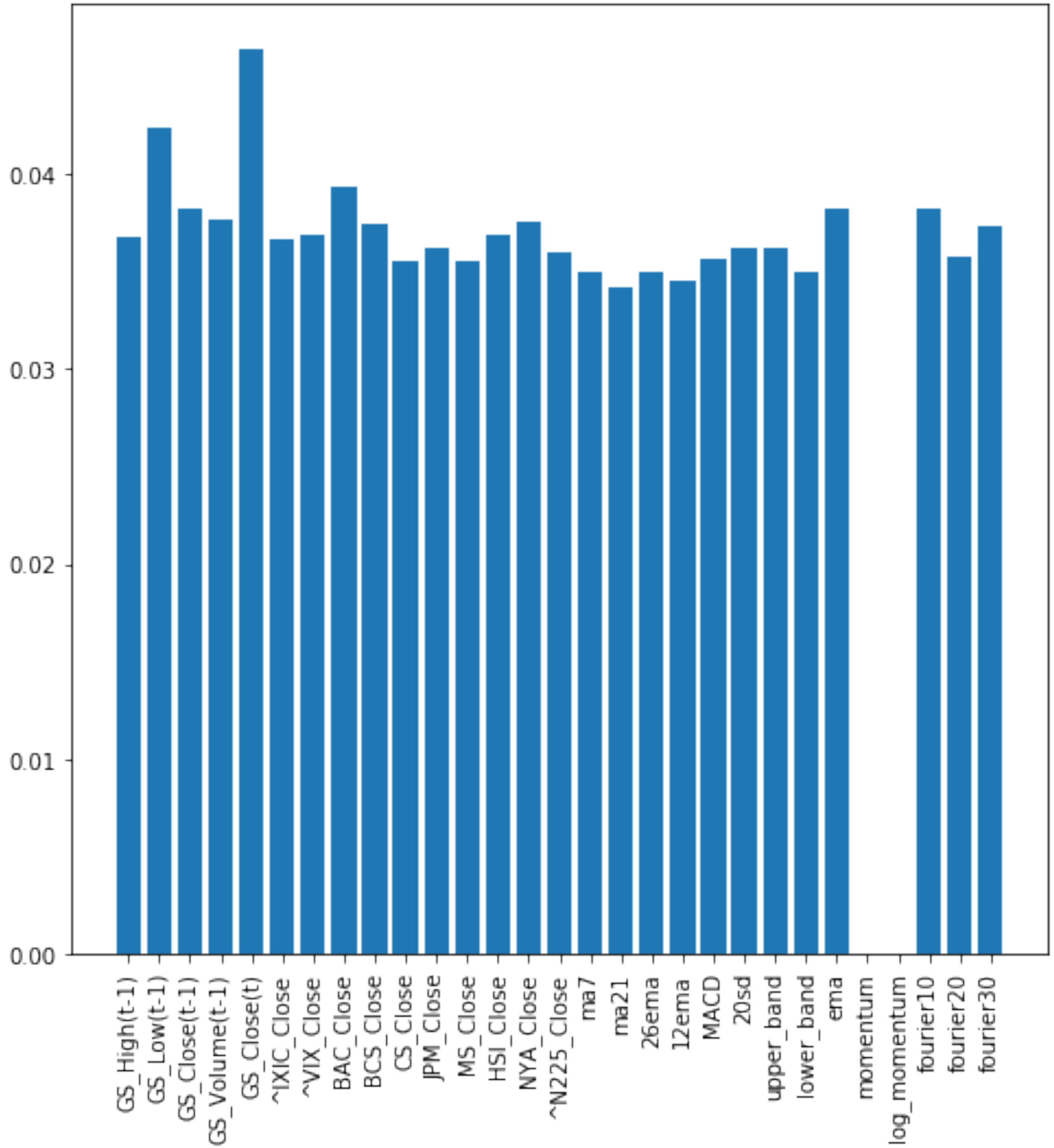
- Subsequent trees, they learn from their predecessors.
- Gradient Descent because we have a minimization problem.
- It is a meta machine learning algorithm that builds a strong model based on many weaker ones sequently.

HERMES TEAM



Training Vs Validation Error
Training Error
Validation Error
RMSE
0
50
100
150
200
0
20
40
60
80
100
120
140
Iterations
TEAM

Figure 6: Feature importance of the technical indicators.



HERMES TEAM

PCA

- We reduce the dimensions from 30 to the 5 most important ones.

HERMES TEAM

PCA

- We reduce the dimensions from 30 to the 5 most important ones.
- How it works:
  a) Calculate the covariance matrix X of data points.

HERMES TEAM

PCA

- We reduce the dimensions from 30 to the 5 most important ones.
- How it works:
  a) Calculate the covariance matrix X of data points.
  b) Calculate eigen vectors and corresponding eigen values.

PCA

- We reduce the dimensions from 30 to the 5 most important ones.
- How it works:
  a) Calculate the covariance matrix *X* of data points.
  b) Calculate eigen vectors and corresponding eigen values.
  c) Sort the eigen vectors according to their eigen values in decreasing order.

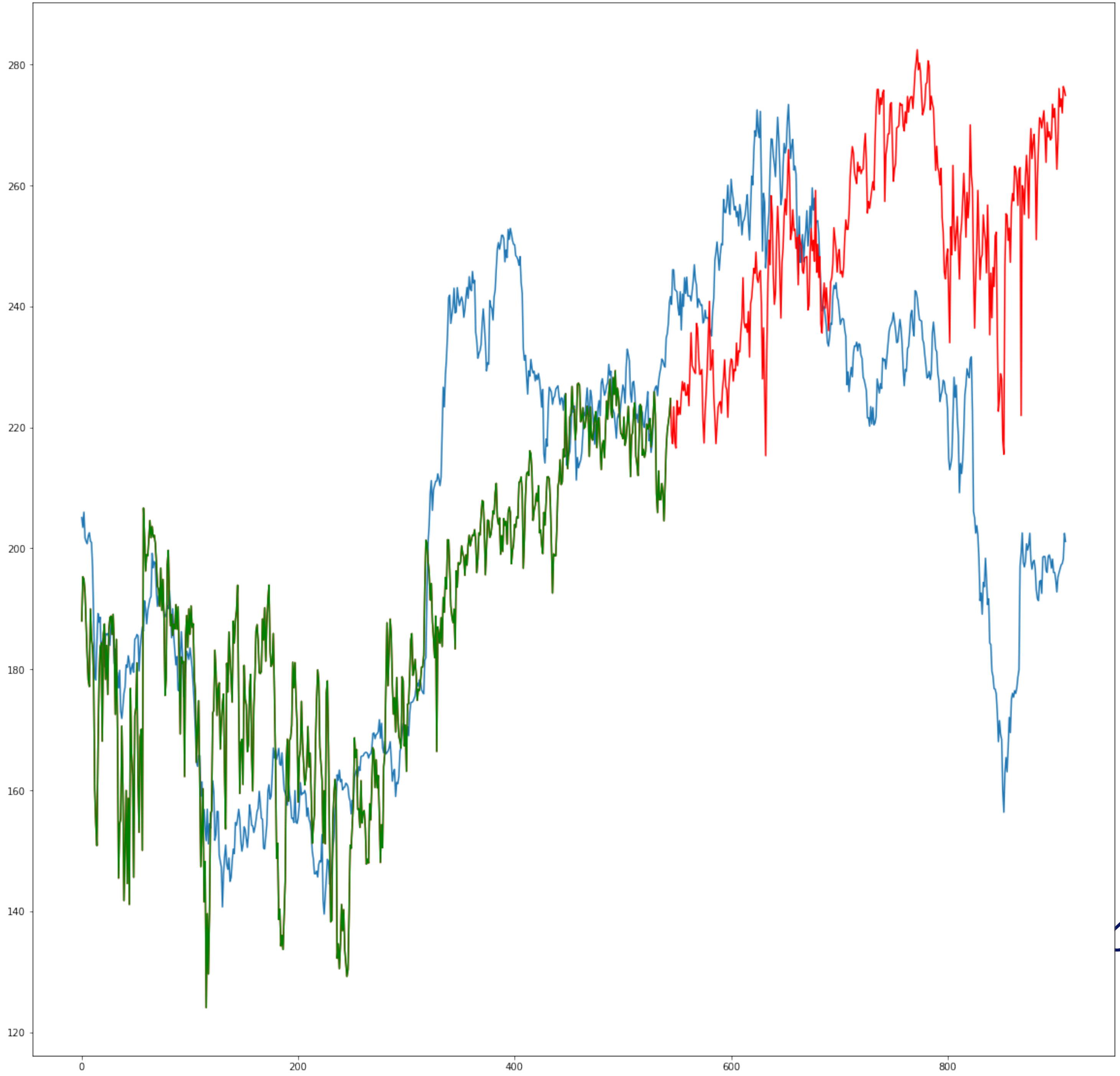
PCA

- We reduce the dimensions from 30 to the 5 most important ones.
- How it works:
  a) Calculate the covariance matrix X of data points.
  b) Calculate eigen vectors and corresponding eigen values.
  c) Sort the eigen vectors according to their eigen values in decreasing order.
  d) Choose first k eigen vectors and that will be the new k dimensions.

PCA

- We reduce the dimensions from 30 to the 5 most important ones.
- How it works:
  a) Calculate the covariance matrix *X* of data points.
  b) Calculate eigen vectors and corresponding eigen values.
  c) Sort the eigen vectors according to their eigen values in decreasing order.
  d) Choose first k eigen vectors and that will be the new k dimensions.
  e) Transform the original n dimensional data points into k dimensions.

Gain=367.00012400000026



HERMES TEAM

Classification

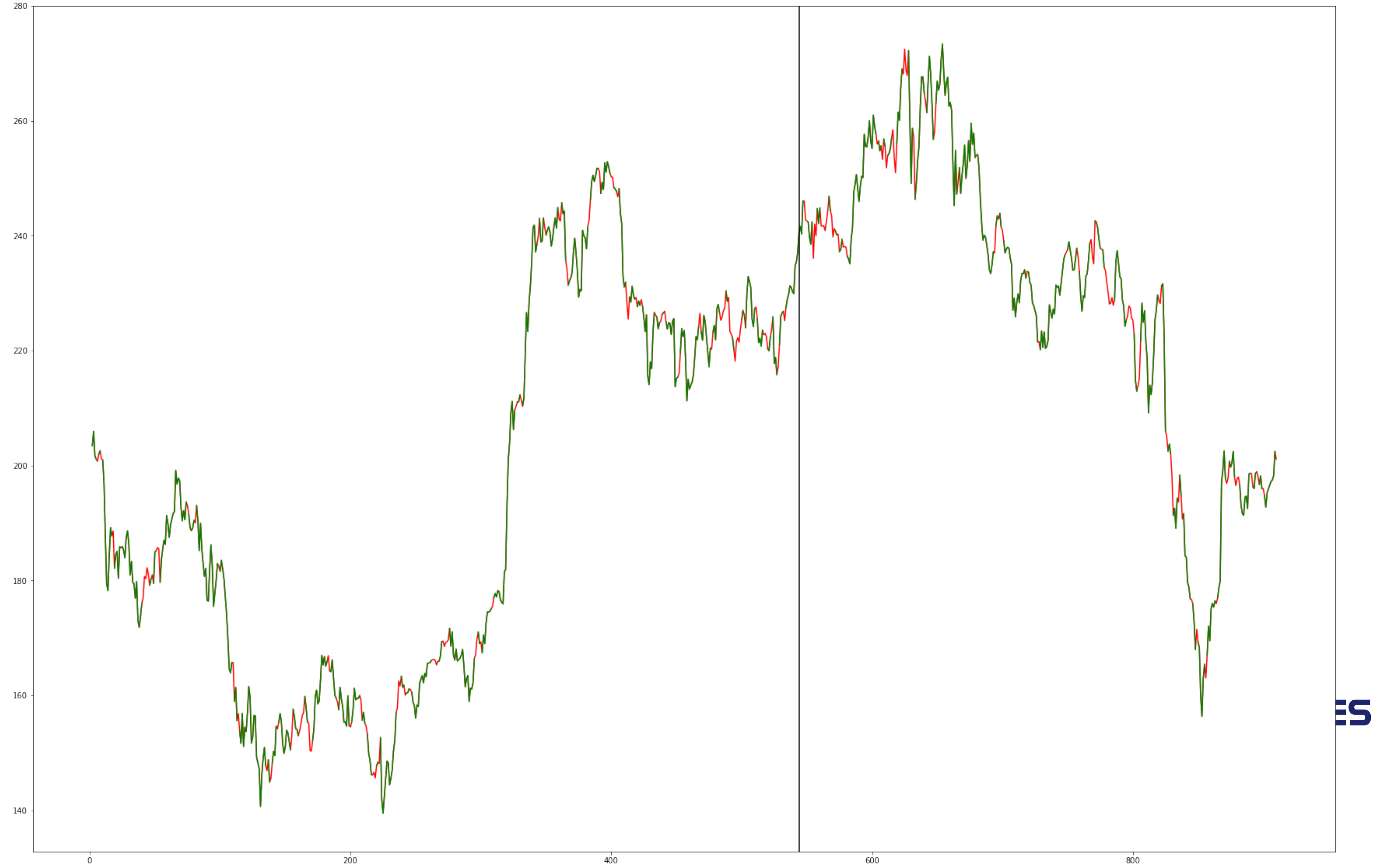
HERMES TEAM

Classification

	GS_Open(t-1)	GS_Close(t-1)	GS_Volume(t-1)	^IXIC_Close	^VIX_Close	BAC_Close	BCS_Close	CS_Close	JPM_Close	MS_Close
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	1.0	1.0	0.0	1.0	1.0	1.0	1.0	1.0
4	0.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
5	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
6	0.0	0.0	1.0	0.0	0.0	1.0	1.0	1.0	1.0	1.0
7	1.0	0.0	1.0	1.0	0.0	1.0	1.0	0.0	1.0	1.0
8	1.0	1.0	1.0	1.0	1.0	1.0	0.0	1.0	1.0	1.0
9	1.0	1.0	0.0	0.0	1.0	0.0	1.0	1.0	1.0	0.0
10	1.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
11	0.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
12	0.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
13	0.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
14	0.0	0.0	1.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0
15	1.0	0.0	1.0	1.0	0.0	1.0	1.0	1.0	1.0	1.0
16	0.0	1.0	0.0	1.0	0.0	1.0	1.0	1.0	1.0	1.0

Classification

HERMES TEAM


280
260
240
220
200
180
160
140
0
200
400
600
800

HERMES TEAM

Future Work

HERMES TEAM

Future Work

- More Complex models

Future Work

- More Complex models
- GAN

Future Work

- More Complex models
- GAN
- Tweets for classification(BERT)

Future Work

- More Complex models
- GAN
- Tweets for classification(BERT)
- Predicting the change of price

HERMES TEAM

Future Work

- More Complex models
- GAN
- Tweets for classification(BERT)
- Predicting the change of price
- Making a model to predict wether we do a trade or not.

HERMES TEAM

Future Work

- More Complex models
- GAN
- Tweets for classification(BERT)
- Predicting the change of price
- Making a model to predict wether we do a trade or not.
- Translate the code into MQL4

HERMES TEAM