



Capstone Project with Spinnaker Analytics

Kaiyu Wang, Simeng Li, Xiangshan Mu, Yihan Jia, Yinghao Wang
Team A6

Abstract and Objectives

Business Objective:

- Find a **tradable signal** in the dataset

Business Application:

- Anticipating the fluctuations of fund prices and tradable signals in advance can help fund traders make better investment decisions. We will make recommendations for investors based on fund types, short-term or long-term given clients' needs and wants.

About the Dataset:

- Our dataset contains sectoral data for 3 separate types of investments made in the US. It represents 60-70% activity in the overall asset classes. The weekly data spans **10** years from 2006 through early 2017.



Data Analytics cont.

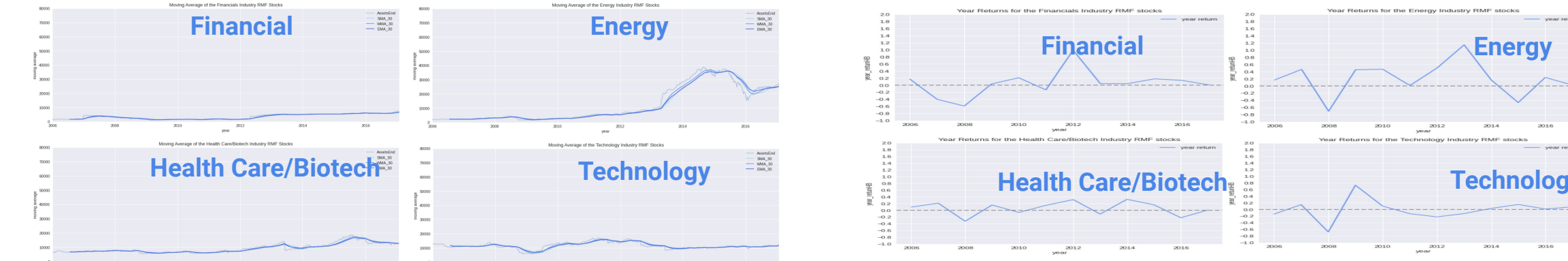
- ETF Type



- IMF Type



- RMF Type



Multi-class Classification Model

Classification Analysis:

- One advantage of classification is we can give a global optimization using only one model.
- From 10 classifiers of sklearn module. BEST classifier is GBDT.
- The final model has **67%** accuracy.

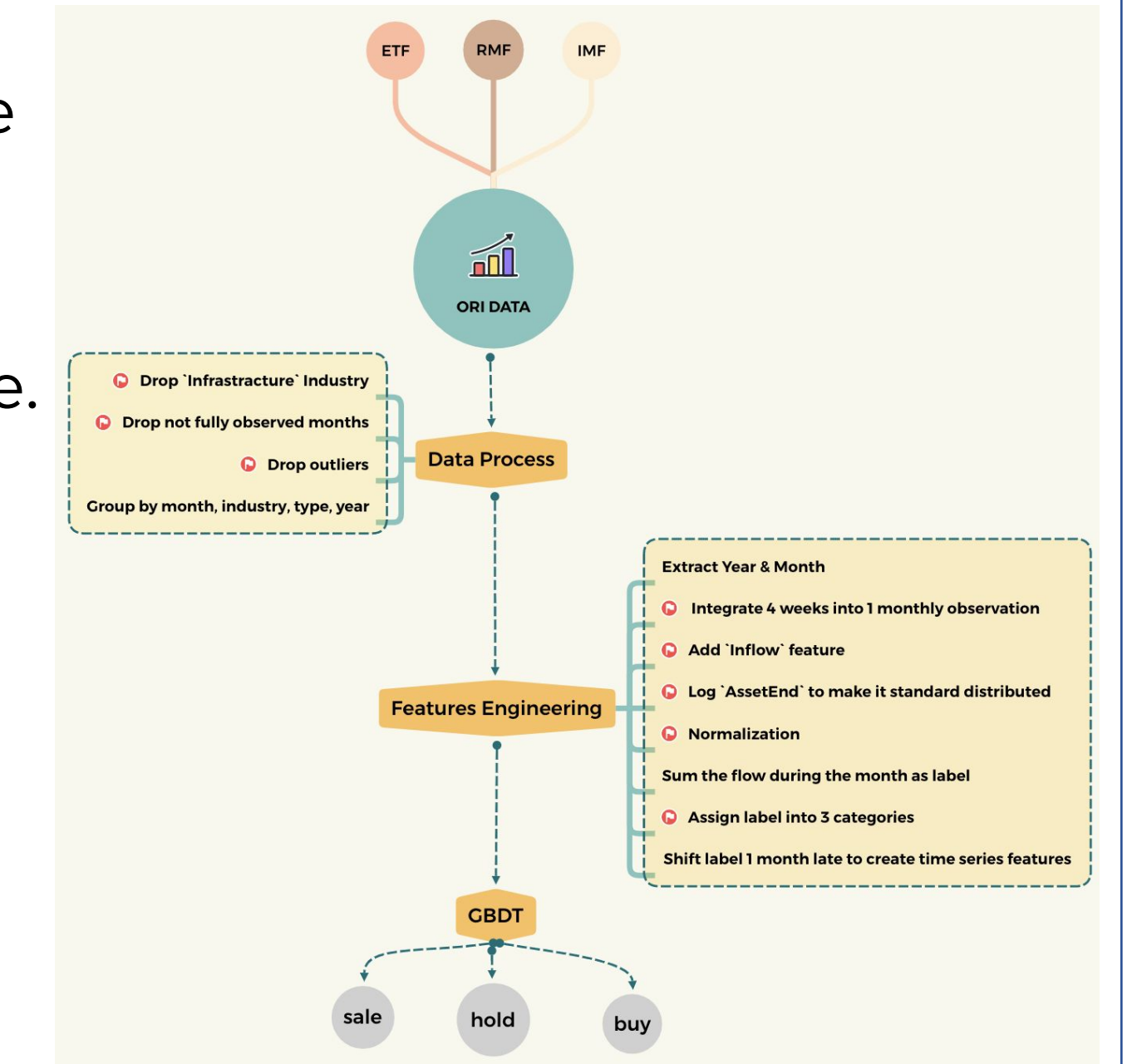
```
Start Training: GBDT
accuracy is 0.672448

precision    recall  f1-score   support

-1         0.64      0.65      0.65      507
 0         0.74      0.69      0.71      459
 1         0.64      0.68      0.66      533

accuracy          0.67      1499
macro avg          0.68      0.67      0.67      1499
weighted avg          0.67      0.67      0.67      1499

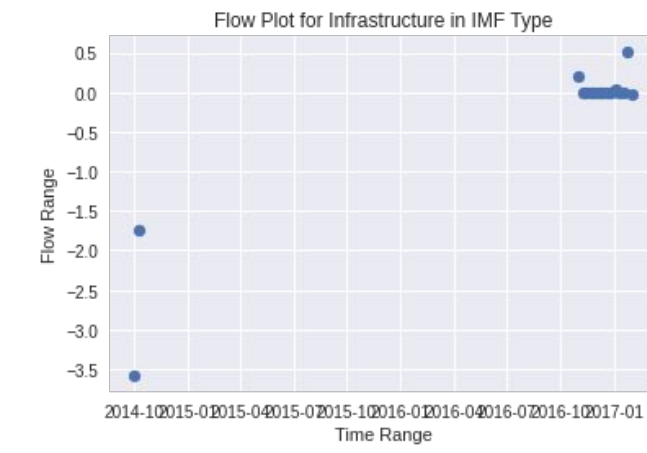
Confusion Matrix...
[[331  46 130]
 [ 74 315  70]
 [109  62 362]]
```



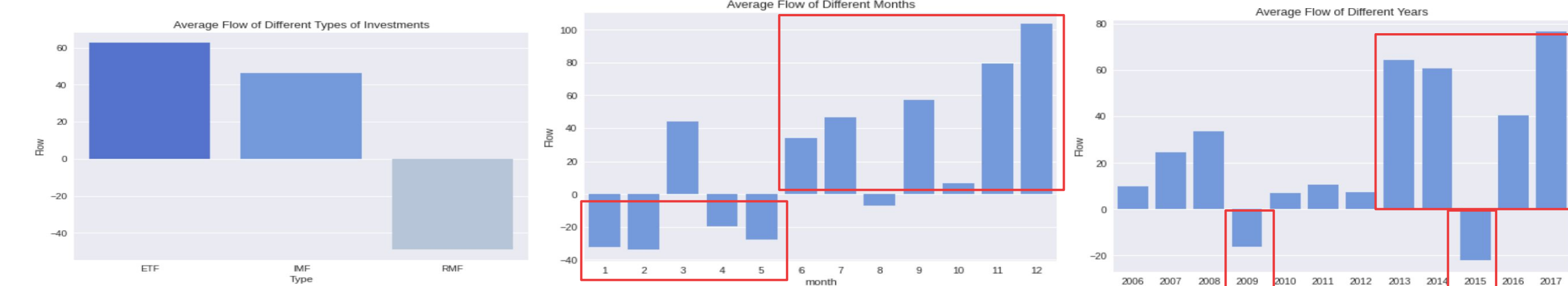
Data Analytics

Continuity Check:

- check there exist missing values in 4 industries:
- Commodities/Materials, Consumer Goods, Industrials, and Infrastructure.



Average Flow:



- We can speculate that most of the funds **flowed in** in the **second half of 2013 - 2017**
- The correlation between **Flow** and **AssetsEnd** is **large** but the correlation between **Flow** and **PortfolioChangePct** is **small**.



PortfolioChangePct:

- The **Financial** and **Energy** industries are more **stable** and suitable for **cautious investors**. Also, they have small fluctuations and large distribution, which means that it is easier for investors to **make profits**.
- **Healthcare** and **Technology** industries are more **risky** and suitable for **open investors**. Also, they're highly volatile but in small distribution, which means that **even a huge drop won't bring huge losses**.

Smoothing & Year Returns:

- The annual returns in the **Health Care/Biotech** and **Technology** industries have shown an **upward trend**.
- In general, the trend of IMF is almost **the same** as that of ETF funds.
- RMF-type funds have the **smallest and smoothes** values.

Time Series Model

Time Series Analysis:

3 Types

16 Industries

48 Subsets

Stationary Check

1. Moving Average
2. Simple Exponential Smoothing
3. AutoRegressive Moving Average
4. Long Short-Term Memory

- with Local Optimization approach

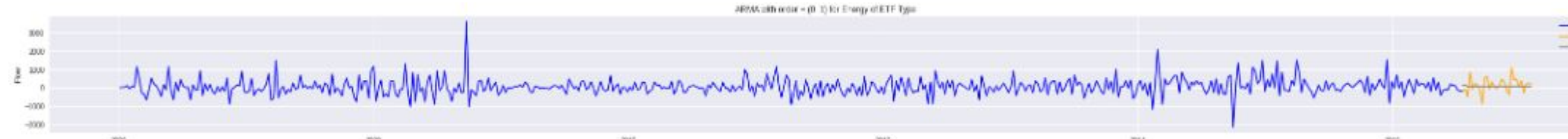
Moving Average:



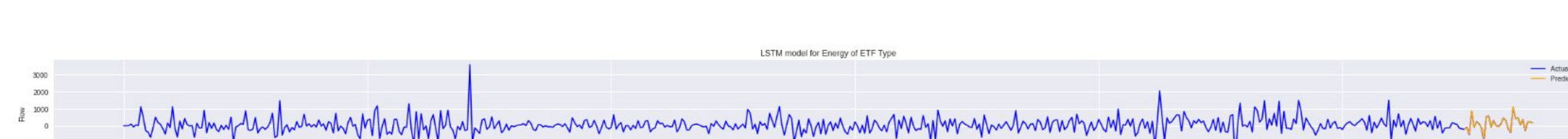
Simple Exponential Smoothing:



AutoRegressive Moving Average:



Long-Short Term Memory:



- LSTM with StandardScaled input generates the most accurate future flow forecast output.

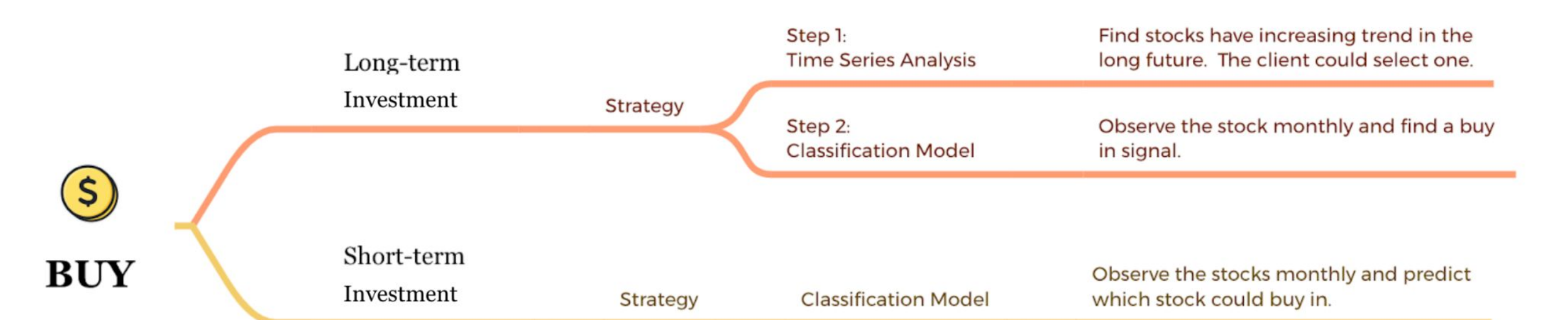
Proposed Business Applications

Model Conclusion:

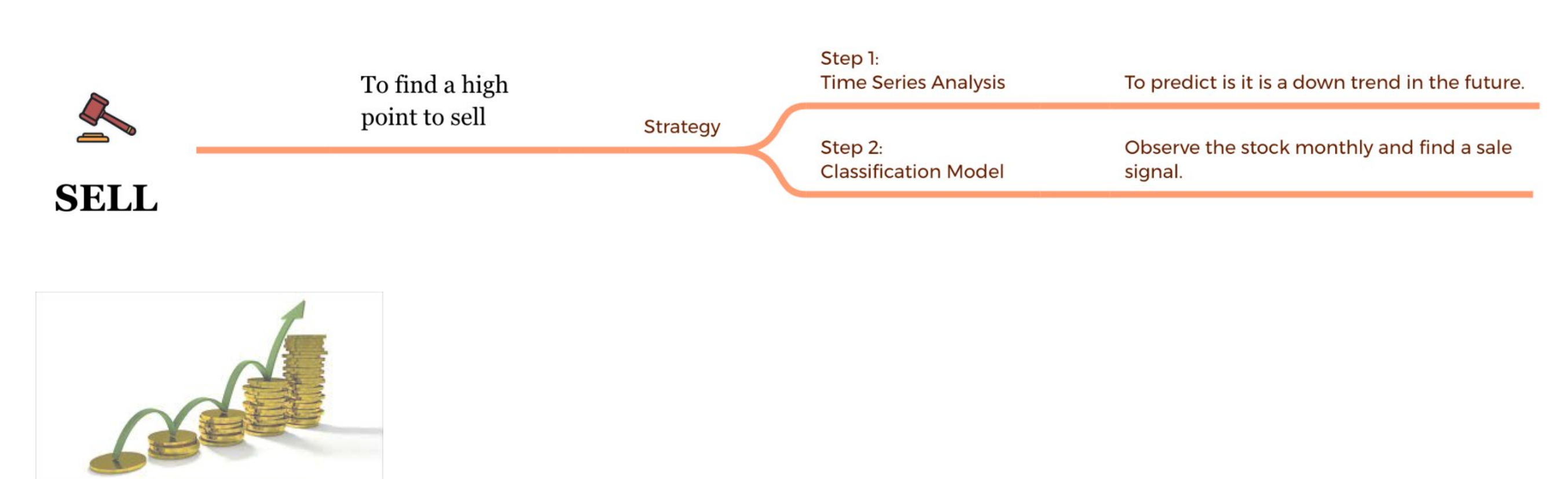
- Time series analysis provide a simple way for short-term forecasting, while the smoothing method offsets the ups and downs of fluctuations.
- The multi-class classification model puts more factors into consideration, and gives a long-term forecast on fund flow trend.
- Stil, we are able to **group funds based on stability, volatility and sensitivity**, and give suggestions to clients for **safe/risk-taken choices, or to diversify portfolio**.

Use-Case Scenario:

- Find a low point to buy



- Find a high point to sell



Contact Information:

Capstone Member: Kaiyu Wang, Simeng Li, Xiangshan Mu, Yihan Jia, Yinghao Wang

Faculty Advisor: Jin Shawn

Address: Questrom Business School of Boston University, Boston MA 02215