

# *Investment Strategy of Jay Street*

*— A machine learning based approach*

*By Yi Gao, Yanni Lu, Sirui Wang, Chen Xi, Hongyu Yao*

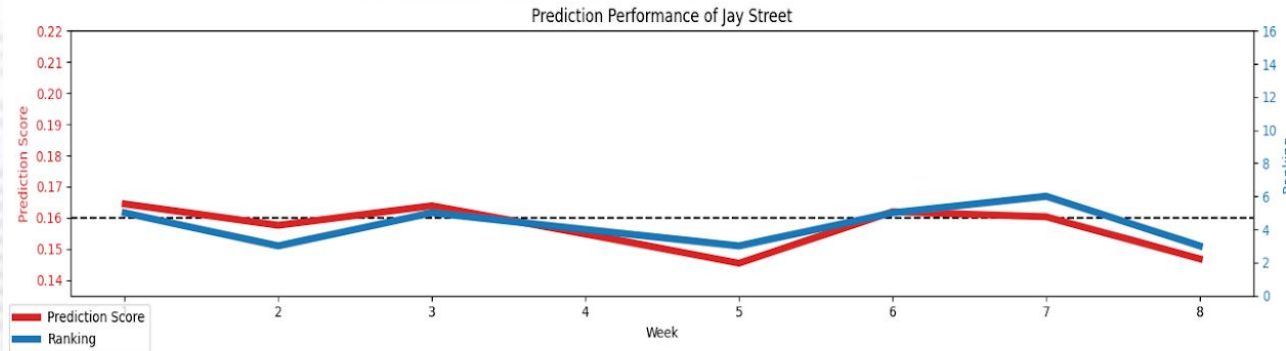
# Summary of Approach & Result



Forecast

## ◆ XGBClassifier

The classification probability across Rank1 to Rank5 serves as forecast value



## ◆ Robust and Excellent Result

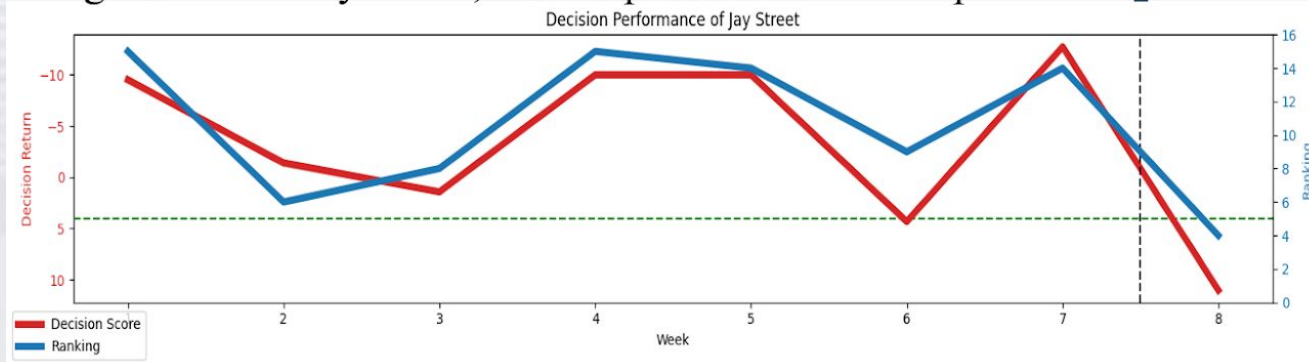
- 1) Forecast score around **0.156**
- 2) Ranked around **2<sup>nd</sup> to 5<sup>th</sup>**



Decision

## ◆ XGBRegressor (Previous Approach)

Regress on weekly return, use the predicted return as position



## ◆ Communication with Professor (New approach)

- 1) Stop point-wise forecasting
- 2) Eliminate position in Crypto
- 3) Position generated from probability of each category directly

$$\text{Position} = 0.2 * \text{Prob}(\text{Rank}_5) + 0.1 * \text{Prob}(\text{Rank}_4) - 0.1 * \text{Prob}(\text{Rank}_2) - 0.2 * \text{Prob}(\text{Rank}_1)$$

## *Investment Strategy of Jay Street*

*— — A machine learning-based approach*



**Feature  
Engineering**



**Model  
Selection**



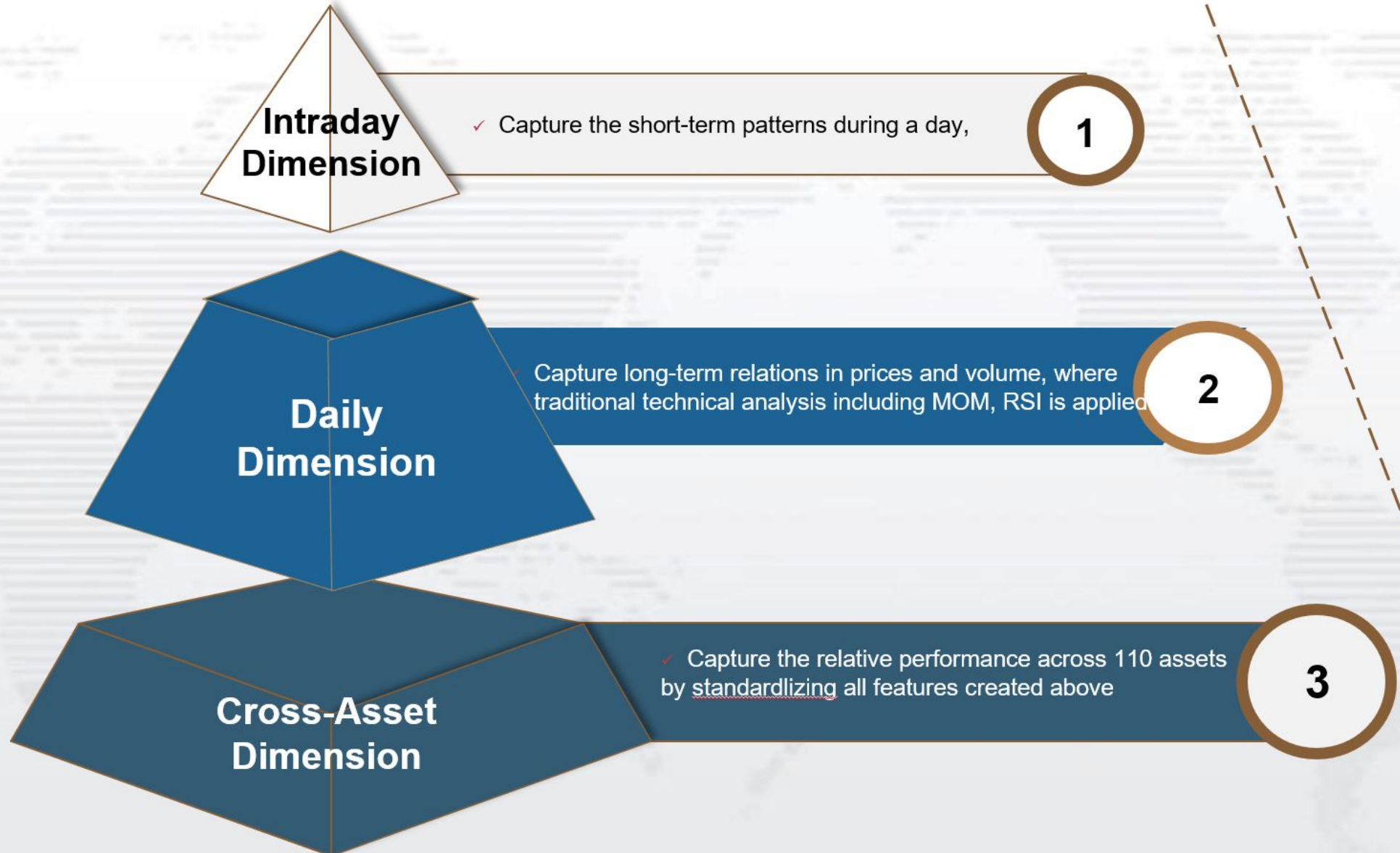
**Backtesting  
Results**



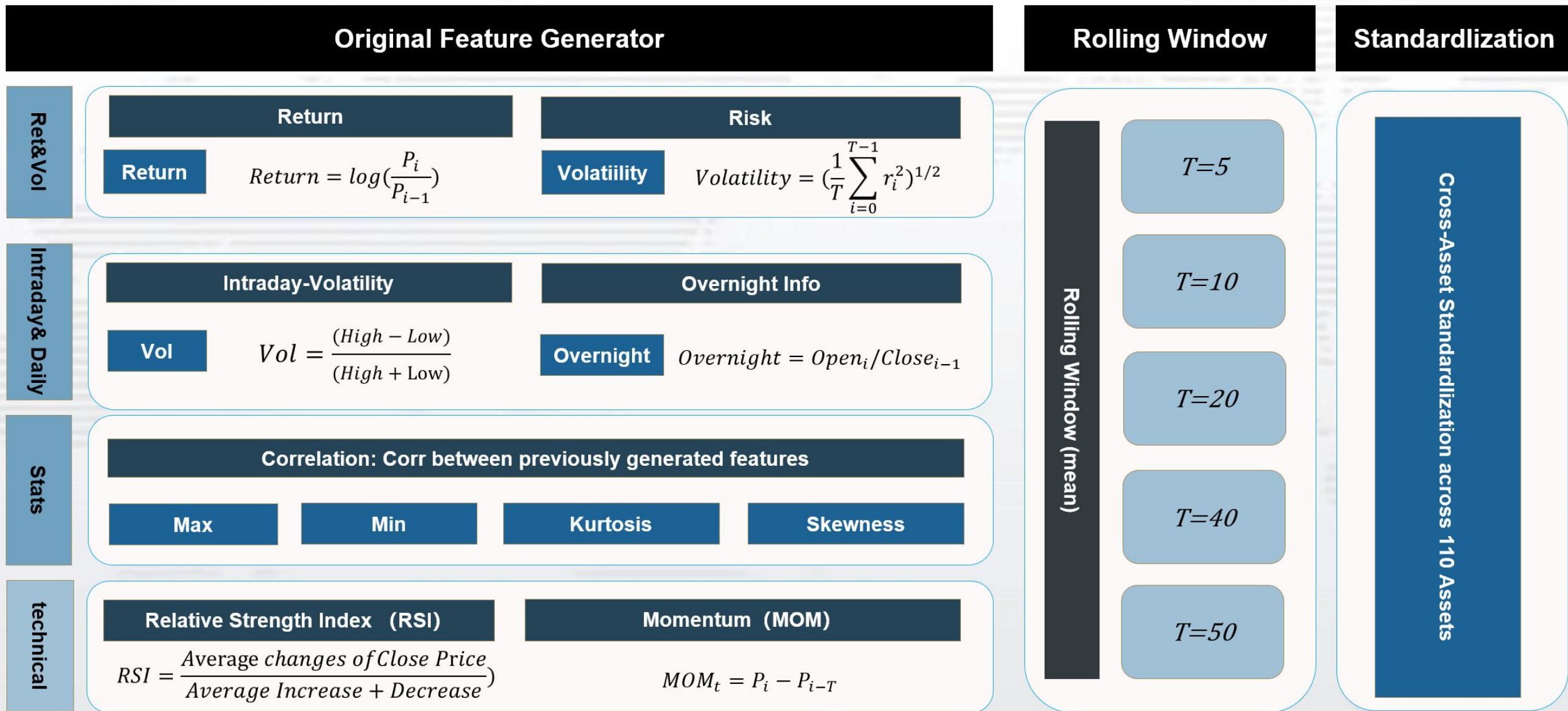
**Conclusion**



# Feature Engineering



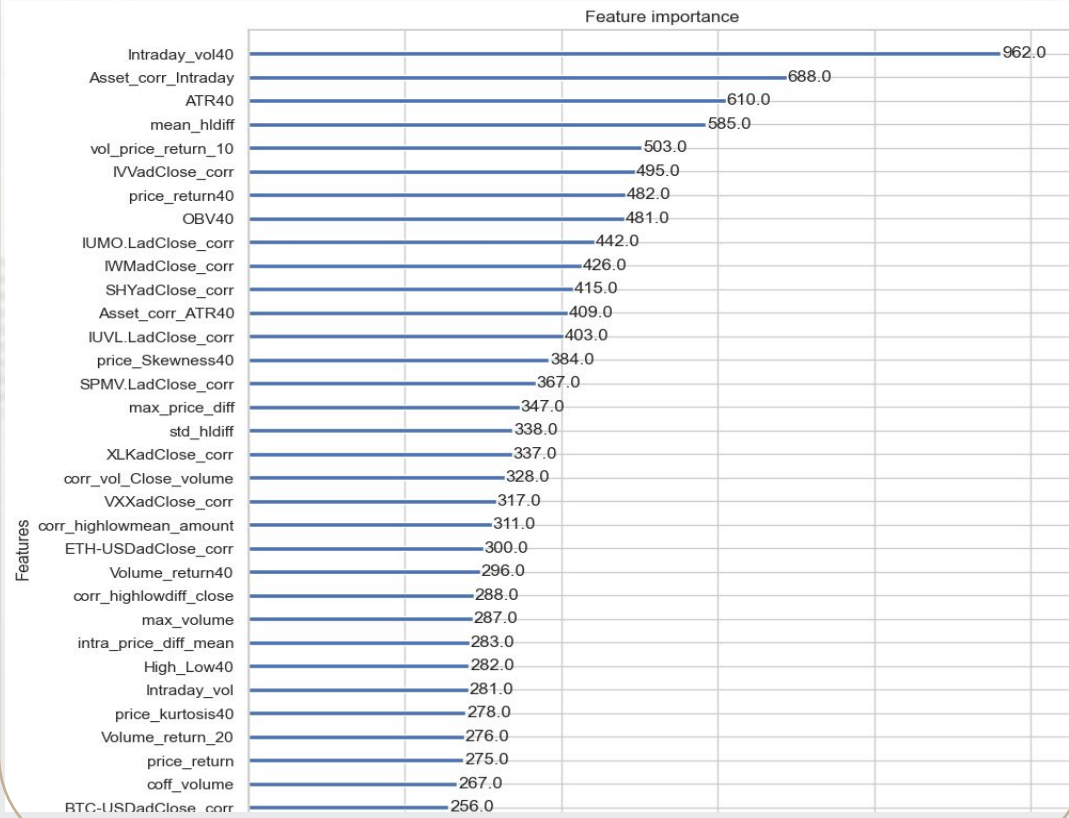
# Feature Engineering



# Feature Selection

## Feature Importance

- ◆ Feature Importance helps us retain the most invaluable features while eliminating redundant features, keeping our model at highest efficiency



## Correlation

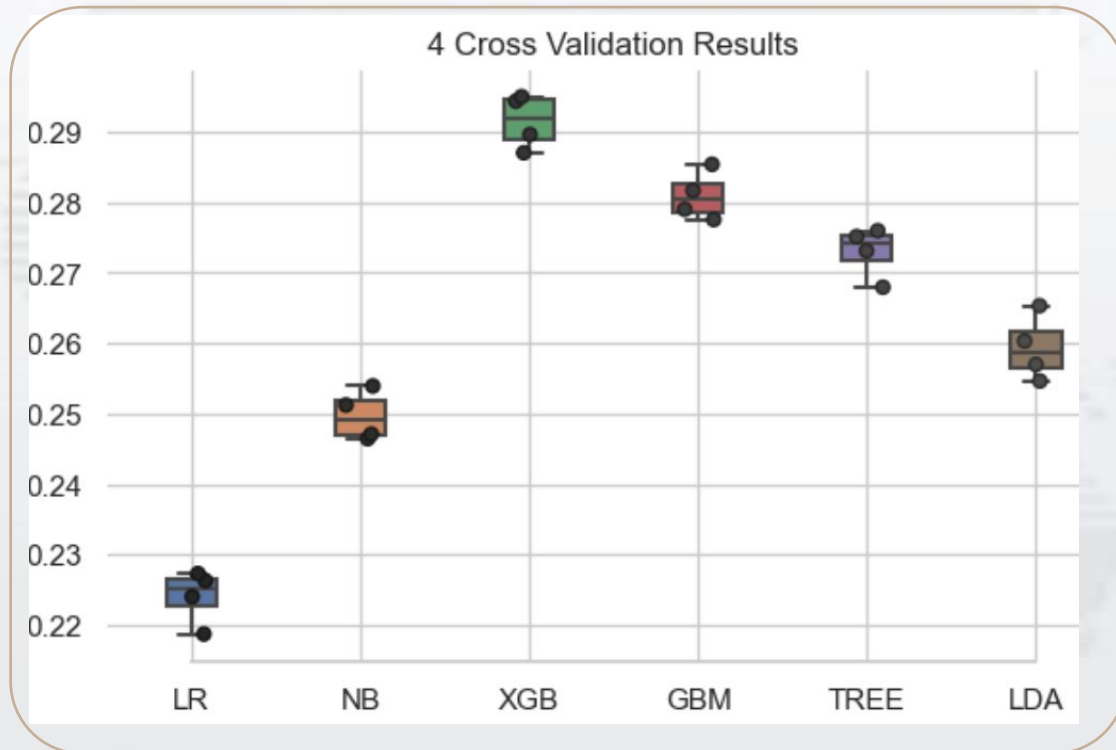
- ◆ We calculated the correlation between the target (i.e. Price Return of each asset we want to predict) and each feature
- ◆ We keep the features that share significant correlations with the target we want to predict, and together with results from feature importance, we eliminate redundant features, therefore making our model as efficient as possible



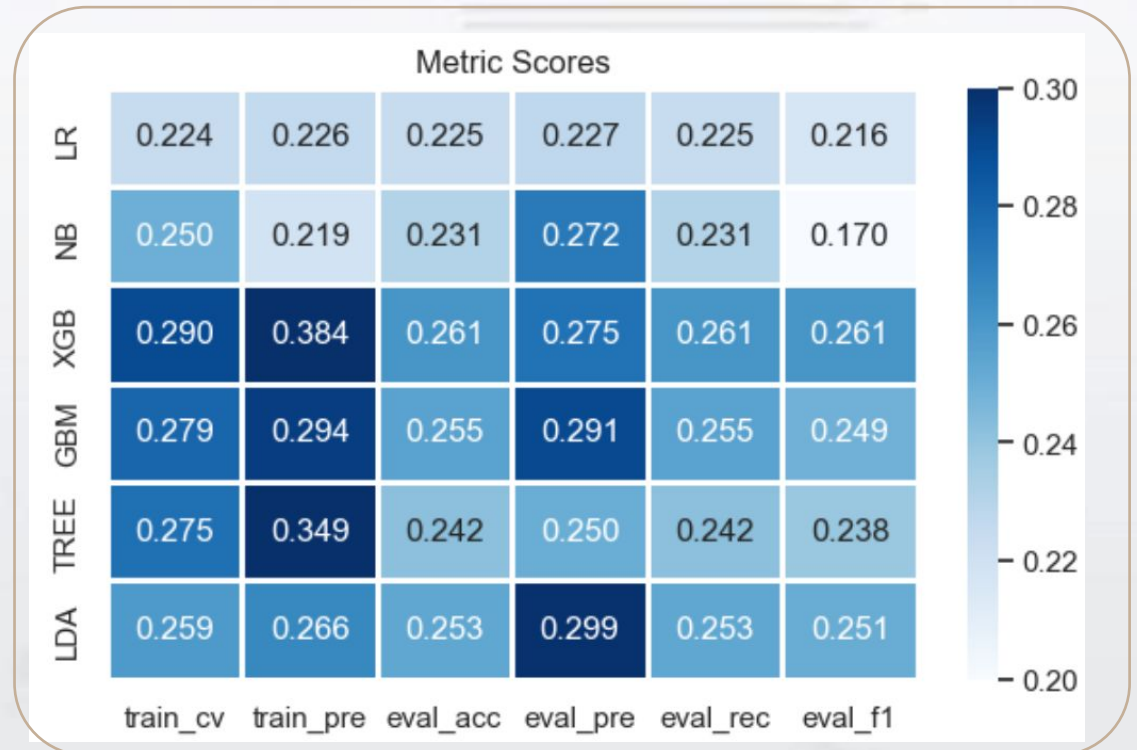
# Forecast: Model Selection

- ◆ Xgboost shows better performance regarding three metrics:
  - 1) Cross Validation Result
  - 2) Confusion Matrix Result (including Precision, Recall and F-1 Score)
  - 3) Backtesting of Prediction Score (which we values the most)

## 1) Cross Validation Results



## 2) Confusion Matrix Result



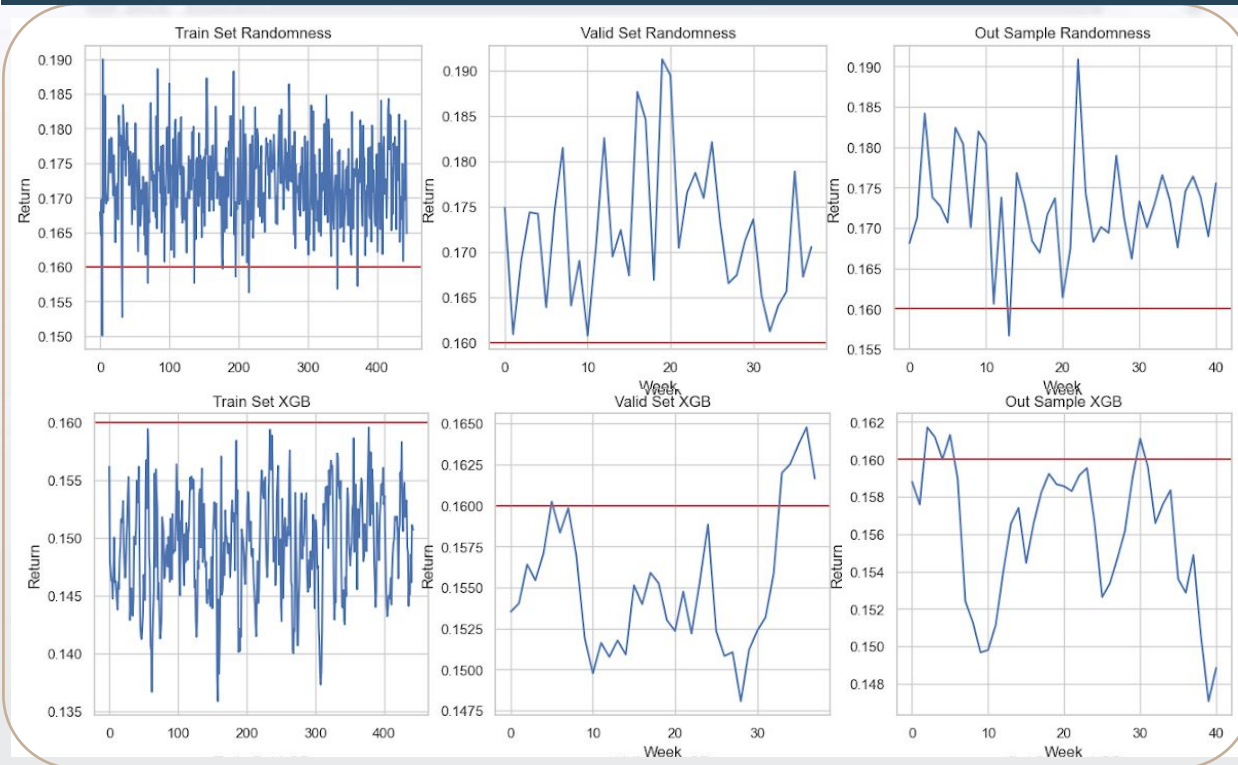


# Forecast: Model Selection & Backtesting

◆ Xgboost shows better performance regarding three metrics:

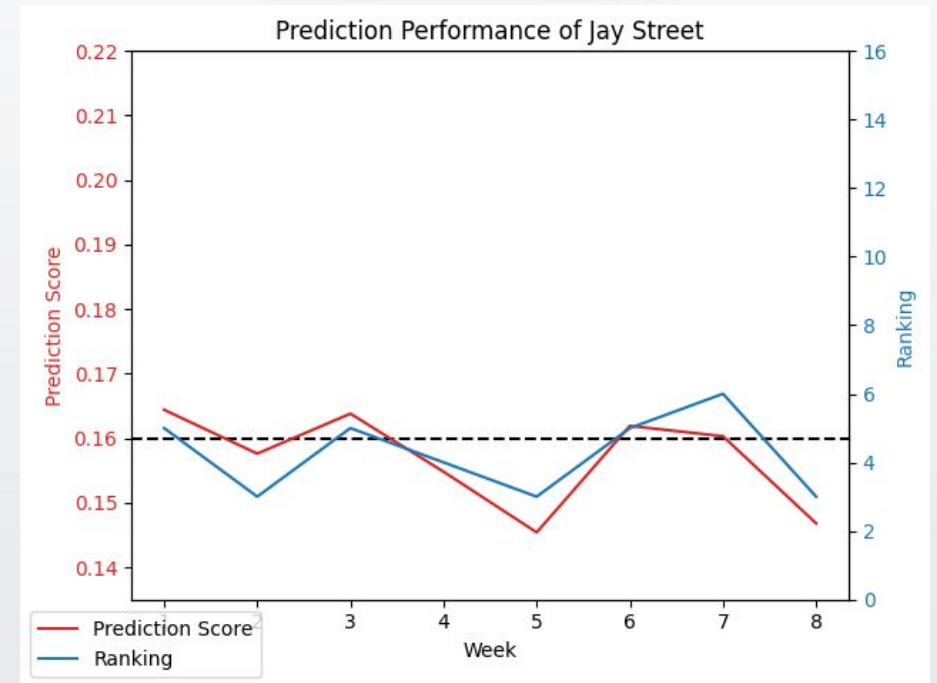
- 1) Cross Validation
- 2) Precision, Recall and F-1 Score
- 3) Backtesting of Prediction Score (**Weighs more**)

## 3) Backtesting Result



## Real World Performance

- ◆ Our forecasting score is around **0.156**, which **aligns with** our backtesting result
- ◆ The **scientific approach** (including feature engineering, model selection, backtesting, etc.) make it possible for us to secure a robust high ranking around **2<sup>nd</sup> to 5<sup>th</sup>**



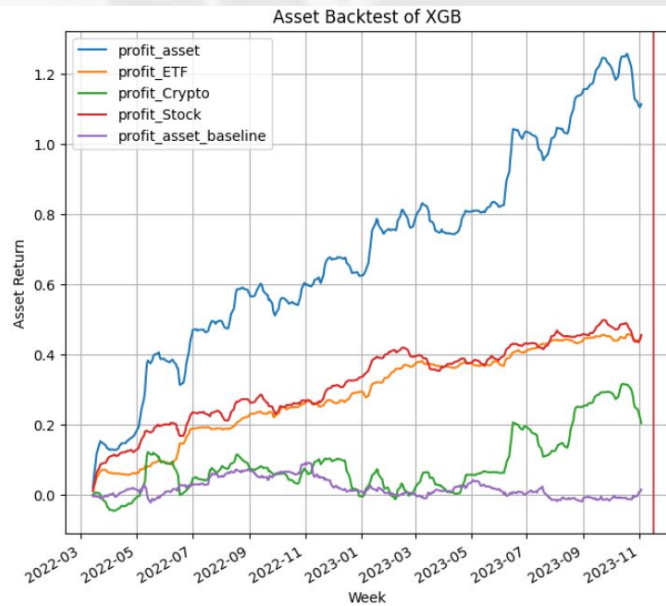


# Decision: Backtesting Results

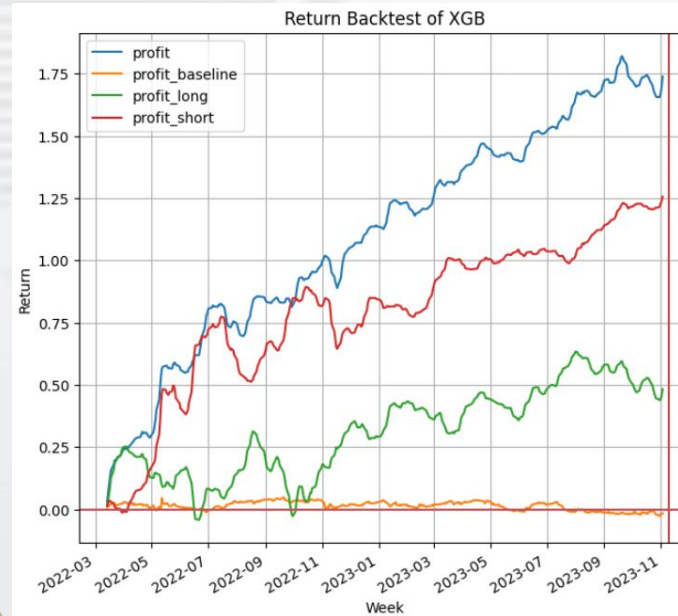
## ◆ Investment Insights

- 1) Crypto Ban: i) Large vol harms *Sharpe*  
ii) Plunge *Return* to negative
- 2) Equal Long/Short Position: Do not take Market Trend

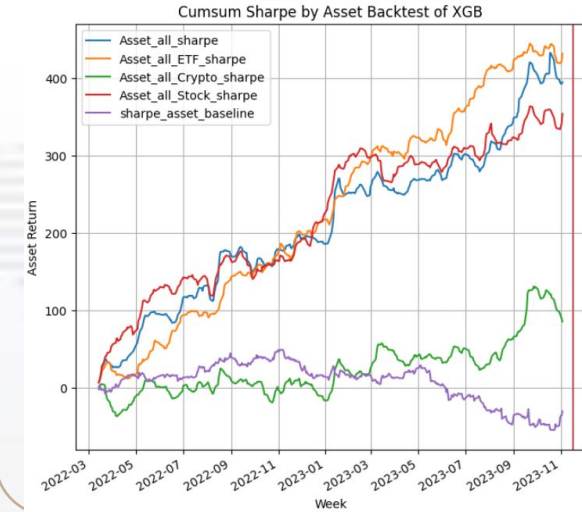
### Returns with Crypto



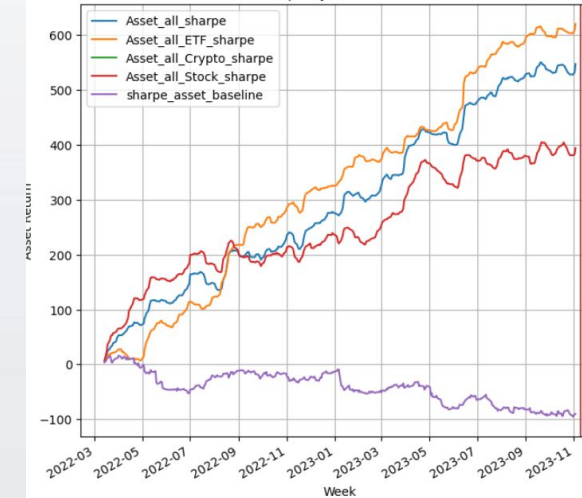
### Returns w/o Crypto



### Sharpe Ratio with Crypto



### Cumsum Sharpe by Asset Backtest of XGB



### Sharpe Ratio w/o Crypto

# Decision: Model Selection & Backtesting

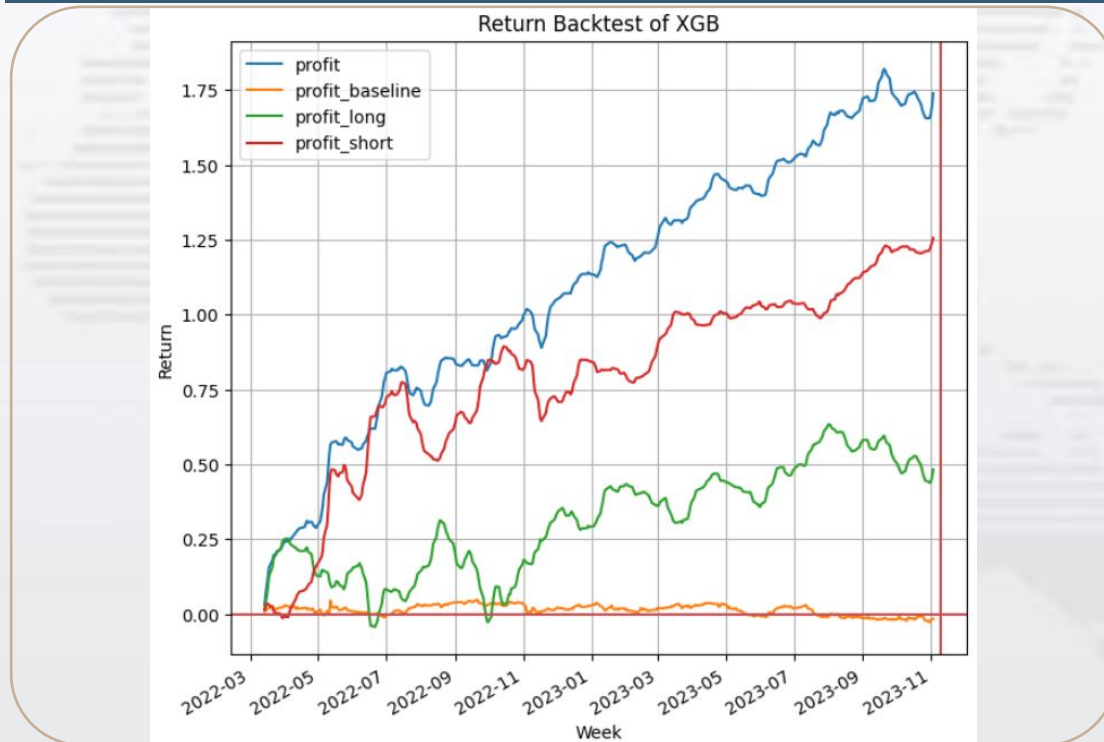
## ◆ Backtesting Perspectives

- 1) Dimension One (Metrics): Info Coeff, Return and Sharpe
- 2) Dimension Two (Position): Long/Short, ETF/Stock/Crypto

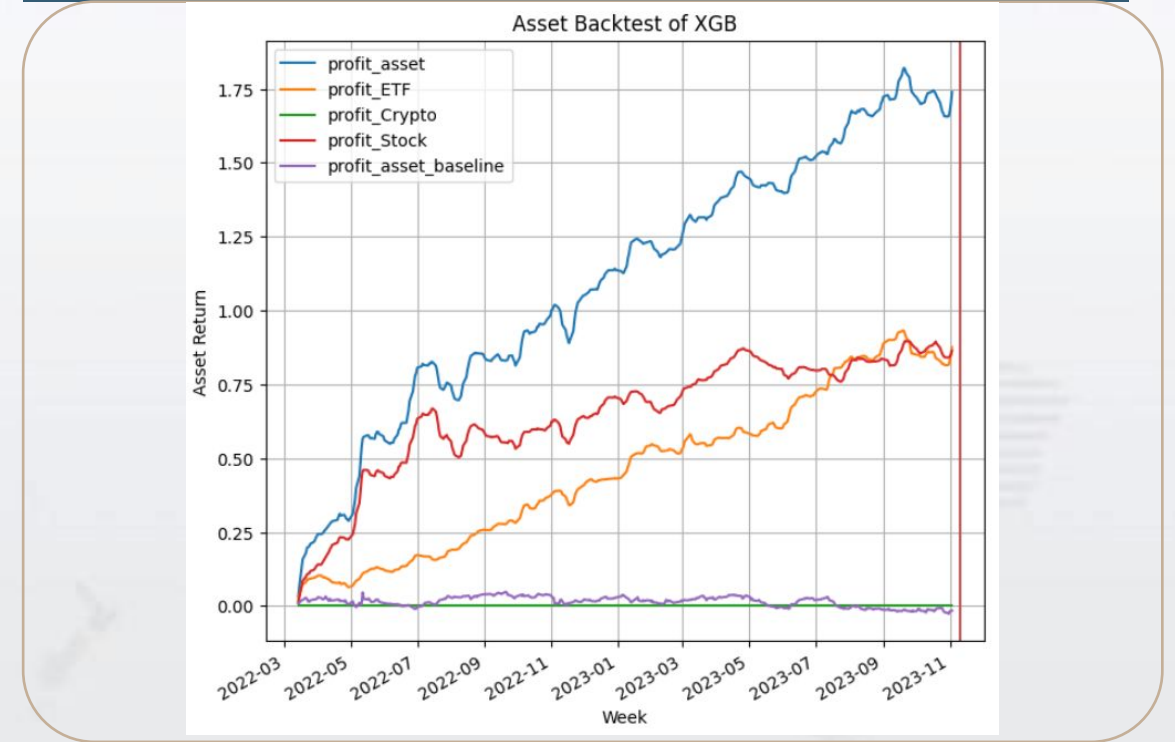
## ◆ Rolling Backtesting: From 2021-09-27 to Today

- 1) Training: 24 weeks
- 2) Test: 8 weeks
- 3) Out Sample: 8 weeks

### 1.1) Return – Long/Short



### 1.2) Return – 3 Assets

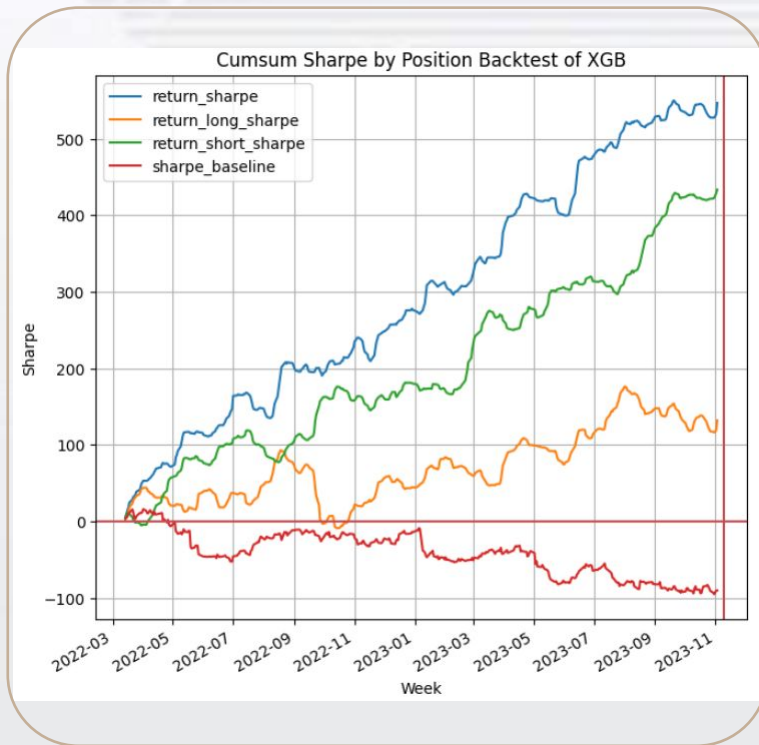


# Decision: Backtesting Results

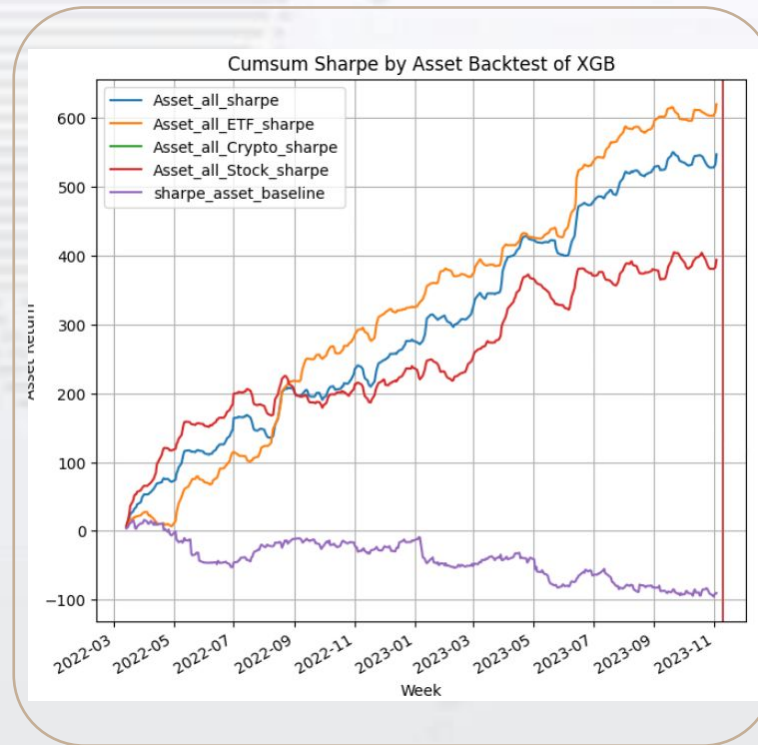
## ◆ Backtesting Perspectives

- 1) Dimension One (Metrics): Information Coefficient, Return and Sharpe
- 2) Dimension Two (Position): Long/Short, ETF/Stock/Crypto

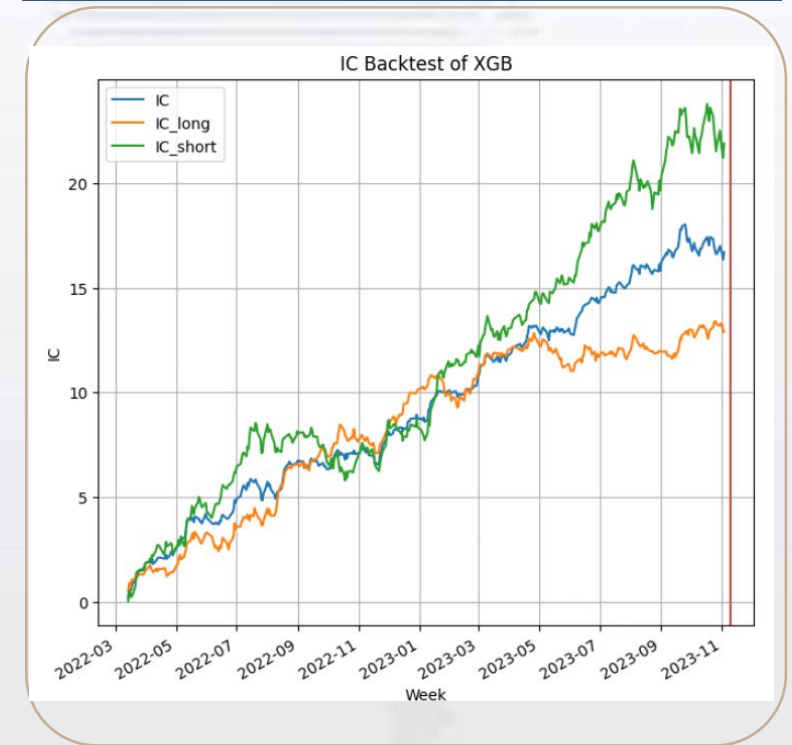
### 2.1) Sharpe – Long/Short



### 2.2) Sharpe – 3 Assets

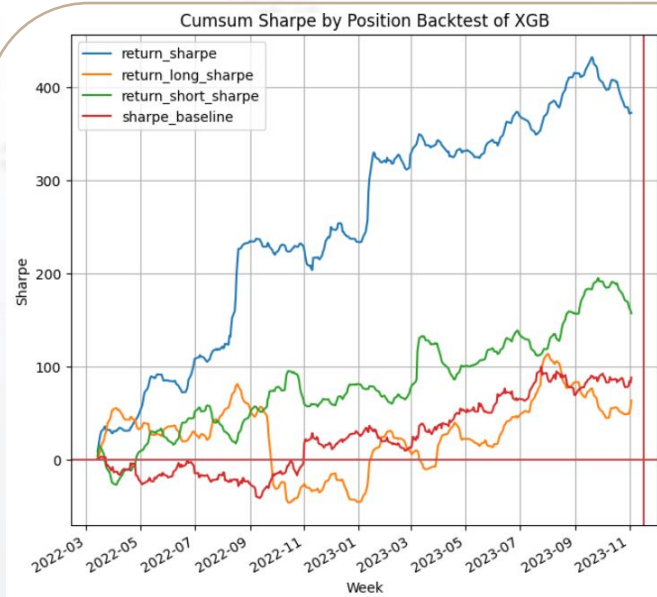


### 3) IC



# Decision: Backtesting Results

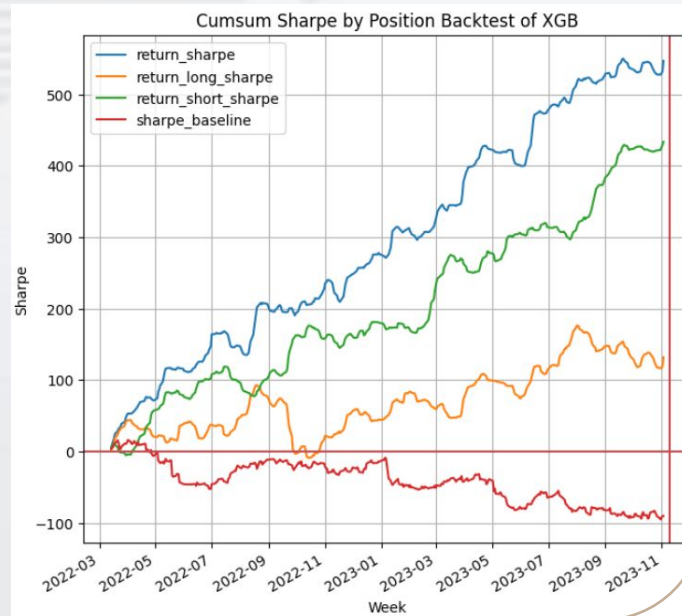
## Comparison of Two different Approach to Decision



- ◆ The Point-wise Approach is susceptible to market volatility, which exhibits significant overfitting on the Out Sample

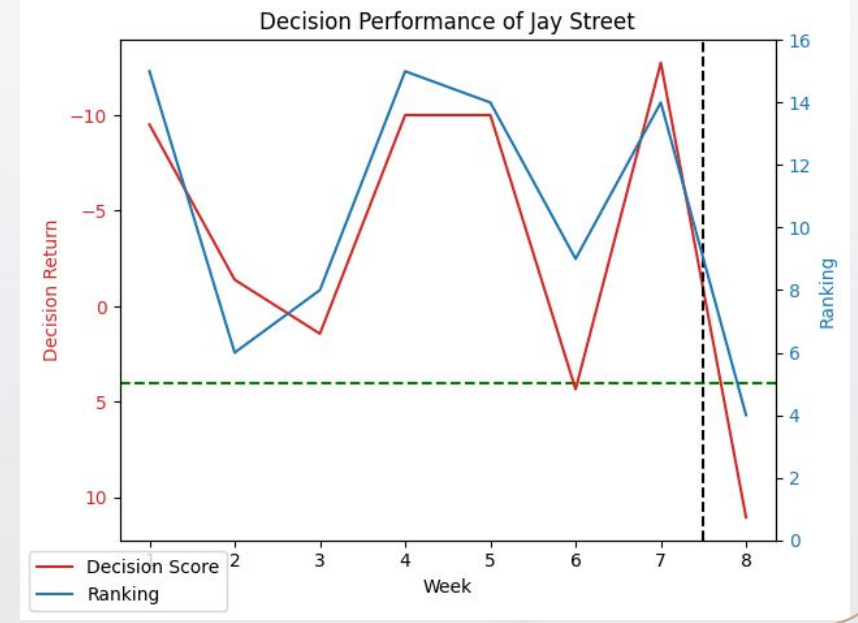
- ◆ The Probability Conversion Approach provides relative stability by stratifying, which strengthens the robustness of investment decisions as a whole

$$\text{Position} = 0.2 * \text{Prob}(\text{Rank}_5) + 0.1 * \text{Prob}(\text{Rank}_4) - 0.1 * \text{Prob}(\text{Rank}_2) - 0.2 * \text{Prob}(\text{Rank}_1)$$



## Real World Performance

- ◆ The **Point-wise** Approach shows significant overfitting and volatility due to unreliable point-wise *Weekly Return* prediction for each asset in a drastically volatile financial world
- ◆ The **Probability Conversion** Approach renders robustness and stability in our investment decision, which enables us to secure a high ranking around **2<sup>nd</sup> to 5<sup>th</sup>** in the latest several submissions





# Conclusion

- ◆ Excellent Performance in *Forecasting* proves the efficacy of our feature engineering and backtesting workflow, which is scientific, reliable and repeatable
- ◆ The limited performance in *Decision* informed us of the volatility of financial markets, and the pursuit of relative performance should prevail over the seeking for absolute accuracy
- ◆ **Always wise to communicate more with the professor!!**

*Thanks for your Attention*

***Jay Street***

*Yi Gao, Yanni Lu, Sirui Wang, Chen Xi, Hongyu Yao*