การนำเสนองานวิจัย “Empirical Asset Pricing using xAI”

โดย ผศ. ดร.สัมพันธ์ เนตยานันท์ มหาวิทยาลัยนเรศวร

ดร.ศิริยศ จุฑานนท์ ตลาดหลักทรัพย์แห่งประเทศไทย

20 กรกฎาคม 2566

Asset Pricing Model: Key Concepts and Assumptions

1.1 Efficient Market Hypothesis (EMH)

- Reflect all available information
- Investors are rational
- Not possible to consistently outperform the market

1.2 Risk and Expected Return

- Investors typically demand a higher return for taking on greater risks

1.3 Capital Asset Pricing Model (CAPM)

- The only risk is systematic risk measured by beta
- The rest is the alpha which refers to fund manager performance

1.4 Factor Model

- Considers risk factors such as size, value, momentum, profitability, etc.

1.5 Behavioral Finance

- Investors may not always behave rationally with their biases and emotions
- Incorporate behavioral factors such as PEAD, VIX index, and survey may improve model

Artificial intelligence (AI) can play a significant role in asset pricing models

Data analysis and feature selection:

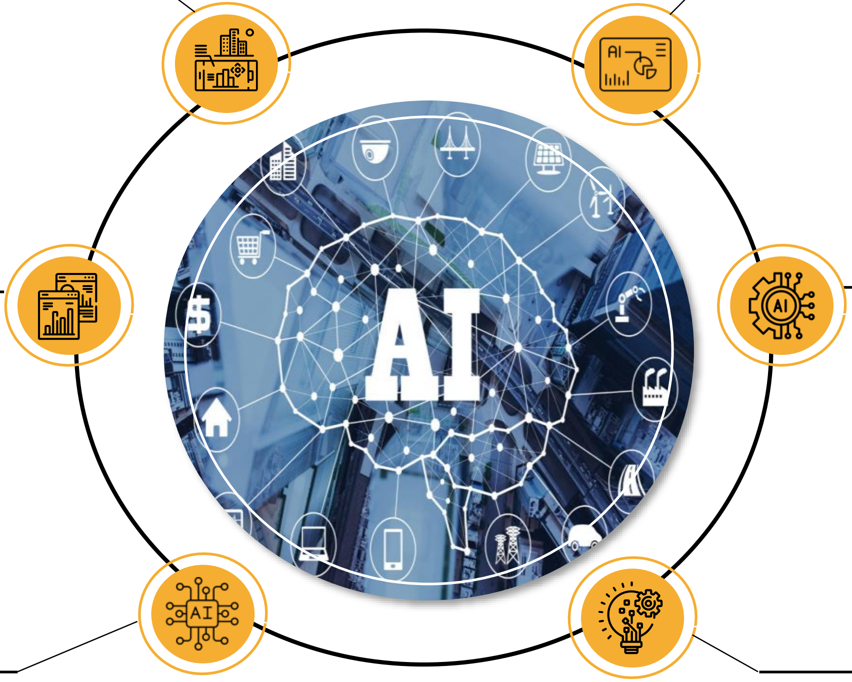
- Analyze large volumes of financial data to identify relevant features that impact asset prices
- Able to uncover non-linear relationships

Pattern recognition and predictive modeling:

- Identify complex patterns and trends in historical asset price data, enabling the development of predictive models

Sentiment analysis:

- Analyze news articles, social media posts, and other textual data to gauge market sentiment



Risk assessment:

- Evaluating and quantifying various risk factors that impact asset prices

Portfolio optimization:

- Optimize portfolio construction and asset allocation by considering multiple factors simultaneously

High-frequency trading:

- Enable the execution of trades at high speeds based on predefined rules and market signals
- Capitalizing on short-term pricing anomalies and improving overall trading performance

Some of the key challenges include:

01 Data quality and availability

- Heavily rely on high-quality and reliable data for accurate predictions. Historical data may contain errors or missing values.
- Data availability can be a constraint for emerging or illiquid markets.

02 High dimensionality and data noise

- Complexity form various indicators such as price, volume and financial ratios. Moreover, some of them are noise in financial market, which can lead to model accuracies.

03 Overfitting and model complexity

- Occurs when a model performs well on historical data but fails to generalize to new, unseen data

04 Changing market dynamics

- Such as shifts in economic factors, policy changes, or unexpected events
- AI models trained on historical data may struggle to adapt to new unpredictable market conditions or unforeseen events

05 Interpretability and transparency

- Often considered "black boxes" because they lack interpretability
- Understanding how the model arrives at its predictions or identifying the key drivers of asset prices can be challenging

Building trust in AI involves several key considerations: Introducing xAI

01

- Unlike traditional computer programs that follow a set of predefined rules to produce an output
- Machine learning algorithms are designed to learn from data and find patterns on their own
- The decision-making process of a machine learning model is not always transparent

02

- Researchers are working to develop methods to make machine learning algorithms more transparent and interpretable
- Clearly communicate how the AI system arrived at a particular recommendation or decision, enabling users to comprehend and verify the reasoning

03

- Explainable AI (xAI): a set of techniques and methods used in artificial intelligence and machine learning that enable users to understand how a model works and why it arrived at a particular decision or prediction
- Users can gain insights into how an AI system arrived at a particular conclusion, and can potentially identify errors or biases

04

xAI techniques that can be used including:

- Feature Importance Analysis
- Decision Trees:
- Partial Dependence Plots
- Rule-Based Systems:

Literature Review xAI

Main Objective	Research Paper	Predicted Variable	Machine Learning	xAI
Economics forecasting	"An interpretable machine learning work flow with an application to economic forecasting" by Buckmann, Joseph, and Robertson (2021)	• Forecast US unemployment one year ahead in a monthly dataset	• Random Forests • Neural Networks • Compare to conventional models.	• Using SHapley Additive exPlanation (SHAP)
	"Interpretable deep learning LSTM model for intelligent economic decision-making" by Park and Yang (2022)	• Predict economic growth rates and crises for major G20 countries	• Deep learning model based on the Long Short-term Memory (LSTM) network	• Using SHapley Additive exPlanation (SHAP)
Stock price prediction	"Explainable stock prices prediction from financial news articles using sentiment analysis" by Gite et al. (2021)	• Predict next-day stock price in the National Stock Exchange (NSE) with Indian finance news headlines	• Suggests a technique involving LSTM	• Using Local interpretable model-agnostic explanations (LIME)
	"Explainable AI for Financial Forecasting" Carta, Podda, Recupero, and Stanciu (2022)	• Predict the next- day returns for stocks in S&P500, CAC, FTSE	• Mean Decrease Impurity (MDI) • Random forests	• Using Local interpretable model-agnostic explanations (LIME)

Literature Review xAI

Main Objective	Research Paper	Predicted Variable	Machine Learning	xAI
Trading strategy	"The best way to select features? Comparing MDA, LIME and SHAP" Man and Chan (2022)	• Predict whether each trade of the strategy will be profitable	• Random forest	• Comparing MDA, LIME and SHAP
Forecast stock market crisis	"Explainable AI (XAI) models applied to planning in financial markets" by Benhamou Ohana, Saltiel, and Guez (2021)	• Identification of the most important variables planning stock market crises during March 2020 equity meltdown	• Gradient boosting decision tree (GBDT)	• Using SHapley Additive exPlanation (SHAP)
Asset pricing model	"Machine Learning Algorithms for Financial Asset Price Forecasting" by Ndikum (2020)	• Explores financial asset price forecasting on U.S equities data	• High performance computing (HPC) infrastructures vs. the traditional CAPM	• None
	"Empirical Asset Pricing via Machine Learning" by Gu, Kelly, and Xiu (2020)	• Comparative analysis of machine learning methods for measuring asset risk premia • Forecast returns using various predictive features at the firm, industry, and macro levels	• Artificial neural networks (ANN) • Boosted regression trees • Random forests	• None

- **Research Question**
  - **How each factor explains portfolio returns in a machine learning setting by using xAI**
- **Contribution**
  - **One of the first to employ the xAI to an expansive list of financial anomalies to illustrate factor importance**

Label and Features

$$E\left(LS_{i,t}\right) = f_{ANN}(\boldsymbol{F_t})$$

- The left-hand side indicates the zero-cost long-short portfolio
- $\boldsymbol{F_t}$ are constructed from the three-by-five portfolios conditioned on the size
- The model structure is similar to Gu, Kelly, and Xiu (2020)
- Right-hand side variables act like the macroeconomic variables in Gu, Kelly, and Xiu (2020)
- The model is also similar to Gu, Kelly, and Xiu (2021): Includes common factors to explain individual and portfolio returns
- Includes the excess market returns
- There are 188 factors/features in total

Data

- We use the data from 1991 until 2021 to consider all anomalies for factor constructions
- Global-q.org
  - 41 momentum
  - 32 value-versus-growth
  - 29 investment
  - 46 profitability
  - 30 intangible
  - 10 friction anomalies
- 69,936 rows of data

Momentum

Zip folders that contain all 41 momentum anomalies for a given frequency

1-way sorts:	Daily	Weekly (calendar)	Weekly (Wednesday-to-Wednesday)	Monthly
2-way sorts:	Daily	Weekly (calendar)	Weekly (Wednesday-to-Wednesday)	Monthly

Explanation of CSV filenames for individual momentum anomalies

1. Abr1 ("abr_1"), cumulative abnormal returns around earnings announcement dates, 1-month holding period;
2. Abr6 ("abr_6"), cumulative abnormal returns around earnings announcement dates, 6-month holding period;
3. Abr12 ("abr_12"), cumulative abnormal returns around earnings announcement dates, 12-month holding period;
4. Cim1 ("cim_1"), customer industries momentum, 1-month holding period;
5. Cim6 ("cim_6"), customer industries momentum, 6-month holding period;
6. Cim12 ("cim_12"), customer industries momentum, 12-month holding period;
7. Cm1 ("cm_1"), customer momentum, 1-month holding period;
8. Cm12 ("cm_12"), customer momentum, 12-month holding period;
9. dEf1 ("def_1"), changes in analyst earnings forecasts, 1-month holding period;
10. dEf6 ("def_6"), changes in analyst earnings forecasts, 6-month holding period;
11. dEf12 ("def_12"), changes in analyst earnings forecasts, 12-month holding period;
12. Ile1 ("ile_1"), industry lead-lag effect in earnings surprises, 1-month holding period;
13. Ilr1 ("ilr_1"), industry lead-lag effect in prior returns, 1-month holding period;
14. Ilr6 ("ilr_6"), industry lead-lag effect in prior returns, 6-month holding period;
15. Ilr12 ("ilr_12"), industry lead-lag effect in prior returns, 12-month holding period;
16. Im1 ("im_1"), industry momentum, 1-month holding period;
17. Im6 ("im_6"), industry momentum, 6-month holding period;

Source: https://global-q.org/testingportfolios.html

Data

Table 1. Statistics of Factors

Anomaly	m	σ	SR	t-stat
abr_1	0.648	1.893	0.342	6.587
abr_6	0.333	1.339	0.249	4.789
abr_12	0.248	0.997	0.249	4.801
aci	0.151	1.903	0.079	1.526
adm	0.192	4.391	0.044	0.842
almq_1	0.389	3.533	0.110	2.122
almq_6	0.472	3.273	0.144	2.780
almq_12	0.344	3.147	0.109	2.104
ato	0.536	2.907	0.184	3.549
atoq_1	0.803	2.594	0.309	5.961
atoq_6	0.783	2.595	0.302	5.814
atoq_12	0.693	2.610	0.266	5.117
beta_1	0.330	5.988	0.055	1.061
bm	0.291	3.833	0.076	1.462
bmj	0.321	4.335	0.074	1.425
bmq_12	0.268	4.274	0.063	1.208

Model

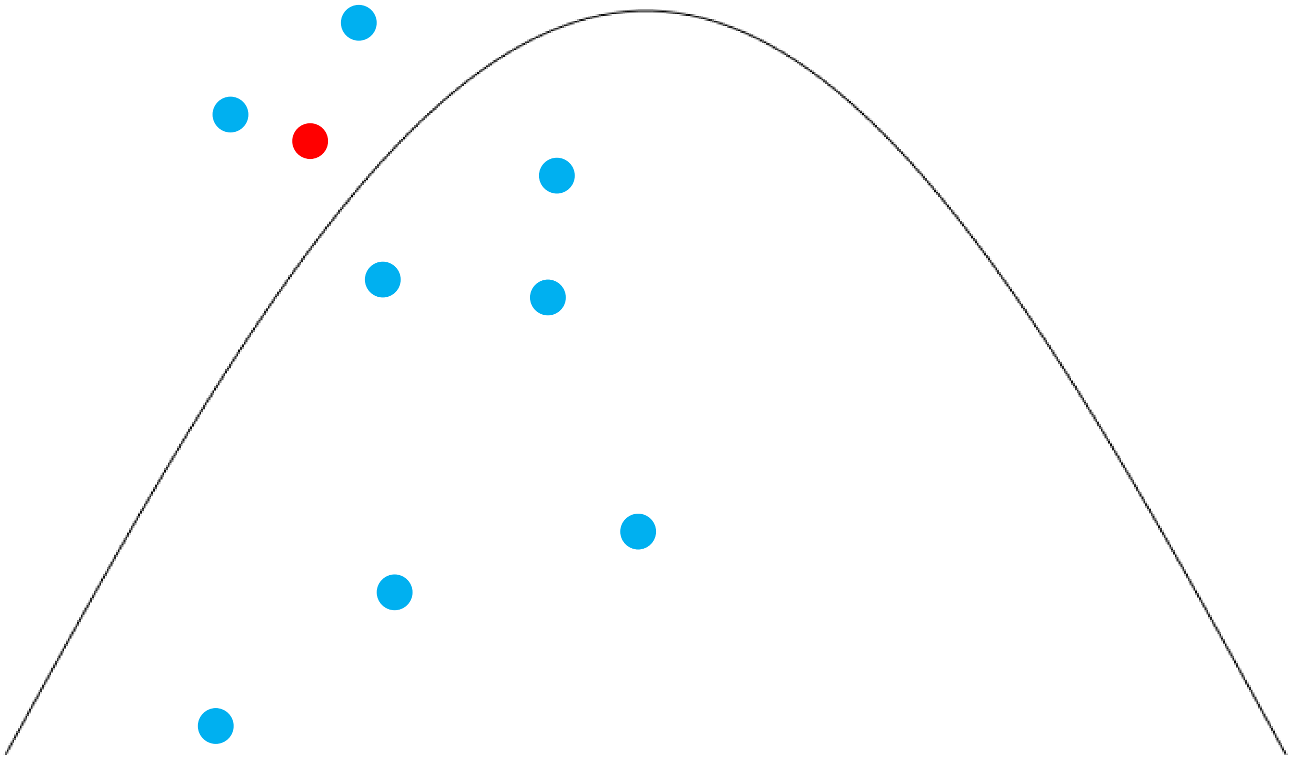
- Using artificial neural networks (ANN), similar to Gu, Kelly, and Xiu (2020)
- Geometric pyramid rule from Master (1993)
- 188 -> 53 -> 15 -> 4 -> 1
- Fully connected
- Sigmoid activation function
- 80% as a training sample 20% as a test sample
- 64 batches and 200 epochs
- Loss function: MSE (0.0023)
- Google Colab and TensorFlows using Python language

SHapley Additive exPlanations (SHAP)

- **Explainable AI (xAI) / Interpretable Machine Learning**
- **Lundberg and Lee (2017)**
- **Use to rank feature importance**
- **The idea is based on Shapley value from game theory**
- **Locally importance for each observation, can extend to global importance**
- **It can take up to nine hours for the calculation of SHAP**

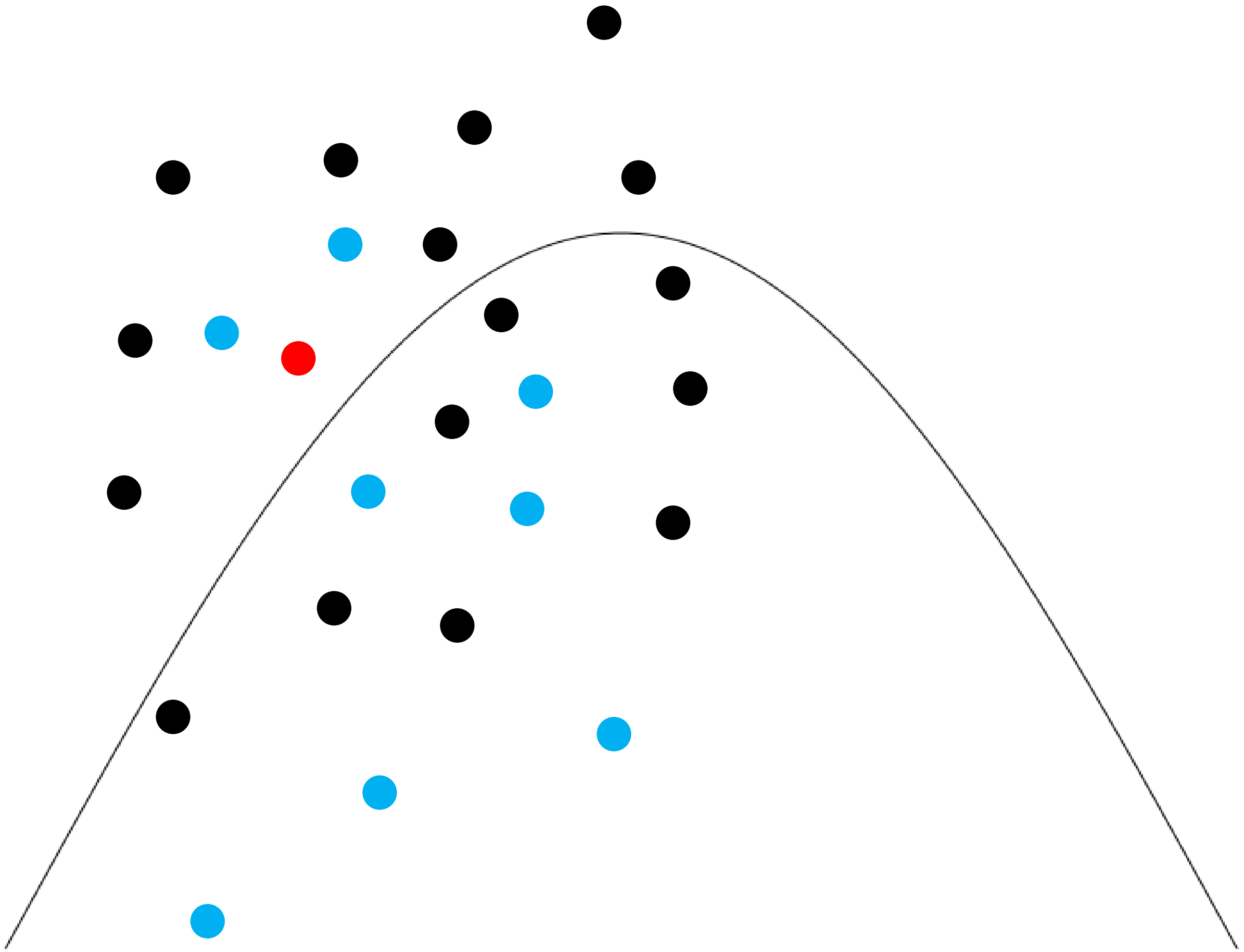
SHapley Additive exPlanations (SHAP)

- **Similar to Local Interpretable Model-agnostic Explanations (LIME) by Ribeiro, Singh, and Guestrin (2016)**



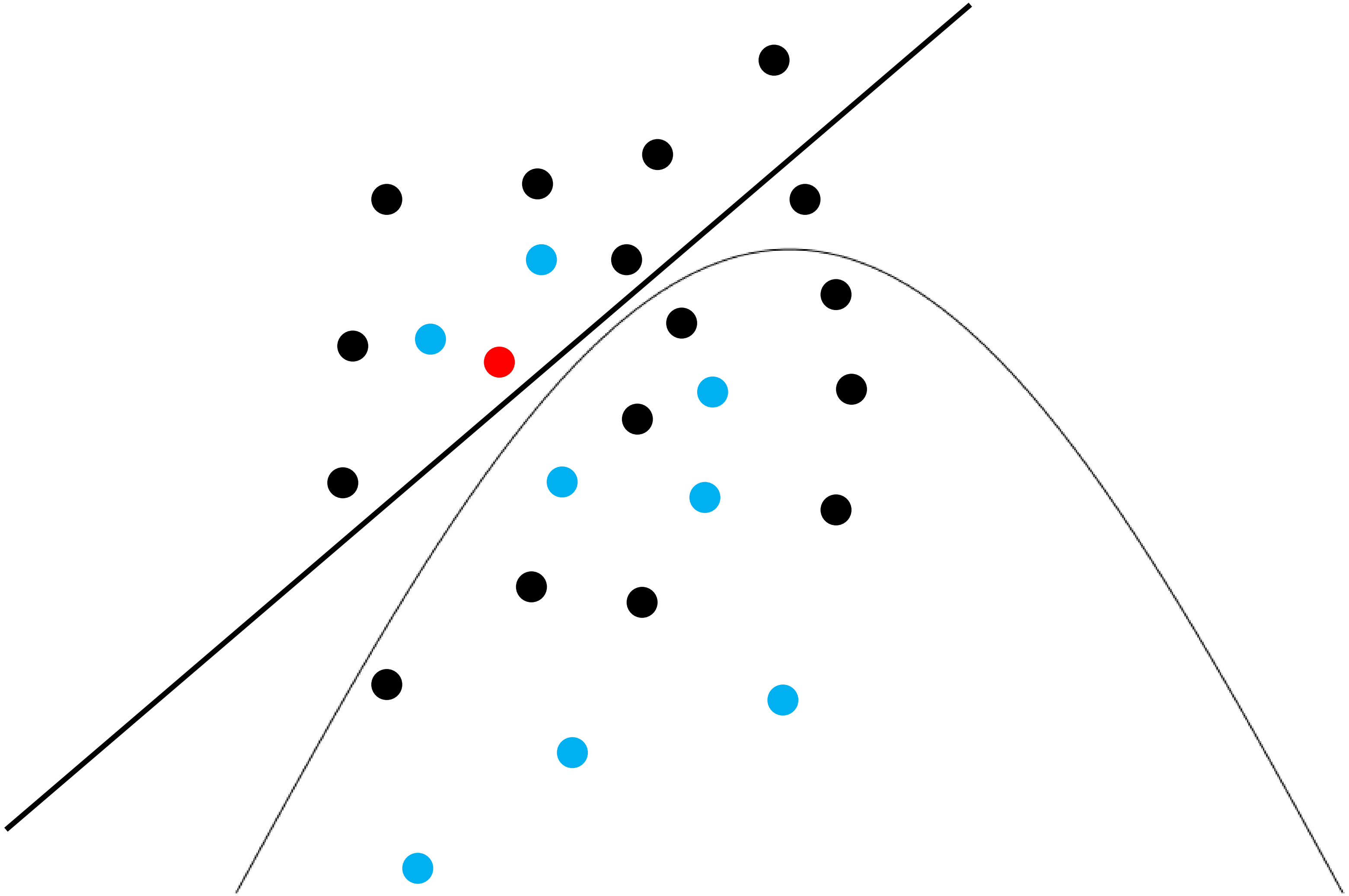
SHapley Additive exPlanations (SHAP)

- Similar to Local Interpretable Model-agnostic Explanations (LIME) by Ribeiro, Singh, and Guestrin (2016)



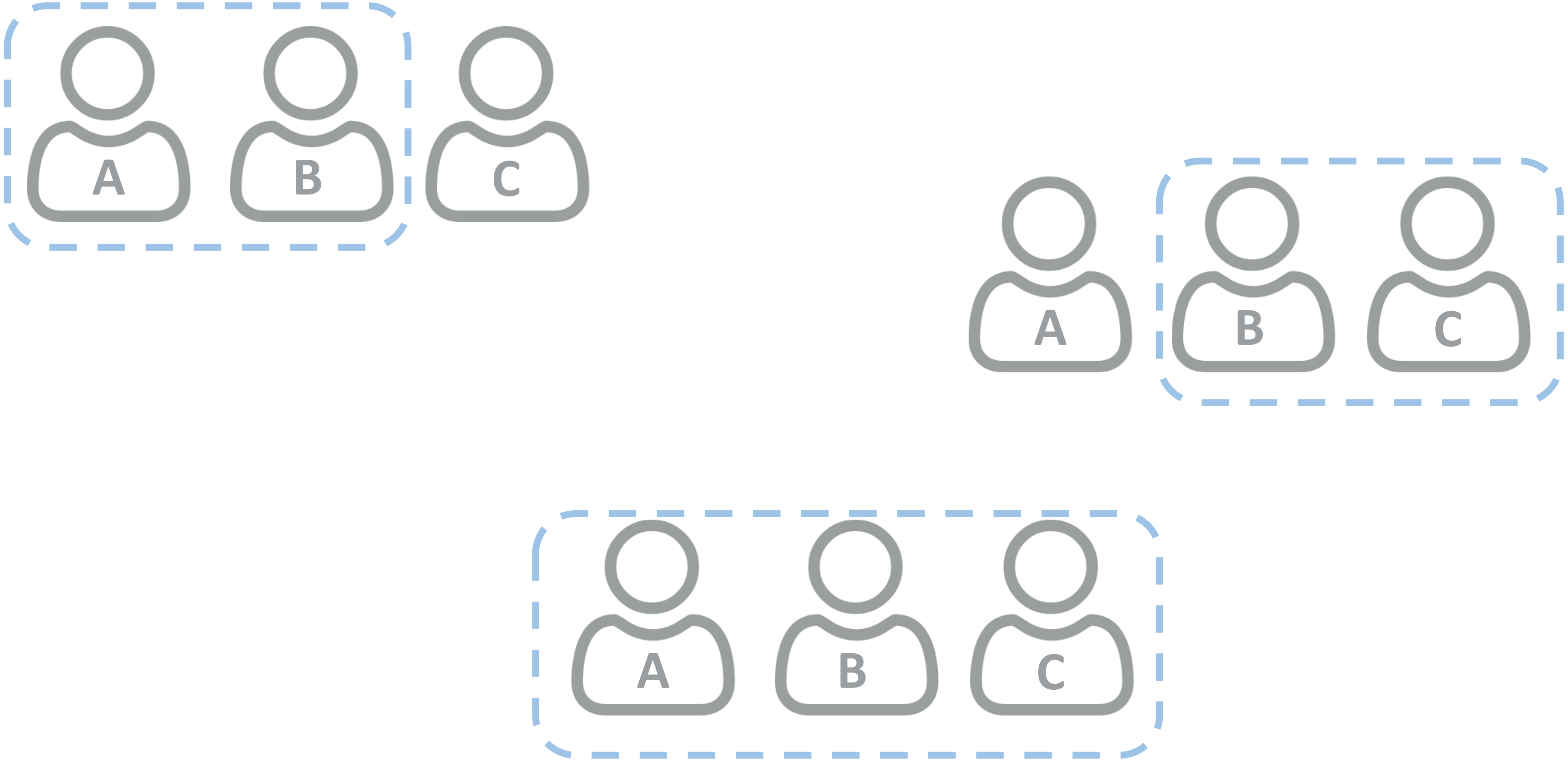
SHapley Additive exPlanations (SHAP)

- **Similar to Local Interpretable Model-agnostic Explanations (LIME) by Ribeiro, Singh, and Guestrin (2016)**



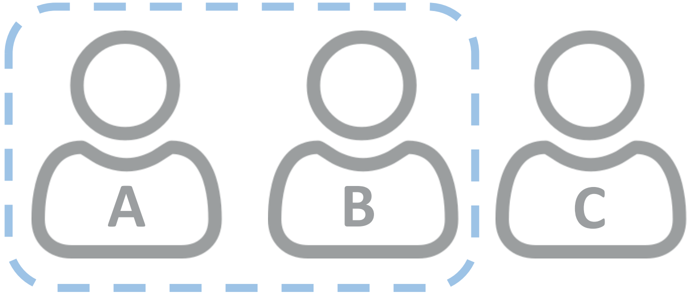
SHapley Additive exPlanations (SHAP)

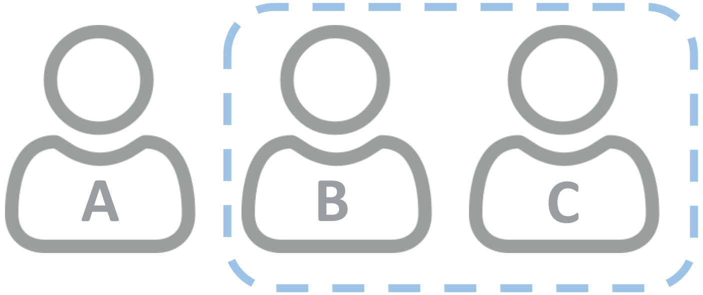
- “Shapley Value” in cooperative game theory



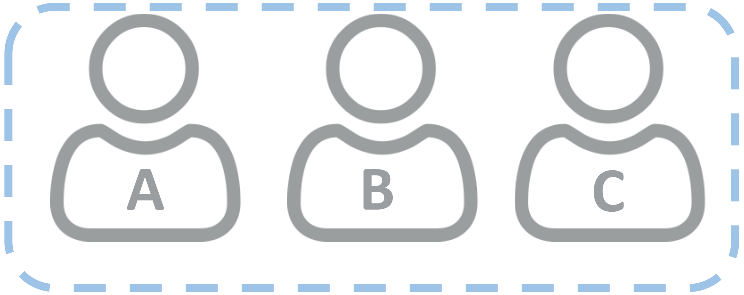
SHapley Additive exPlanations (SHAP)

- “Shapley Value” in cooperative game theory

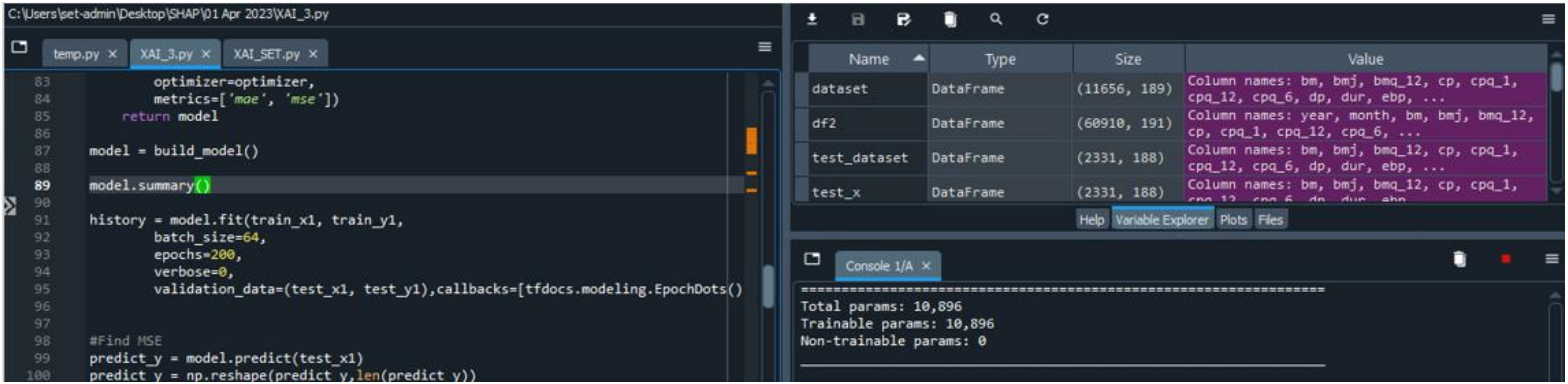




  - Kernel SHAP
  - Tree SHAP
  - Deep SHAP

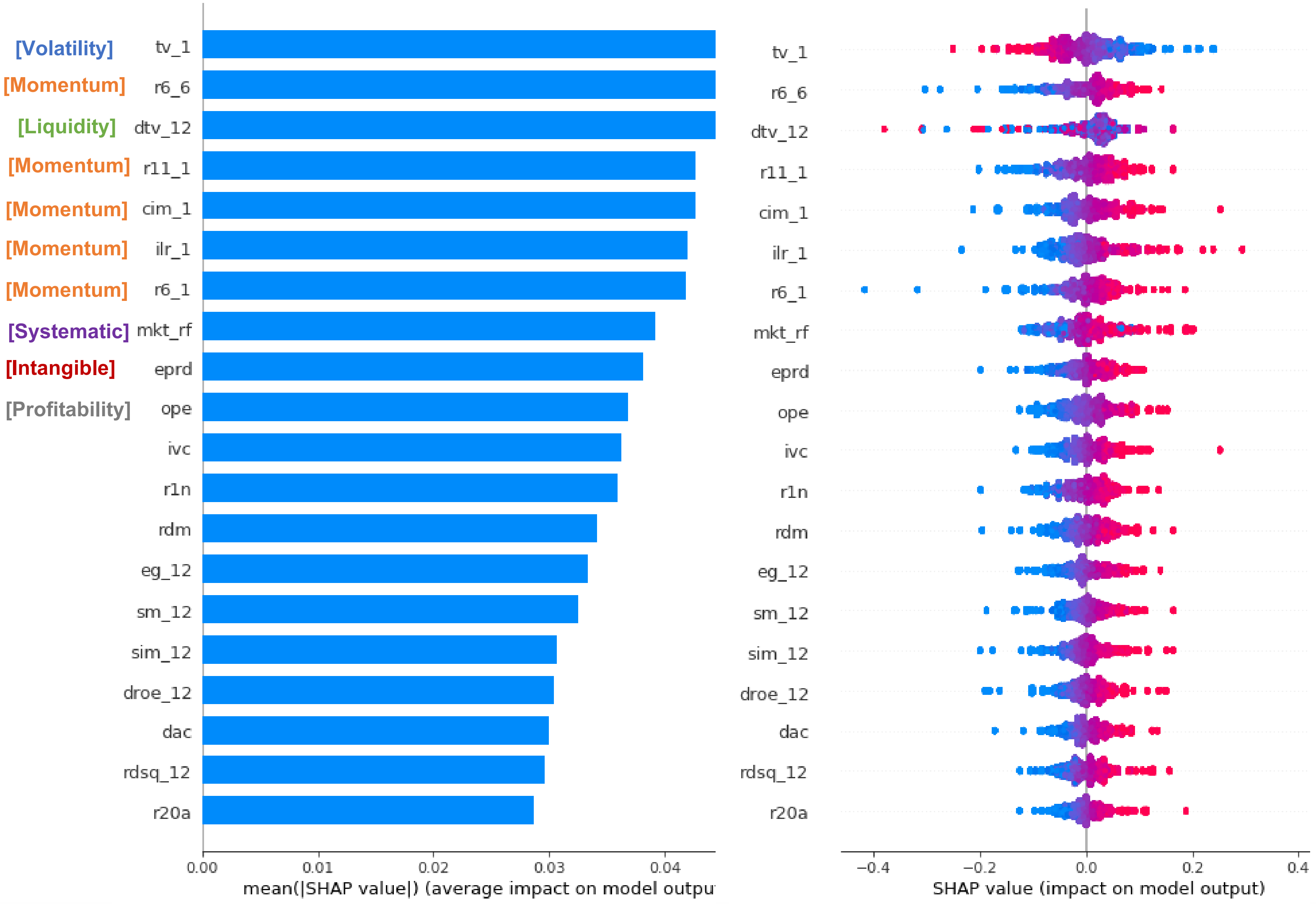


Running SHAP in real life!!



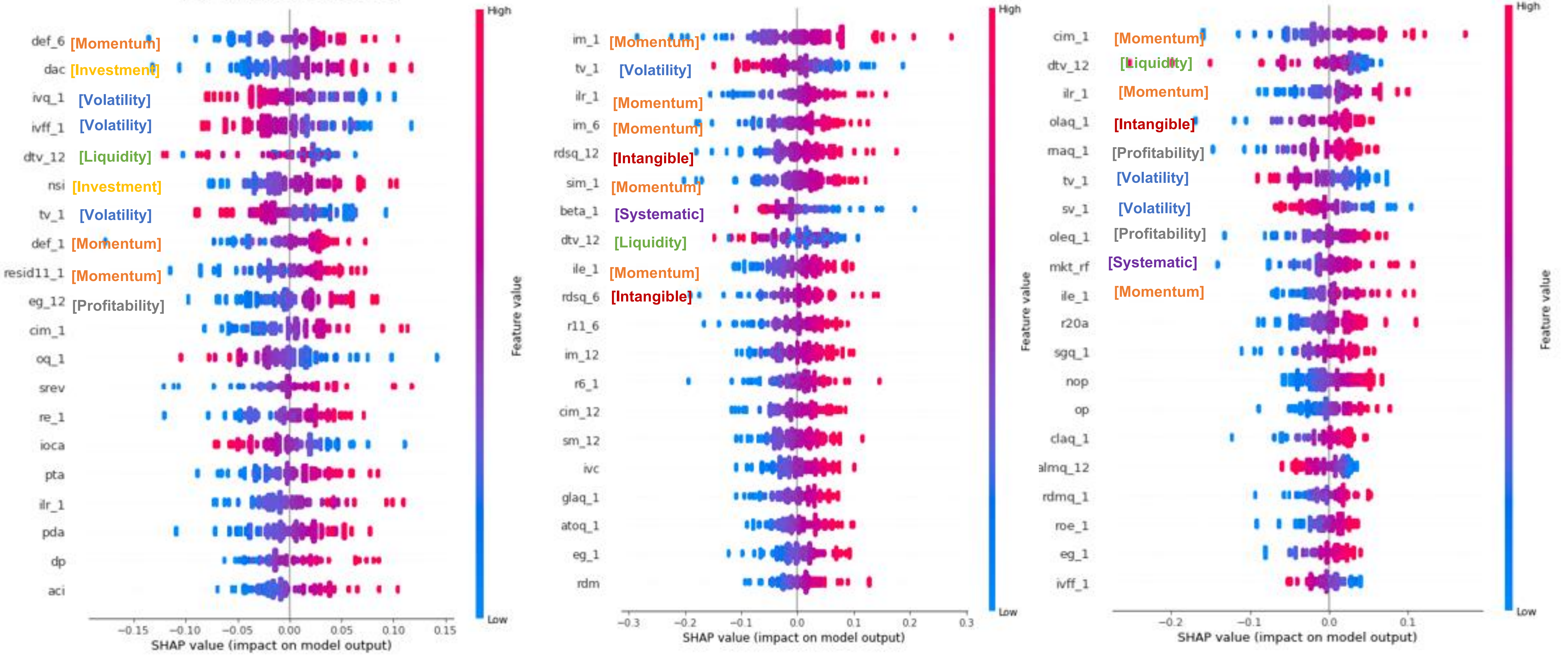
Overall Periods

Figure 1. Factor Importance for ANN in the Overall Periods



Subperiods

Panel A: 1991-2000 | Panel B: 2001-2010 | Panel C: 2011-2020



Conclusions

- Employ the SHAP method to explain returns
- Rank feature/factor importance
- Find that the top factors explaining returns in overall and subperiod periods differ
- Individual or institutional investors can use SHAP to explain the machine learning model
- Stock exchanges can explore the SHAP explanation and use it to explain how factors move asset returns in the market

SET | Make it Work for Everyone

www.set.or.th