XAI in Supervised Learning Model for Stock Price Manipulation Detection: Phase 2

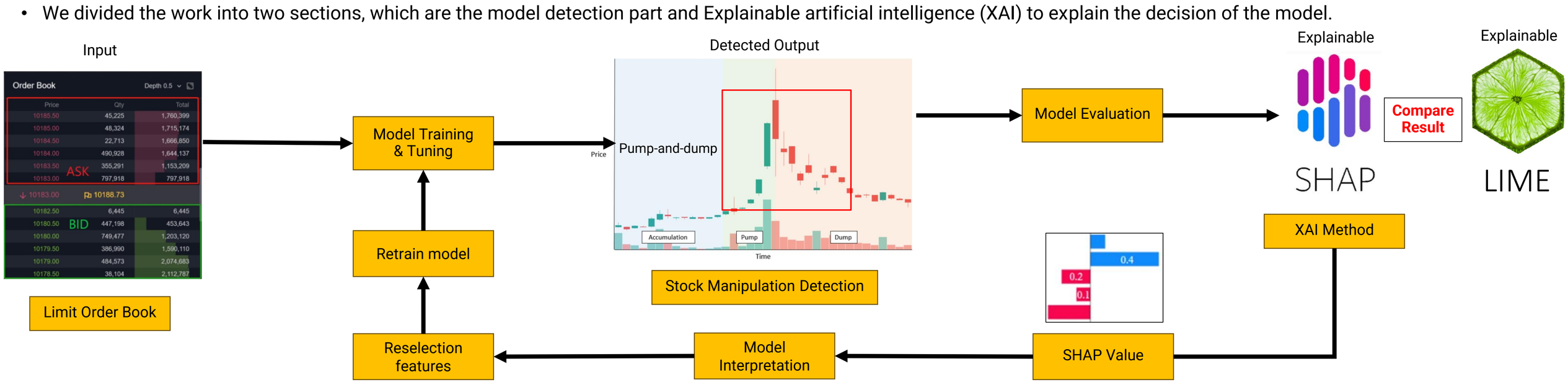
CONTROL SYSTEMS AND INSTRUMENTATION ENGINEERING PROGRAM
Thanapak Chantarasiriwattana, Varavit Kongsinrat, Advisor : Dr.Teema Leangarun

Abstract

Stock manipulation is an illegal action that happens by manipulators who want to make a profit. Pump-and-dump pattern is one of the most popular patterns. Manipulators will try to pump the stock price and collect a capital gain from dumping it. Many investors have been trapped in this action and lost their money. Today artificial intelligence has a role in many things including stock manipulation detection. The problem with new technology is reliability. We proposed the explainable artificial intelligence (XAI) combined supervised learning model, which used the dataset from the Stock Exchange of Thailand (SET). The global interpretation shows The top four important features have dominated the model prediction. There are Rate of change of best bid price, Rate of change of best ask price, Canceled bid volume, and Match volume. Furthermore, we compare our results with the XAI phase one. The results in the XAI phase one were the same as in the global interpretation. We also did local interpretation to illustrate that manipulation does not occur every second, and perhaps there is more than pump-and-dump but another manipulation pattern.

System Overview

- We divided the work into two sections, which are the model detection part and Explainable artificial intelligence (XAI) to explain the decision of the model.



Method

The interpretation of the model using SHAP values can categorized into two types:

1.Global interpretation: This is an interpretation to see the overviews of all model predictions.
2.Local interpretation: This is an interpretation to focus only one model prediction at a time.

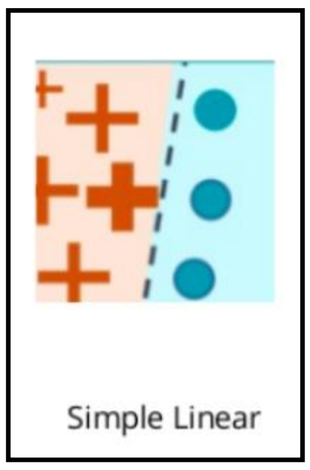


Figure 1: Local interpretation

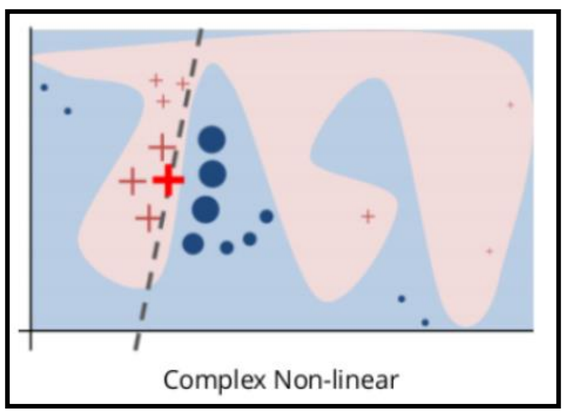


Figure 2: Global interpretation

We used Shapley additive explanations (SHAP) with the Long Short-Term Memory (LSTM) model, which is a supervised learning model. Estimating the actual model by SHAP can be done in several ways. In this research, the DeepSHAP technique has been applied. It is a suitable technique for approximating SHAP values of deep-learning models because it uses (DeepLIFT + SHAP values). DeepLIFT uses backpropagation to evaluate SHAP values.

The results of global interpretation can show many plots of the instance by SHAP summary plot, and the results of local interpretation can be seen by plotting the SHAP force plot. The feature's importance results will compared to the LIME results from the phase 1 experiment. The SHAP results will used to train new models and compare the performance of the models.

Result

The results of global interpretation provided by SHAP and LIME were similar in the Top four most contribution features which are 1. Rate of change of best bid price, 2. Rate of change of best ark price, 3. Cancelled bid volume, and 4. Match volume.

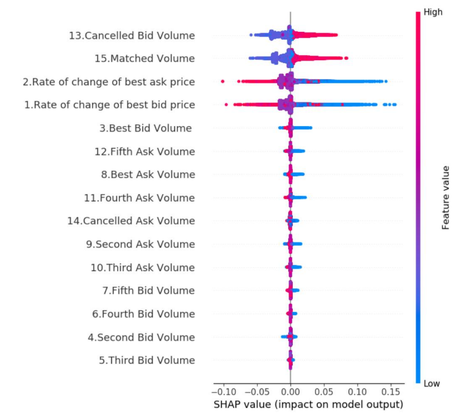


Figure 3: Global interpretation by SHAP

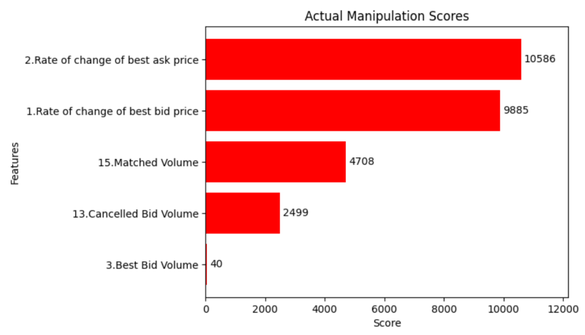


Figure 4: Global interpretation by LIME

In the re-training model session, the main process is to reselect the features and separate them into three categories. We named three categories by their method,
1. Top 4 and 2. Top 5 from SHAP summary plots, and 3. Correlation analysis plus SHAP from the REFRESH method [3]. We re-trained the models and collected their performance.

Dataset	Epoch	Synthesize manipulation data				Real manipulation data
		Accuracy	Precision	Recall	F1-score	Stock Detected
Original	210	0.9878	0.9999	0.9370	0.9674	4
Top 4	55	0.9847	0.9999	0.9211	0.9589	4
Top 5	36	0.9438	1.00	0.7102	0.8305	3
CA + SHAP	3	0.9294	1.00	0.6359	0.7775	5

Table 1. we made a comparison of every model to the original model. The comparison showed that the Top 4 category is the best group.

Conclusion

The interpretation results from SHAP and LIME in the Top four contribution features were the same. SHAP is strong in global interpretation, which has a summary plot to be a great visualization of the overall feature's contribution. LIME is proficient in local interpretation, which can provide deeper information, such as the probability of each window. According to re-trained results, we could summarize that the Top 4 is the best re-train category. This performance is outstanding enough to conclude that using only the top four contribution features can maintain the performance of the original model.

Reference

[1] Chullamonthon, P. and Tangamchit, P., 2023, “Ensemble of supervised and unsupervised deep neural networks for stock price manipulation detection”, Expert Systems with Applications, Vol.220, No.1.
[2] Tan B, Gan Z. and Wu Y. (2023). The measurement and early warning of daily financial stability index based on XGBoost and SHAP: Evidence from China, ScienceDirect , School of Economics and Management, China Three Gorges University, Yichang 443002, Chin
[3] Shubham Sharma, Sanghamitra Dutta, Emanuele Albini, Freddy Lecue, Daniele Magazzeni, Manuela Veloso., 2024 “REFRESH: Responsible and Efficient Feature Reselection Guided by SHAP Values”