MARKET MANIPULATION DETECTION by SUPERVISED LEARNING

Project Supervisor: Dr. Liu Qi **Second Examiner**: Dr. Zou Difan

Team FYP23058 : Ho Megan Qian Hua (3035832749), Li Wo Him (3035783986), Tsang Hoi Wei (3035785879)

MOTIVATION

As our world becomes increasingly sophisticated, it simultaneously offers more avenues for individuals with fraudulent intentions to exploit. **Fraud** has become an alarmingly pervasive issue, occurring at an astonishing rate of approximately **every 15 seconds**. The focus of the current market surveillance system is on sudden fluctuations or unusual behaviors in share prices or trading volumes. However, **this current investigation process still heavily relies on manual tracking and rule-based systems**. The methods used may need to become more efficient in detecting market manipulation.

METHODOLOGIES

Step 1 Data Collection from Shanghai and Shenzhen Stock Exchange (Price, Volume, PE Ratio, Beta, Realised Volatility, Current Ratio, Quick Ratio).

Step 2 EDA Analysis.

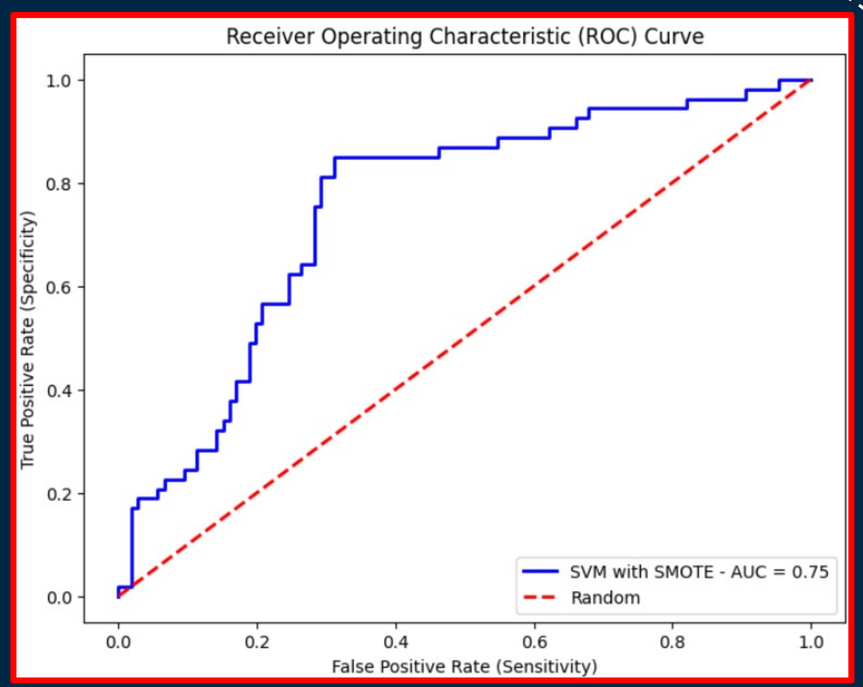
Step 3 Model Training (Support Vector Machines SVM, Decision Trees DT, Naive Bayes NB, Logistic Regression LR).

Step 4 Hyperparameters Tuning (e.g. DT - Max Depth, Loss Function, Random Forest).

Step 5 Perform oversampling method, SMOTE technique.

RESULTS

	AUC	F-Score	Accuracy
Support Vector Machine (SVM)	0.62	0.43	0.67
SVM-SMOTE	**0.75**	**0.63**	**0.72**
Decision Tree (Boosting)	0.74	0.57	0.71
Decision Tree (Random Forest)	0.69	0.38	0.63
Naïve Bayes	0.62	0.58	0.54
Logistic Regression (LR)	0.67	0.04	0.72
LR-SMOTE	0.53	0.58	0.72



- The **SVM with SMOTE** outperforms the rest of the models with the **highest accuracy and AUC**.
- **Balanced sensitivity and specificity** in determining the true positives and true negatives.
- The **high AUC** demonstrated that the model is **suitable for binary classification**, which aligns with our dataset of detecting violated cases and non-violated cases.

DISCUSSION

From the model training process and results obtained, we suggest some measures that can be done to enhance model performance:

Social Media NLP - Some types of market manipulations are largely contributed by retail investors, the Chinese stock forum, Guba (股吧) is considered a popular forum for stock market.

Separate Model Training - Refined to tailoring models to specific types of market manipulations identified in database (e.g. assume violation type P2501 is most easily detected by SVM).

Anomalies Real Time Detection - Dynamically adjusted probabilities of market manipulation detected from different models and report the trade if the probabilities are high among various models.