Crypto currency Return and Price Prediction with Machine Learning

Project Group 4

Shuyue Jia

shuyuejia3-c@my.cityu.edu.hk

Data Pre-processing and Feature Selection

- \rightarrow **Data cleaning**: drop all rows that contain **NaN** values
- \rightarrow **Data sampling**: Sample one-minute interval into **one-hour data** to <u>decorrelate and reduce redundancy</u>
- \rightarrow Feature Selection:

✓ Select 4 features out of 8 since the open, high, low, close, and VWAP are *highly correlated*

✓ Applied a PCA to the 8 features and <u>reduce dimensionality</u> to 5 features \leftarrow Better Performance

 \rightarrow Feature Normalization:

 \checkmark **Robust Scalar** to <u>remove outliers</u> based on the first and third quantile range

✓ **Min-max Standardization** to *normalize features between 0 and 1*

 \rightarrow Data Splitting and Cross-validation:

 \checkmark 90% Training set, 5% Validation set, and 5% Testing set

√ 10-fold Cross-validation for *hyper-parameter tuning and avoid training bias*

→ **Evaluation Metrics**: PLCC, SRCC, MSE, RMSE, MAE

Return Prediction via Single-value Function Approximation



Price Prediction via Correlation Approximation



Conclusions and Insights

✓ **Insight**: Returns (*relative factor*) Prediction \leftarrow Single-value function approximation

Price (*absolute factor*) Prediction \leftarrow Correlation approximation

✓ **Insight**: For time-series signals, **overfitting** should be seriously considered.

✓ **Dedicated Data Pre-processing** and **Feature Selection** are the necessary prerequisites.

✓ Dimensionality Reduction, e.g., PCA, and Data Argumentation improve performances.

✓ Various models should be compared, especially the Deep Learning-based models.