Cryptocurrency High-Frequency Liquidity Strategy based on Orderbook Behavior

Lynn, 朱致伶

2024/10/07 ML_FinTech Project Slide

Outline

1. Motivation
2. Design of Analysis
3. Data
    - Pre-processing
    - Feature engineering
    - Exploratory data analysis
4. Method
5. Experiment and Result
6. Conclusion

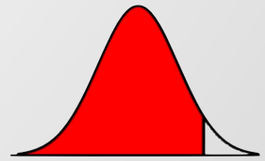
Motivation

- Cryptocurrency market is renowned for its high volatility and fragmented liquidity across different exchanges.
- **Liquidity** plays a crucial role in determining the efficiency and stability of the market.
- Liquidity is **not only a trading consideration but also a crucial element of risk management**.
- Orderbook Behavior which reflects real-time market, also liquidity conditions.

Motivation

- Goal: Develop a **predictive model for cryptocurrency liquidity** using orderbook data.
- Why the project is important:

- Improving Market Efficiency
- Risk Management through Liquidity
- Reducing Trading Costs



Design of Analysis

Data Collection and Preprocessing

Crawl order book data From Binance API, implement rolling window approach.

Data Featuring

Design and craft meaningful feature from raw data.

Model Training and Optimization

Design and train different models, including:

- Linear regression
- Super vector machine
- XGBoost
- Long-short term machine

Model Evaluation

Assess model performance employed metrics MSE, RMSE, MAE, R^2.

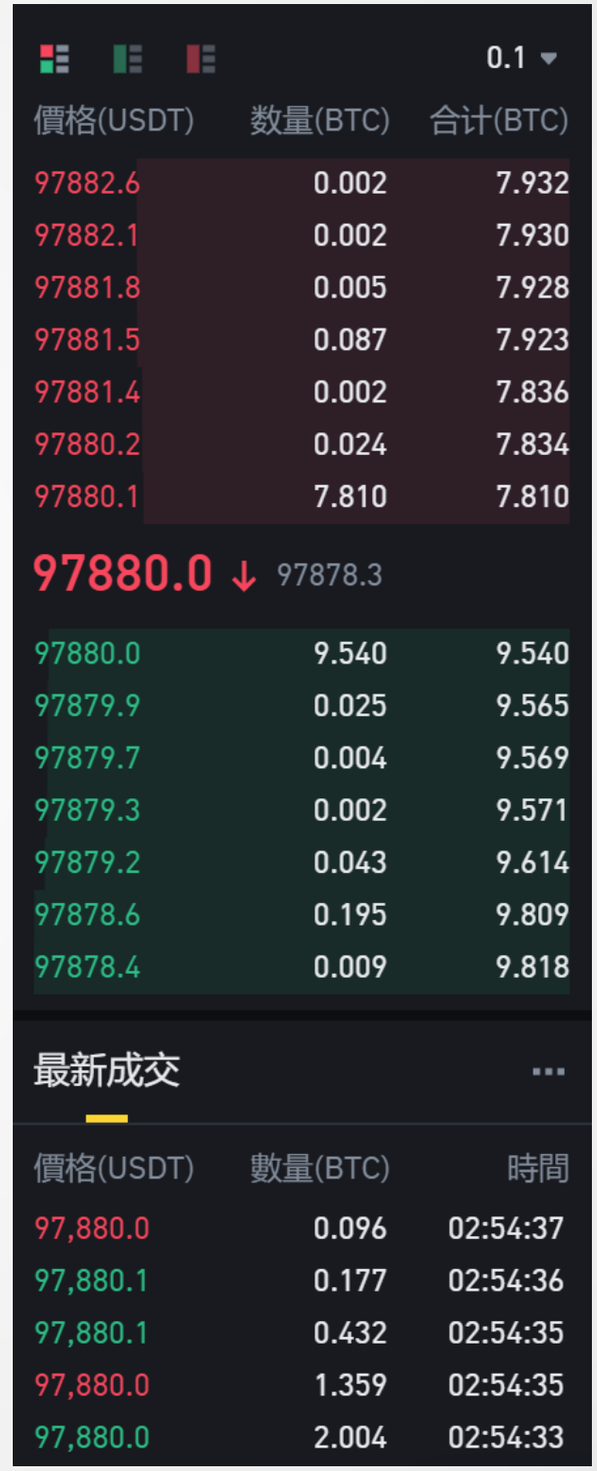
Conclusion

Interpret model performance and propose a strategy.

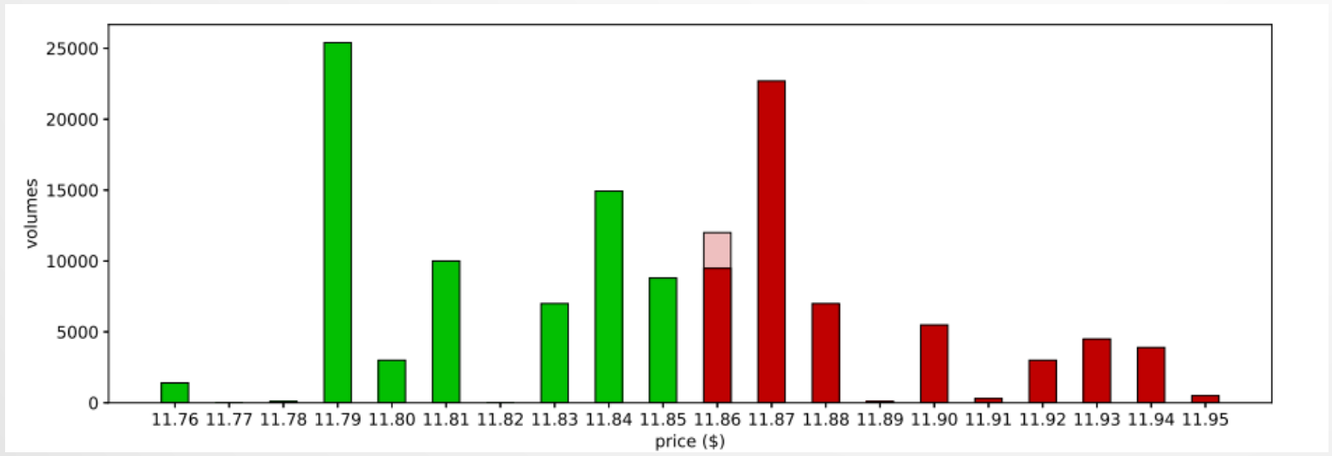
Data

- Binance Order Book API : BTCUSDT
- Response example

```
{
  "T": 1589436922972,   // transaction time
  "u": 37461            // update id
  "bids": [             // Buy order
    [
      "1000",           // Price
      "0.9"             // Quantity
    ]
  ],
  "asks": [             // Sell order
    [
      "1100",           // Price
      "0.1"             // Quantity
    ]
  ]
}
```

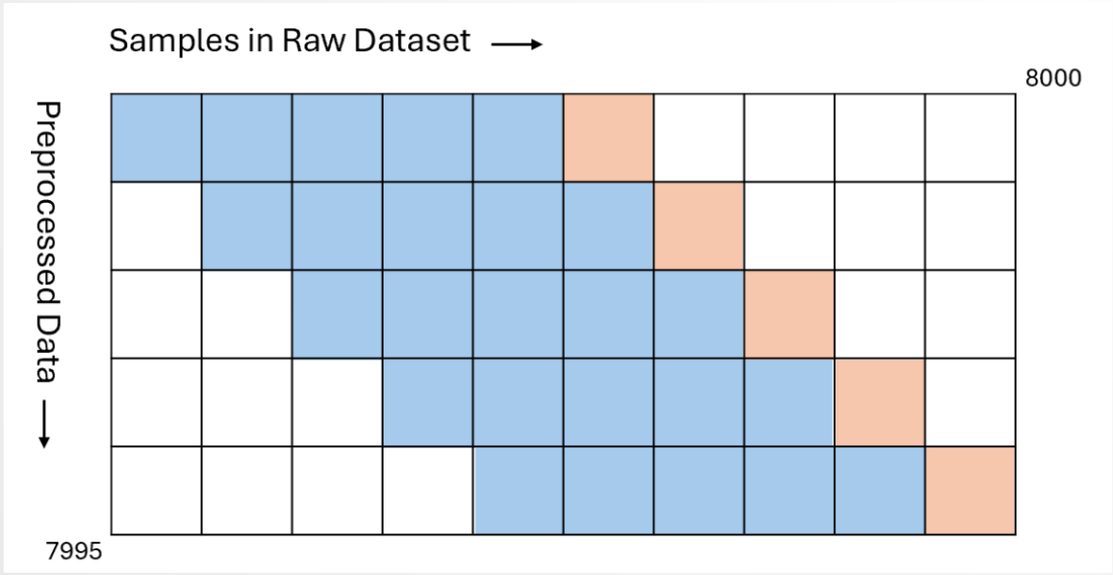


- Order book data schematic diagram



Data and Pre-processing

- Bitcoin (BTC) paired with Tether (USDT)
  - Market Dominance
  - Liquidity
  - Price Stability
  - Data Quality
- Implement rolling window approach

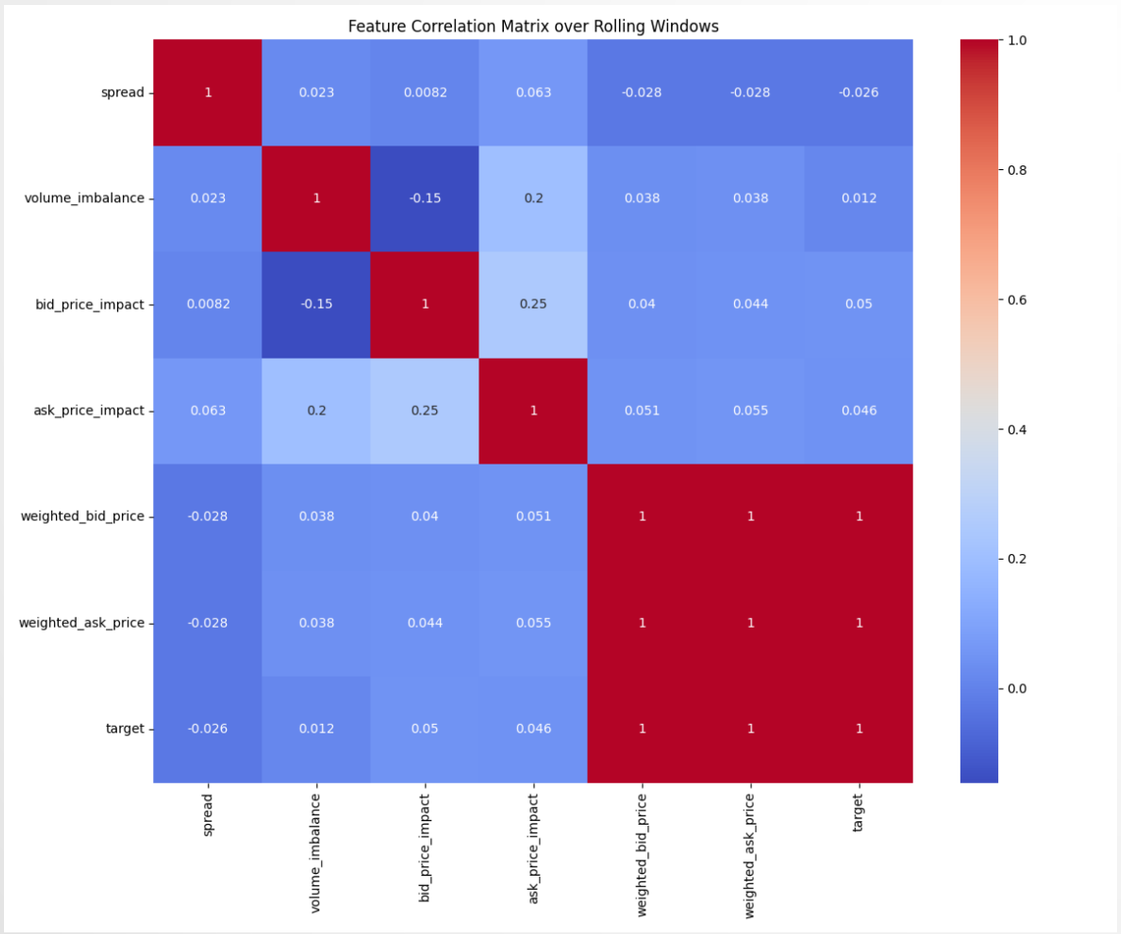


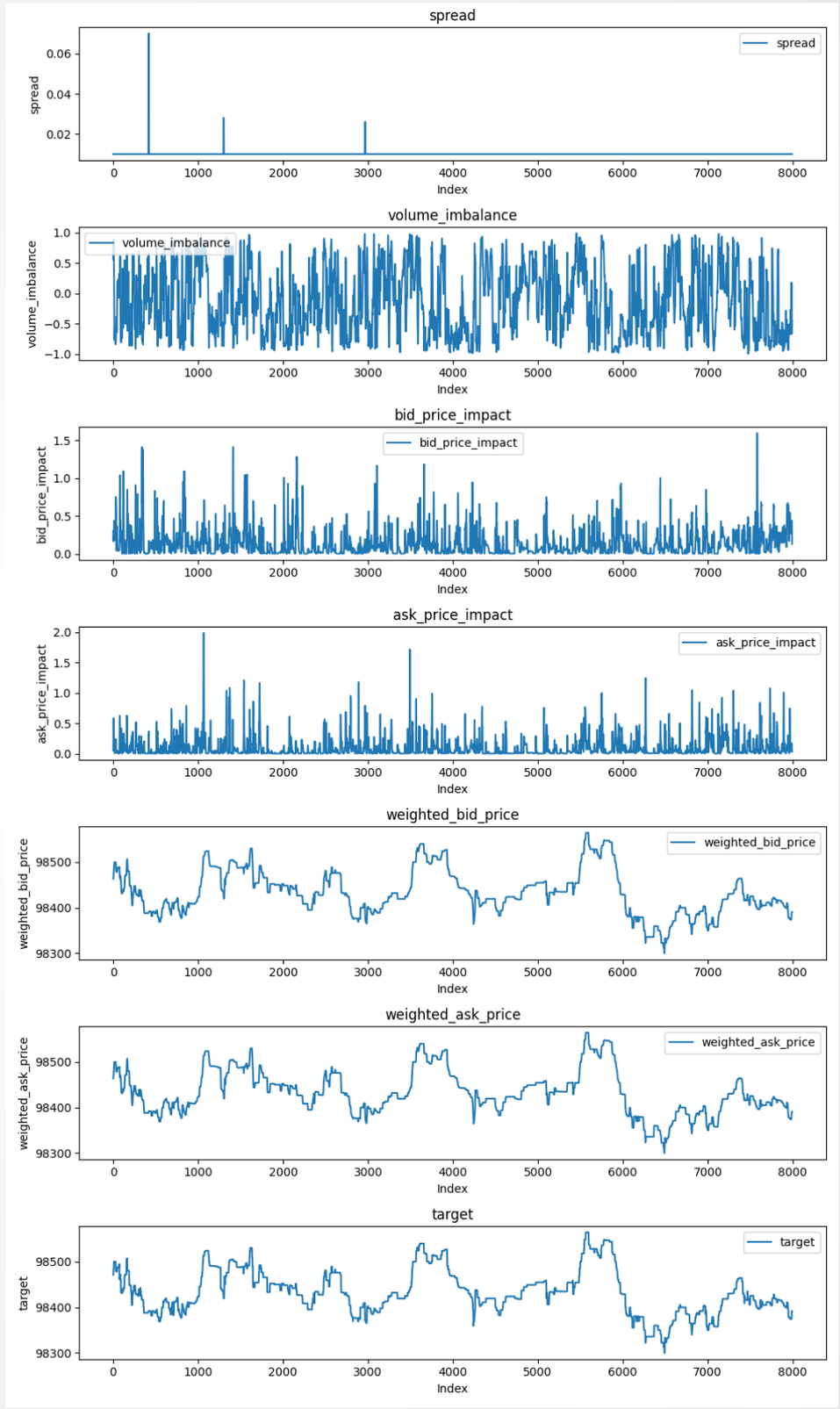
Feature Engineering

- Extract several features from the raw order book dataset for better model learning performance.
  - $Mid\ Price(Target) = {}^{P_{ask}+P_{bid}}/_{2}$
  - $Spread = P_{ask} - P_{bid}$
  - $Volume\ Imbalance = \frac{\sum Vol_{ask} - \sum Vol_{bid}}{\sum Vol_{bid} + \sum Vol_{ask}}$
  - $Price\ Impact = \frac{\sum(V_i \times |P_i - P_{mid}|)}{\sum V_i}$
  - $Weighted\ Price = \frac{P_i \times V_i}{\sum V_i}$

Exploratory Data Analysis

- Data distribution
- Correlation heatmap



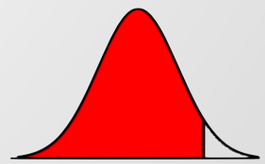


Method

- Split dataset to training(70%) and testing(30%)
- Train and optimize the below machine learning model
  - Linear regression
  - Super-vector regression
  - XGBoost
  - Long-short term machine

Slide Title

- Use
- All
  - w2
  - w3

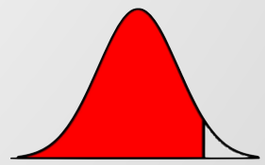


Logo and Links to Quantinar Courselets

- Use Quantinar icon and name as source

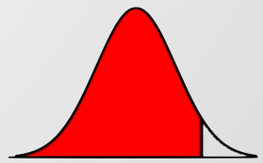
Logo and Links to Quantlet/GitHub

- Use Quantlet icon and name as source
- Hyperlink both to GitHub repository  Styleguide
- Change the presentation logo in the master slide (see View/Edit Master Slide, shortcut: Shift-Command-E)



LvB notations 1

- Use the formula creator within keynote ‘Insert/Equation’
- All operators are to be defined by \operatorname{}
  - without operatorname: ~~$argmax_i f(x_i)$~~
  - with operatorname: $\operatorname{argmax}_i f(x_i)$
- Equations covering multiple lines may be written aligned
- Use bracket sequence $[\{(\dots)\}]$
- Conventional bracket rules represent and exemption of the rule above. For example: $Y \sim \mathcal{N}(\mu(X), \sigma(X))$



LvB notations 2

- Use ^{\top} to write the transpose symbol: $x^\top x = \| x \|$
- Use \ldots to write the three dots symbol: $x \in \{1, \ldots, n\}$ $\hat{}$ $\tilde{}$
- Use \widehat{} and \widetilde{} rather than \hat{}, \tilde{}: $\widehat{Y}, \widetilde{Y}$
- Write norms via \|: $\| x \|$

LvB notations 3

- The for convergence may be written with \mathcal{O}: $\mathcal{O}$
- The operator for exponential terms with Euler's number as the base is defined by \exp: $\exp(1) \approx 2.718$
- Use \overset{\mathcal{L}}{\rightarrow} to write the symbol for convergence in distribution and denote the normal distribution by \mathcal{N}, this produces $X \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$
- Use \overset{\operatorname{as.}}{\sim} to write the symbol for asymptotic distribution $X \overset{\text{as.}}{\sim} \chi^2$
- To define a function, variable etc. use \overset{\operatorname{def}}{=} $f(x) \overset{\text{def}}{=} ax + b$

LvB notations 4

- Use \log for the natural logarithm: $\log\{\exp(1)\} = 1$
- Use \mathsf{E} for expectation: $\mathsf{E}[X] = \mu$
- Use \operatorname{P} to write the symbol for probability: P
- Use \operatorname{\mathbf{I}} for the indicator function: $\mathbf{I}\{x < 1\}$
- Use \varepsilon instead of epsilon: $\not\epsilon \rightarrow \varepsilon$

Tables

- Follow the Cambridge University Press Style
- Round appropriately (as much information as necessary, as little as possible)
- Align decimal points

d	10	11	12
10%	2.2886	2.4966	2.6862
5%	2.5268	2.7444	2.9490
1%	3.0339	3.2680	3.4911

Figures

- Give informative axis labels
- If x- and y-axis are on the same domain, the plot should be square
- Use same color scheme for multiple plots if they show the same content.

Reference

- Binance Order Book API: https://developers.binance.com/docs/derivatives/option/market-data/Order-Book
- The Short-Term Predictability of Returns in Order Book Markets: a Deep Learning Perspective

TEN Template

Your Name

Your affiliation

Your Webpage

Repeat on last slide the lead picture