1 Signal Selection and Comparison (Volume)

1.1 Project Background and Objective

This project aims to develop quantitative trading strategies based on historical market data. The core objective is to identify effective technical signals that can be integrated into real-time trading systems. We focus on a single trading instrument: **BTCUSDT (Bitcoin vs. USDT)** to ensure consistency between the strategy design and live deployment.

The historical data set is obtained from the **Binance public API**, covering the period from May 1, 2024 to May 1, 2025, with a frequency of 1-minute candlesticks. Each record includes open, high, low, closed prices, and volume, totaling approximately 520,000 rows. This rich data set provides a solid foundation for signal generation, backtesting, and performance evaluation.

In this stage, we concentrate on five volume-based technical indicators.

- OBV (On-Balance Volume)

- AD (Accumulation/Distribution Line)

- ADOSC (Chaikin Oscillator)

- MFI (Money Flow Index)

- BOP (Balance of Power)

We construct separate buy-sell signals for each indicator and evaluate their quality through statistical analysis and visualization.

1.2 Signal Definitions and Construction Logic

In our core strategy function `generate_signals(df)`, we analyze the market every 3 minutes and assign one of three signal states (1 = buy, -1 = sell, 0 = hold) for each indicator based on directional changes or thresholds. The specific logic is as follows:

- **OBV Signal:** We compare the current OBV value with that from 3 minutes before. An upward move signals a buy, while a downward move signals a sell.

- **AD Signal:** Similarly, we assess the trend of the AD indicator over the last 3 minutes. An increase indicates a buy signal and a decrease indicates a sell signal.

- **ADOSC Signal:** Since this indicator oscillates around zero, we use zero as the decision boundary: positive values indicate buy signals, and negative values indicate sell signals.

- **MFI Signal:** We use a quantile-based thresholding approach: the 20% quantile (Q_{20}) of all MFI values serves as the buy threshold, and the 80% quantile (Q_{80}) as the sell threshold. A value below Q_{20} triggers a buy signal, and above Q_{80} triggers a sell signal.

Figure 1: MFI Quantile Thresholds: Buy < 31.93, Sell > 67.89

- **BOP Signal:** This is also thresholded at zero. A positive value indicates a buy signal, and a negative value indicates a sell signal.

All signals are evaluated every 3 minutes (i.e., every third row) to reduce signal frequency and enhance robustness.

1.3 Signal performance evaluation

We designed a simple backtesting system for each signal, with the initial capital set at $10,000. Whenever the signal changes, perform a buy or sell operation and record the cumulative account value.

To further understand the dynamic performance of each volume signal, we have plotted the overall trend of the indicators over time as follows:
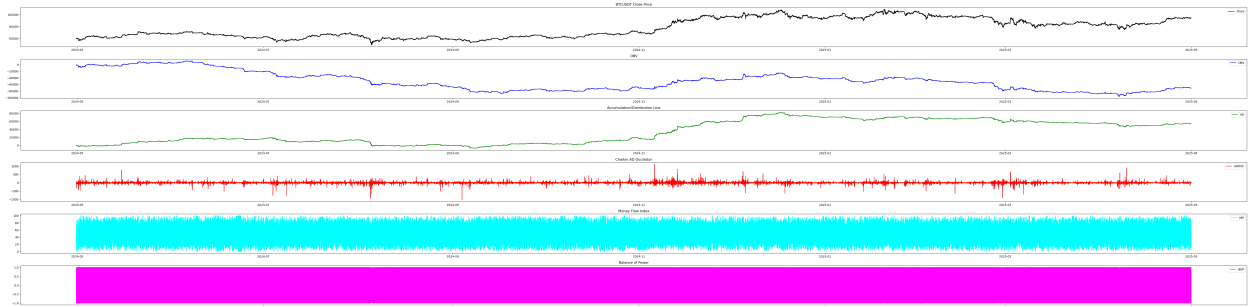

Figure 2: The historical evolution of BTCUSDT and five volume signals

This chart shows the response characteristics of various technical indicators at different price stages, intuitively reflecting their leading position, lagging position or noise level, providing background support for subsequent performance evaluation.

1.4 Backtesting results and analysis

The final account balances of each signal are as follows:

- MFI: $18,918.57 (**Best Performer**)

- AD: $17,356.93

- OBV: $6,364.60 (Loss)

- ADOSC: $5,868.97 (Loss)

- BOP: $5,275.44 (Loss)

```
--- Top 3 ---
1. MFI - Final_Balance: $18918.57
2. AD - Final_Balance: $17356.93
3. OBV - Final_Balance: $6364.60
```

Figure 3: Top 3 Performing Signals Summary

MFI: This signal combines price and trading volume information to capture changes in capital flows and is suitable for highly volatile markets or those with overbought/oversold characteristics. In the BTC data we use, its performance indicates that there are structural signals of capital inflows and outflows in the market at multiple stages, especially in the area of wave reversals.

```
Signal_MFI Final Balance: $18918.57
(Bought 6343 times in total, sold 6342 times
```

Figure 4: The backtest net value curves of MFI

AD: As a typical representative of volume-price divergence, AD is more suitable for trend trading environments. As a highly volatile asset, BTC's trend structure enables AD to identify the divergence state of "price rising but volume not following" or "price falling but volume supporting", thus remaining stable in medium and long-term backtesting.

```
Signal_AD Final Balance: $17356.93
(Bought 43759 times in total, sold 43759 times
```

Figure 5: The backtest net value curves of AD

OBV: The OBV indicator is susceptible to extreme trading volume interference, especially when the market fluctuates sharply, it may be misleading. Therefore, although its performance is positive, the pullback is large and the signal frequency is relatively high.

```
signal = df[signal_col][i]
Signal_OBV Final Balance: $6364.60
(Bought 43585 times in total, sold 43585 times
```

Figure 6: The backtest net value curves of OBV

ADOSC: It has a relatively high signal sensitivity and is prone to generating frequent false signals during sideways fluctuations, especially during periods when BTC is light, such as the Asian afternoon.

```
Signal_ADOSC Final Balance: $5868.97
(Bought 20865 times in total, sold 20865 times
```

Figure 7: The backtest net value curves of ADOSC

BOP: It measures the strength of bulls and bears, but lacks specific reference for price fluctuations, has poor stability, is sensitive to noise, and is prone to failure when the volatility of BTC reverses sharply or rapidly.

```
Signal_BOP Final Balance: $5275.44
(Bought 44175 times in total, sold 44175 times
```

Figure 8: The backtest net value curves of BOP

Analysis: MFI and AD performed the most outstandingly. MFI can better capture the overbought/oversold range and combines the dual characteristics of price and trading volume. AD also reflects the divergence relationship between price and trading volume relatively stably. Although OBV is profitable, it fluctuates greatly. The net returns of ADOSC and BOP are negative, which might be due to being overly sensitive or not adapting to the current market structure.

Conclusion: The comprehensive performance of MFI and AD is the best. It is recommended to incorporate it into the subsequent strategy development or connect it to the live trading. The OBV can continue to be observed. ADOSC and BOP are not recommended for live trading due to stability and risk control issues.

2 Different Signal Strategies and Live Trading Access

2.1 Strict Five Signal Strategies and Analysis

We initially designed five signal strategies based on historical BTCUSDT data and conducted preliminary real-time testing. These strategies are illustrated below:

- **Strategy 1: RSI Divergence + Hammer**
  Logic: RSI is rising while the price is falling, combined with a Hammer pattern at the bottom to confirm potential reversal.
  Advantage: Suitable for identifying trend reversals, especially effective when a Hammer appears at local lows.
  Motivation: To capture rebound opportunities following RSI-price divergence.

- **Strategy 2: MACD Flip + Shooting Star**
  Logic: MACD histogram flips from positive to negative, accompanied by a Shooting Star at the top.
  Advantage: Good for spotting short-term tops, particularly effective for volatile assets like BTC.
  Motivation: Aims to detect short-term peaks using both momentum and candlestick reversal signals.

- **Strategy 3: Bollinger Rebound + RSI + Bullish Engulfing**
  Logic: Price approaches or falls below the lower Bollinger Band, RSI indicates oversold, and a Bullish Engulfing pattern appears.
  Advantage: Combines volatility bands, momentum, and price action; ideal for identifying technical rebounds.
  Motivation: Designed to catch "oversold" rebounds and avoid chasing dips.

- **Strategy 4: ATR Spike + Volume Surge + RSI < 30**
  Logic: Sudden spike in volatility (ATR), surge in trading volume, and RSI drops below 30.
  Advantage: Highly responsive and frequent signal; good at capturing rebounds after sharp sell-offs.
  Motivation: Targets short-term bottom-fishing opportunities, ideal for active traders.

- **Strategy 5: RSI > 70 + MACD Cross + Bearish Engulfing**
  Logic: Overbought zone (RSI > 70), MACD bearish crossover, and Bearish Engulfing pattern.
  Advantage: Triple confirmation of top reversal; helps detect trend exhaustion.
  Motivation: Designed to capture exit signals near market tops, especially after strong uptrends.

We put these five signal strategies into the database for testing, and can obtain the number of signals captured by each strategy and the number of correct predictions. The specific data is shown in the following figure:

```
Backtesting...: 100%|████████████| 525998/525998 [01:34<00:00, 5553.93it/s]

=== Backtest Summary ===
                                Strategy  Signal Count  Correct Predictions  \
0                 RSI Divergence + Hammer             0                    0
1                MACD Flip + ShootingStar            24                   14
2      Bollinger Rebound + RSI + Engulfing           60                   35
3        ATR Spike + Volume Surge + RSI<30         2686                 1476
4  RSI>70 + MACD Cross + BearishEngulfing           24                   10

   Accuracy
0       N/A
1    58.33%
2    58.33%
3    54.95%
4    41.67%
```

Figure 9: Strict Single-Signal Strategies

We can easily find that there are over 500,000 pieces of data in the database, but only a very small amount of data can be captured, especially Strategy One, which has not captured any effective signals. From this, it can be seen that these strategies were overly strict, leading to low trigger frequencies. To enhance real-time applicability, we relaxed some of the conditions in the next stage.

2.2 Relaxed Signal Strategies and Evaluation

After relaxing the strategy conditions, the number of signals increased significantly. Since we generate signals every 5 minutes in real-time, the original conditions were too restrictive. The relaxed strategy rules include:

- **Strategy 1:** Hammer $= 100$ relaxed to Hammer > 0.

- **Strategy 2:** MACD histogram flip relaxed to MACD line $<$ Signal line; ShootingStar < 0 instead of exactly -100.

- **Strategy 3:** Close $<$ Lower Band $\times 1.01$, RSI > 45, Engulfing > 0.

- **Strategy 4:** ATR $> 1.1 \times$ mean, Volume $> 1.2 \times$ mean, RSI < 35.

- **Strategy 5:** RSI > 65, BearishEngulfing < 0.

Compared to the strict versions, the relaxed signal strategies loosen the filtering conditions to significantly increase signal frequency.

While the strict rules aimed for high precision with exact pattern matches (e.g., `Hammer == 100`, RSI > 70, MACD histogram flip), the relaxed versions allow a broader definition (e.g., `Hammer > 0`, RSI > 65, MACD $<$ Signal) to improve real-time applicability.

This trade-off leads to higher signal coverage but slightly reduced accuracy, making the relaxed strategies more suitable for real-time trading where signal availability is crucial.

The updated signal performance is illustrated below.

```
Backtesting...: 100%|███████████| 525997/525997 [35:16<00:00, 248.57it/s]

=== Backtest Summary ===
                              Strategy  Signal Count  Correct Predictions  \
0                RSI Divergence + Hammer           992                  482
1                MACD Flip + ShootingStar          1199                  645
2      Bollinger Rebound + RSI + Engulfing        10520                 5418
3        ATR Spike + Volume Surge + RSI<30        11660                 6114
4  RSI>70 + MACD Cross + BearishEngulfing          135                   70

   Accuracy
0    48.59%
1    53.79%
2    51.50%
3    52.44%
4    51.85%
```

Figure 10: Relaxed Single-Signal Strategies

While accuracy slightly decreased, the improved signal detection rate significantly enhanced usability. We observed that different strategies have distinct characteristics, which motivated us to explore combined strategies to improve performance.

2.3   Combined Strategies (Strict Version)

We first tested combinations of strict (unrelaxed) strategies such as Strategy 1 + Strategy 3, Strategy 2 + Strategy 5, Strategy 2 + Strategy 4, Strategy 3 + Strategy 4, and Strategy 1 + Strategy 2 + Strategy 4.

We first tested combinations of strict (unrelaxed) strategies, including:

- **Strategy 1 + Strategy 3:** RSI Divergence + Hammer combined with Bollinger Rebound + RSI + Bullish Engulfing

- **Strategy 2 + Strategy 5:** MACD Flip + Shooting Star combined with RSI > 70 + MACD Cross + Bearish Engulfing

- **Strategy 2 + Strategy 4:** MACD Flip + Shooting Star combined with ATR Spike + Volume Surge + RSI < 30

- **Strategy 3 + Strategy 4:** Bollinger Rebound + RSI + Bullish Engulfing combined with ATR Spike + Volume Surge + RSI < 30

- **Strategy 1 + Strategy 2 + Strategy 4:** RSI Divergence + Hammer, MACD Flip + Shooting Star, and ATR Spike + Volume Surge + RSI < 30

```
Backtesting Strategy 1 + 3 (strict): 100%|████████| 525997/525997 [01:16<00:00, 6916.72it/s]
Backtesting Strategy 2 + 5 (strict): 100%|████████| 525997/525997 [00:53<00:00, 9818.91it/s]
Backtesting Strategy 2 + 4 (strict): 100%|████████| 525997/525997 [01:01<00:00, 8597.60it/s]
Backtesting Strategy 3 + 4 (strict): 100%|████████| 525997/525997 [01:06<00:00, 7898.87it/s]
Backtesting Strategy 1 + 2 + 4 (strict): 100%|████████| 525997/525997 [01:18<00:00, 6663.74it/s]

=== Combined Strategy Backtest Summary (Strict Conditions) ===
                       Strategy  Signal Count  Correct Predictions Accuracy
0        Strategy 1 + 3 (strict)           60                   35   58.33%
1        Strategy 2 + 5 (strict)           48                   24   50.00%
2        Strategy 2 + 4 (strict)         2710                 1490   54.98%
3        Strategy 3 + 4 (strict)         2746                 1511   55.03%
4    Strategy 1 + 2 + 4 (strict)         2710                 1490   54.98%
```

Figure 11: Combined Strategies - Strict Version

2.4 Combined Strategies (Relaxed Version)

We then implemented the same combinations but using the relaxed versions of the strategies:

Strategy	Strict Signal Count	Strict Accuracy	Relaxed Signal Count	Relaxed Accuracy
1 + 3	60	58.33%	20,258	50.66%
2 + 5	48	50.00%	1,334	58.62%
2 + 4	2,710	54.98%	12,843	52.64%
3 + 4	2,746	55.03%	22,180	52.34%
1 + 2 + 4	2,710	54.98%	22,547	51.12%

Table 1: Signal Volume and Accuracy of Combined Strategies

Strategy	Recommendation
1 + 3	Useful as a high-confidence filter.
2 + 5	Best post-relaxation performance, recommended for live testing.
2 + 4	Balanced in volume and accuracy, considered a core real-time strategy.
3 + 4	Remains stable after relaxation, strongly recommended for deployment.
1 + 2 + 4	High signal volume but reduced accuracy, suitable for data generation or model training.

Table 2: Strategy Recommendations Based on Combined Results

Final Selection: Considering accuracy, signal frequency, and ease of deployment, we finally selected **Strategy 2 + 4 (Relaxed)** and **Strategy 2 + 5 (Relaxed)** for integration into our real-time signal visualization system.

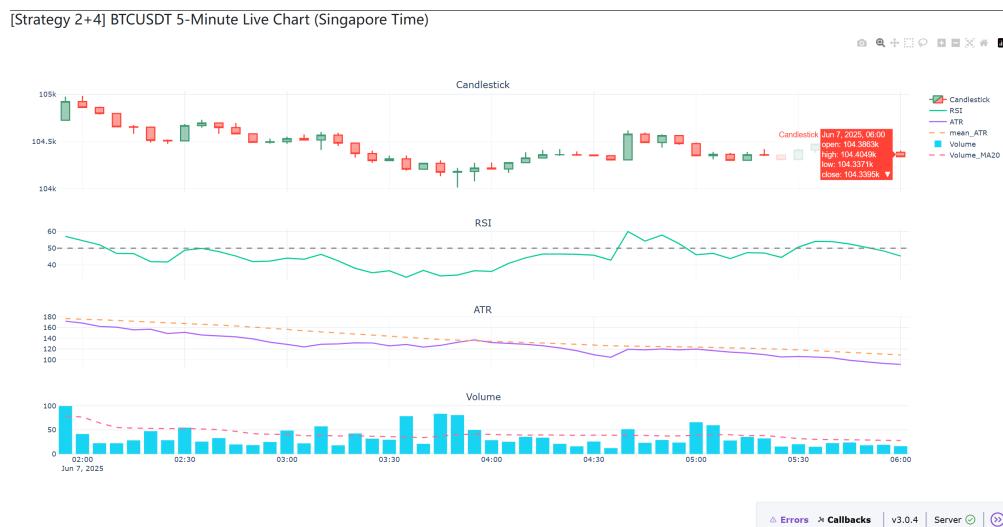## 2.5   Comparison of Strategies for Live Trading Access
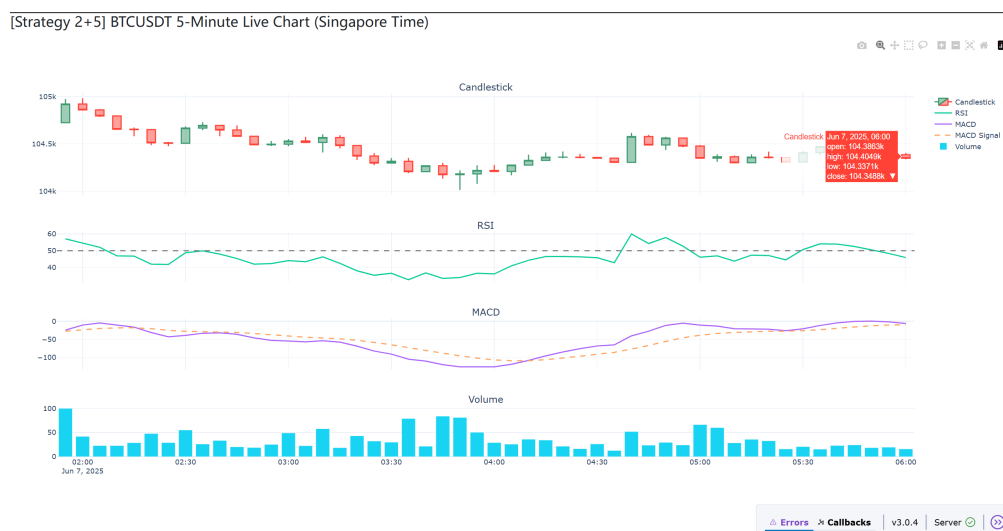

Figure 12: Strategy 2 + 4 (Relaxed)

Figure 13: Strategy 2 + 5 (Relaxed)

In both figures above, around 6:00am (June 7, 2025), a red downward-pointing triangle. It indicates that on this 5-minute K-line, the short selling condition of Strategy 2+4 and 2+5 was triggered.

```
[LightGBM] [Info] Number of positive: 2427, number of negative: 2427
[LightGBM] [Info] Auto-choosing col-wise multi-threading, the overhead of testing was 0.001719
seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 9947
[LightGBM] [Info] Number of data points in the train set: 4854, number of used features: 40
[LightGBM] [Info] [binary:BoostFromScore]: pavg=0.500000 -> initscore=0.000000
[Classification Report]
              precision    recall  f1-score   support

        -1.0       0.18      0.86      0.30       596
         0.0       1.00      0.89      0.94     20437

    accuracy                           0.89     21033
   macro avg       0.59      0.87      0.62     21033
weighted avg       0.97      0.89      0.92     21033

[Confusion Matrix]
[[  511    85]
 [ 2276 18161]]
```

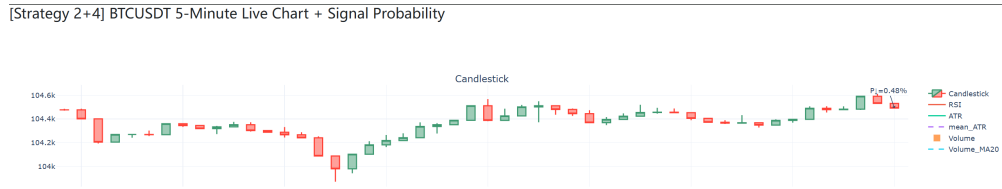Figure 14: LightGBM model parameters for Strategy 2+4

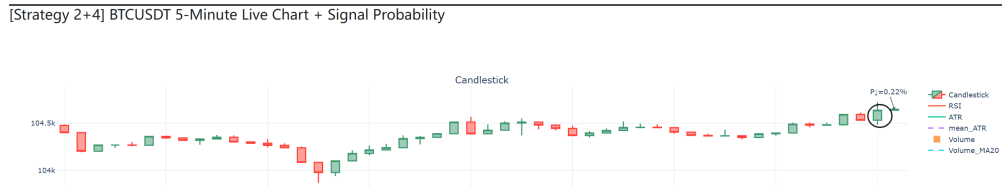Figure 15: Probability (p) of Strategy 2+4 at time t

Figure 16: Probability and actual signal of Strategy 2+4 at time $t + 5min$

The above figures illustrate signal prediction for Strategy 2+4. Since Strategy 2 and Strategy 4 are designed to detect bearish signals, we trained a LightGBM classifier to predict whether a bearish signal (-1) or no signal (0) would occur.

The value of p shown in the top right corner represents the model's predicted probability that a bearish signal will appear on the next candlestick. This probability is directly derived from the LightGBM model's output for class -1 (i.e., the "bearish" signal as defined by Strategy 2+4).

```
[LightGBM] [Info] Number of positive: 1332, number of negative: 3845
[LightGBM] [Info] Auto-choosing col-wise multi-threading, the overhead of testing was 0.000572
seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 2553
[LightGBM] [Info] Number of data points in the train set: 5177, number of used features: 11
[LightGBM] [Info] [binary:BoostFromScore]: pavg=0.257292 -> initscore=-1.060092
[LightGBM] [Info] Start training from score -1.060092
[Classification Report]
              precision    recall  f1-score   support

         0.0       0.86      0.88      0.87      1009
         1.0       0.55      0.51      0.53       286

    accuracy                           0.80      1295
   macro avg       0.71      0.69      0.70      1295
weighted avg       0.79      0.80      0.80      1295

[Confusion Matrix]
[[890 119]
 [141 145]]
Model saved as lgb_model_strategy25.pkl
```

Figure 17: LightGBM model parameters for Strategy 2+5

Figure 18: Probability (p) of Strategy 2+5 at time t

Figure 19: Probability and actual signal of Strategy 2+5 at time $t + 5min$

The above figures illustrate signal prediction for Strategy 2+5. Since Strategy 2 and Strategy 5 are designed to detect bullish signals, we trained a LightGBM classifier to predict whether a bullish signal (1) or no signal (0) would occur.

The value of p shown in the top right corner represents the model's predicted probability that a bullish signal will appear on the next candlestick. This probability corresponds to the output probability for class 1 (i.e., the "bullish" signal defined by Strategy 2+5).

Based on the four figures above, we observe that for Strategy 2+4, the model predicts the next 5-minute candle to be bullish (with a bearish probability of only 0.48%). For Strategy 2+5, the predicted bullish probability is 40.64%, which leans toward a neutral signal. However, based on continuous monitoring, we found that when the bullish probability in

Strategy 2+5 exceeds 25%, it often implies a potential upward movement.
Therefore, we believe that both strategies indicate a bullish signal within the next 5 minutes, and the subsequent charts confirm this view. However, such predictions are not always accurate. Long-term observation shows the prediction accuracy fluctuates around 50%, which is consistent with the historical signal accuracy observed during model training.

2.6   Summarize and Improve the Direction

- **Strategy 2 + 4**: Bearish signals based on MACD with Shooting Star (S2) and high volatility + low RSI (S4).

- **Strategy 2 + 5**: Bullish signals from MACD with Bullish Engulfing (S2) and high RSI + Bullish Engulfing (S5).

Both strategies were applied to Binance 5-minute data and validated with annotated charts. We trained two LightGBM classifiers using one year of 1-minute data (resampled to 5-minute), achieving high predictive accuracy and deployed them in a live dashboard showing real-time signals and probability forecasts.
Key Takeaways:

- Built and visualized two signal-based strategies for real-time trading.

- Trained LightGBM models to classify upcoming bullish/bearish signals.

- Integrated predictions into a live monitoring dashboard.

Future Work:

- Add backtesting and PnL tracking to evaluate model-driven signals.

- Expand features with trend/multi-timeframe indicators and volume sentiment.

- Explore more advanced models (e.g., LSTM) for sequential signal prediction.