Features from OHLCV Data: 5mins interval

May 29, 2025

1 Lagged Returns

- **return_1**: One-period (5 min) return:

$$\text{return}_1 = \frac{\text{Close}_t}{\text{Close}_{t-1}} - 1.$$

Remarks: This feature captures immediate price momentum. Positive values indicate upward movement since the last bar, which can be used to identify short bursts of trend or mean reversion opportunities.

- **return_3**: Three-period (15 min) return:

$$\text{return}_3 = \frac{\text{Close}_t}{\text{Close}_{t-3}} - 1.$$

Remarks: By aggregating over three bars, this return smooths out single-bar noise and highlights slightly more sustained intraday moves. It balances responsiveness and stability.

2 Rolling Statistics (Window = 5 bars ≈ 25 min)

- **MA_5**: Simple moving average of the last 5 closes:

$$\text{MA}_5 = \frac{1}{5}\sum_{i=0}^{4} \text{Close}_{t-i}.$$

Remarks: Moving averages smooth out random fluctuations, providing a clearer picture of the trend. Short windows react quickly but can be noisy.

- **STD_5**: Standard deviation of the last 5 closes:

$$\text{STD}_5 = \sqrt{\frac{1}{4}\sum_{i=0}^{4}(\text{Close}_{t-i} - \text{MA}_5)^2}.$$

Remarks: Standard deviation quantifies recent volatility. Higher values mean more risk or potential opportunity, useful for dynamic position sizing or volatility-based signals.

3 Price Range & Candlestick Features

- **HL_range**: Total intrabar movement:

$$\text{HL\_range} = \text{High}_t - \text{Low}_t.$$

Remarks: Large ranges often accompany news events or breakouts, signaling heightened trader activity.

- **upper_shadow**: Wicks above the candle body:

$$\text{upper\_shadow} = \text{High}_t - \max(\text{Open}_t, \text{Close}_t).$$

Remarks: A long upper shadow indicates that prices rallied but were pushed back, suggesting selling pressure at highs.

- **lower_shadow**: Wicks below the candle body:

$$\text{lower\_shadow} = \min(\text{Open}_t, \text{Close}_t) - \text{Low}_t.$$

Remarks: A long lower shadow shows buyers stepping in at the lows, often seen as a bullish signal if it follows downward pressure.

4 Moving Averages & EMA

- **SMA_12**: 12-bar simple moving average.
- **EMA_12**, **EMA_26**: Exponential moving averages with spans 12 and 26:

$$\text{EMA}_n(t) = \alpha \cdot \text{Close}_t + (1-\alpha) \cdot \text{EMA}_n(t-1), \quad \alpha = \frac{2}{n+1}.$$

Remarks: EMAs give more weight to recent prices, making them faster to respond. The lag trade-off between SMA and EMA can be leveraged to detect trend changes.

5 MACD & Signal Line

- **MACD**: Difference between EMAs:

$$\text{MACD}_t = \text{EMA}_{12}(t) - \text{EMA}_{26}(t).$$

Remarks: The MACD line shows the convergence/divergence of short and medium EMAs. When the line crosses zero, it signals potential shifts in momentum.

- **MACD_signal**: 9-period EMA of the MACD series. *Remarks:* Crossovers between MACD and its signal line generate trade signals: a bullish crossover (MACD rising above signal) suggests an entry, while a bearish crossover implies exit.

6 RSI (Relative Strength Index, 14-period)

Define:

$$\Delta_t = \text{Close}_t - \text{Close}_{t-1}, \quad \text{gain}_t = \max(\Delta_t, 0), \quad \text{loss}_t = \max(-\Delta_t, 0).$$

Compute:

$$\text{RS}_t = \frac{\text{AvgGain}_{14}(t)}{\text{AvgLoss}_{14}(t)}, \quad \text{RSI}_{14}(t) = 100 - \frac{100}{1+\text{RS}_t}.$$

Remarks: RSI oscillates between 0 and 100. Readings above 70 indicate overbought conditions; below 30 indicate oversold. Divergences between RSI and price can prefigure reversals.

7 Bollinger Bands (20-period)

$\text{BB\_mid} = \text{MA}_{20}$, $\text{BB\_std} = \text{STD}_{20}$, $\text{BB\_upper} = \text{BB\_mid} + 2\times\text{BB\_std}$, $\text{BB\_lower} = \text{BB\_mid} - 2\times\text{BB\_std}$.

Remarks: Bands expand when volatility rises and contract during quiet periods (a "squeeze"). Breakouts after squeezes often lead to strong trends.

8 ATR (Average True Range, 14-period)

Define true range:

$$\mathrm{TR}_t = \max\big(\mathrm{High}_t - \mathrm{Low}_t,\ |\mathrm{High}_t - \mathrm{Close}_{t-1}|,\ |\mathrm{Low}_t - \mathrm{Close}_{t-1}|\big).$$

Then:

$$\mathrm{ATR}_{14}(t) = \frac{1}{14}\sum_{i=0}^{13} \mathrm{TR}_{t-i}.$$

Remarks: ATR measures market volatility and is often used to set stop-loss levels: wider ATR implies wider stops to avoid whipsaws.

9 VWAP (5-period)

$$\mathrm{VWAP}_5 = \frac{\sum_{i=0}^{4}(\mathrm{Close}_{t-i} \times \mathrm{Volume}_{t-i})}{\sum_{i=0}^{4} \mathrm{Volume}_{t-i}}.$$

Remarks: VWAP gives more weight to high-volume bars, serving as a benchmark for trade execution quality. Institutional traders aim to buy below VWAP and sell above.

10 OBV (On-Balance Volume)

$$\mathrm{OBV}_t = \sum_{i=1}^{t} \big[\mathrm{sign}(\mathrm{Close}_i - \mathrm{Close}_{i-1}) \times \mathrm{Volume}_i\big].$$

Remarks: Rising OBV while price rises confirms uptrends; if OBV diverges (ticks down while price rises), it may signal weakening momentum.

11 Time Features

- **hour**: Hour of the day (0–23). Encodes intraday seasonality, such as peak trading hours.
- **minute**: Minute of the hour (0–59). Useful for capturing finer time buckets (e.g., quarter-hour trends).
- **weekday**: Day of week (0=Monday, ..., 6=Sunday). Captures weekly patterns; weekends often see different liquidity and volatility.