

OrderBook Feature Engineering and Mid-Price Prediction Using LightGBM

Overview

This project predicts short-term **mid-price direction (Up/Down)** using **LightGBM** on engineered **Limit Order Book (LOB)** microstructure features.

LightGBM is chosen because it:

- Handles large LOB datasets efficiently
- Works well with tabular microstructure features
- Captures non-linear relationships
- Provides high interpretability with feature importance

Dataset Description

The dataset contains top-5 level order book data:

Raw Columns

- BidPrice1–BidPrice5
- BidQty1–BidQty5
- BidNumOrders1–BidNumOrders5
- AskPrice1–AskPrice5
- AskQty1–AskQty5
- AskNumOrders1–AskNumOrders5
- LTP, LTQ
- Timestamp

Feature Engineering

The following engineered features are used for LightGBM:

Price Features

- MidPrice
- MicroPrice
- Spread
- Relative Spread

Order Imbalance & Pressure

- Top-level imbalance
- 5-level depth imbalance
- Order Flow Imbalance (OFI)
- Weighted Mid Price (WMP)

Liquidity & Depth

- Total Bid Depth
- Total Ask Depth
- Depth Ratio

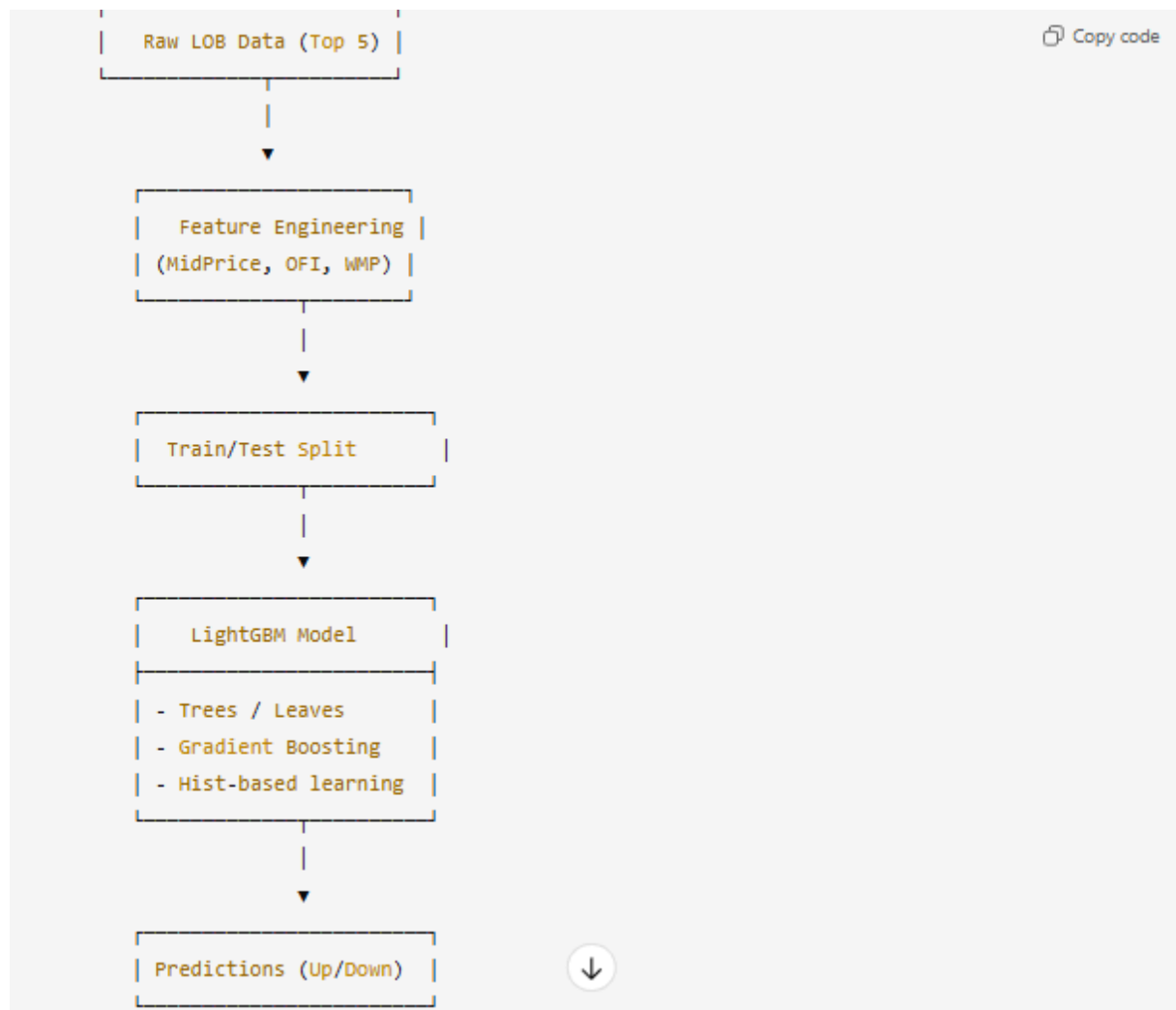
Volatility

- Rolling Volatility
- EWMA Volatility
- Parkinson Volatility

These features significantly improve LightGBM accuracy.

Model: LightGBM Classifier

```
.
├─ data/
│   └─ orderbook.csv
├─ features/
│   └─ feature_engineering.py
├─ models/
│   └─ lightgbm_model.pkl
├─ plots/
│   ├── feature_importance.png
│   └─ architecture_lightgbm.png
├─ results/
│   └─ metrics.json
├─ training/
│   └─ train_lightgbm.py
├─ README.md
└─ requirements.txt
```



Final Predictions

