Regime_Detection_Report

Objective

The objective of this project is to segment financial market behavior into distinct market regimes using unsupervised learning methods. The regimes are characterized based on:

- Trending vs Mean-Reverting
- Volatile vs Stable
- Liquid vs Illiquid

The analysis uses order book (depth20) and trade volume (aggTrade) data to extract meaningful features and detect market patterns without any labels.

1. Feature Engineering

We engineered a comprehensive set of features to capture the dynamics of liquidity, volatility, and price movement:

Liquidity & Depth Features:

- Bid-Ask Spread: ask_1_price - bid_1_price
- Order Book Imbalance: (bid_qty_1 - ask_qty_1) / (bid_qty_1 + ask_qty_1)
- Microprice: (bid_1_price * ask_qty_1 + ask_1_price * bid_qty_1) / (bid_qty_1 + ask_qty_1)
- Cumulative Depth: Sum of quantities across top 20 bid/ask levels

Volatility & Price Features:

- Rolling Mid-Price Returns: log(mid_t / mid_t-1)
- Rolling Volatility: Standard deviation of returns over past 10s, 30s

Volume Features:

- Volume Imbalance: Difference between buy and sell volumes
- VWAP Shift: Changes in volume-weighted average price over time windows

2. Data Normalization and Dimensionality Reduction

All features were normalized using MinMaxScaler to bring them to a comparable scale. Principal Component Analysis (PCA) was applied for dimensionality reduction, primarily to assist in better clustering and visualization.

3. Clustering Techniques

Multiple unsupervised learning techniques were employed to uncover natural structure in the data:

- K-Means Clustering: Optimal number of clusters determined via the elbow method. Evaluated using Silhouette Score and Davies-Bouldin Index.
- Gaussian Mixture Models (GMM): Provided probabilistic soft clustering. Useful in identifying overlapping regimes.
- HDBSCAN: Captured non-spherical clusters and labeled noise points. No need to predefine cluster count.

4. Regime Labeling and Analysis

Based on average values of volatility, spread, depth, and volume across clusters, each regime was interpreted and labeled. Some discovered regimes include:

- Volatile & Illiquid
- Stable & Liquid
- Trending & Volatile
- Mean-Reverting & Stable

These labels were derived by analyzing:

- Spread size and microprice dynamics (Liquidity)
- Standard deviation of returns (Volatility)
- Cumulative depth and volume imbalance (Market action)

5. Visualizations

To enhance interpretability:

- t-SNE and PCA were used to project clusters in 2D
- Line plots showed regime evolution over time
- Price overlaid plots validated alignment between regime shifts and market behavior

6. Regime Transition Analysis

A transition matrix was constructed to understand the likelihood of moving from one regime to another. Insights into regime predictability and cycle patterns were obtained, which could be beneficial for trading strategy development.

Conclusion

This unsupervised learning-based framework successfully segmented market behavior into meaningful regimes. The extracted features captured microstructure patterns well, and clustering techniques revealed actionable behavioral patterns.

Future work may involve:

- Integrating more trade features (e.g., aggressor side)
- Testing regime-aware trading strategies
- Using Hidden Markov Models for transition prediction