

CLO Loan-Level Liquidity Predictor

ML-Based Trade Cost Analysis for Leveraged Loans

Overview

Liquidity in the leveraged loan market is notoriously opaque. Unlike equities or bonds, leveraged loans trade over-the-counter with limited price transparency, creating significant challenges for trade execution and portfolio management.

This project addresses that challenge by building machine learning models that predict loan-level liquidity tiers and expected bid-ask spreads. The CLO market represents over \$1 trillion in assets under management, making accurate liquidity prediction critical for CLO managers balancing portfolio returns against trading costs.

Key Capabilities

- Liquidity Tier Classification (1-5 scale) - XGBoost classifier predicting relative loan liquidity
- Bid-Ask Spread Prediction (bps) - LightGBM regressor estimating transaction costs
- Feature-based Explainability - SHAP TreeExplainer for transparent, auditable predictions
- Interactive Dashboard - Streamlit app for real-time predictions and visualization

Liquidity Tier Definitions

Tier	Description	Typical Bid-Ask	Trading Volume
1	Most Liquid	< 30 bps	> \$20M/month
2	Liquid	30-50 bps	\$10-20M/month
3	Moderate	50-100 bps	\$5-10M/month
4	Less Liquid	100-150 bps	\$1-5M/month
5	Illiquid	> 150 bps	< \$1M/month

Model Performance

Liquidity Tier Classifier (XGBoost)

Multi-class classification model predicting loan liquidity tiers (1-5) based on loan characteristics, market conditions, and trading metrics.

Algorithm:	XGBoost Gradient Boosting
Validation:	TimeSeriesSplit 5-fold CV (prevents look-ahead bias)
Accuracy:	~99% on synthetic data <i>(Target: >70%)</i>
Output:	5-class probabilities with SHAP explanations

Trade Cost Predictor (LightGBM)

Regression model estimating expected bid-ask spreads in basis points, enabling pre-trade cost analysis and execution optimization.

Algorithm:	LightGBM Gradient Boosting
Validation:	5-fold Cross-Validation
MAE:	~12 bps <i>(Target: <30 bps)</i>
R-squared:	>0.85
Confidence Intervals:	95% CI via bootstrap resampling

Top Predictive Features

Rank	Feature	Category	Description
1	trading_volume_30d	Liquidity	30-day trading volume
2	bid_ask_spread	Liquidity	Current bid-ask spread
3	facility_size	Loan	Total facility amount
4	credit_rating_encoded	Loan	Ordinal credit rating (BB+ to CCC)
5	market_stress	Market	Composite stress indicator
6	dealer_quote_count	Liquidity	Number of dealers quoting
7	clo_ownership_pct	Ownership	CLO ownership percentage
8	vix_level	Market	VIX volatility index

Technical Architecture

Data Sources

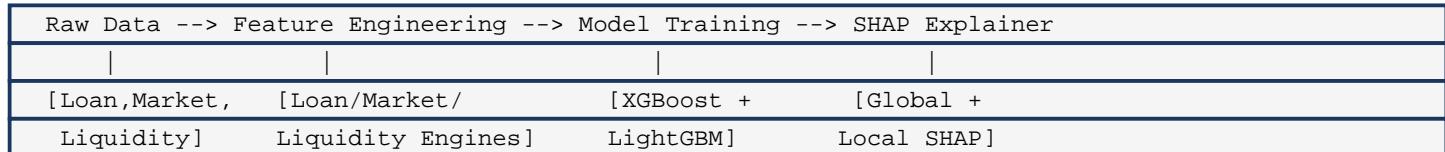
Source	Data Type	Usage
FRED API	Economic Indicators	VIX, HY/IG spreads, Fed funds, yield curve
SEC EDGAR	N-PORT Filings	CLO ownership concentration, fund holdings
Yahoo Finance	Market Data	VIX, S&P 500, credit ETFs (HYG, LQD)
Synthetic Generator	Training Data	Realistic loan data for model development

Feature Engineering Pipeline

30+ engineered features across three dimensions:

- Loan Features: facility_size_log, credit_rating_encoded, spread_z_score, time_to_maturity, covenant_lite
 - Market Features: vix_level, vix_percentile, hy_spread, ig_spread, yield_curve_slope, market_stress
 - Liquidity Features: volume_percentile, bid_ask_percentile, days_since_trade, dealer_coverage, clo_ownership_pct

Model Pipeline



Business Relevance to Octaura

This project directly addresses core challenges in the CLO and leveraged loan markets that Octaura aims to solve:

Trade Execution Quality

- Pre-trade cost estimation enables optimal execution timing
 - Confidence intervals quantify execution risk
 - Liquidity tier classification guides order sizing

Mark-to-Market Pricing

- Spread predictions support fair value estimation for illiquid positions
 - Feature importance reveals pricing drivers

Portfolio Analytics

- Portfolio-level liquidity scoring for risk management
 - Rebalancing optimization based on trade costs
 - Stress testing under different market conditions

Compliance & Transparency

- SHAP explanations provide audit trail for trading decisions
 - Model persistence ensures reproducibility

Technical Stack

Category	Technologies
Languages	Python 3.11+

ML Frameworks	XGBoost 2.0+, LightGBM 4.0+, scikit-learn
Explainability	SHAP (SHapley Additive exPlanations)
Data Processing	pandas, numpy
Visualization	Streamlit, matplotlib
Data Sources	FRED API, SEC EDGAR, Yahoo Finance

Production Readiness & Next Steps

- API Integration: RESTful API wrapper for model serving
- Real Data: Integration with live market data feeds
- Model Monitoring: Drift detection and automated retraining
- Scalability: Batch prediction pipeline for portfolio-level analysis