Quantitative Trading Strategy with Machine Learning

ML-Enhanced Algorithmic Trading System for NIFTY 50

Executive Summary: Project Overview

- End-to-end quantitative trading system for NIFTY 50 (5-min intraday)
- 1-year historical data (Oct 2021 - Oct 2022)
- 3-layer strategy: Trend signals → Regime filter → ML enhancement
- Demonstrates ML as probabilistic trade filter, not price predictor
- Combines 5/15 EMA crossover + HMM regimes + XGBoost classifier
- Goal: Improve risk-adjusted returns while reducing drawdowns

Executive Summary: Key Findings

- ✓ Regime filter reduces false signals by filtering sideways periods
- ✓ ML enhancement improves win rate and reduces drawdown
- ✓ Sharpe ratio improvement: [See results]
- ✓ Max drawdown reduction: [See results]
- ✓ Scalable architecture: Modular code for production deployment
- ✓ Comprehensive analysis: Outlier detection + statistical testing

Data Pipeline Architecture

- Step 1: Data Acquisition
- • NIFTY 50 Spot: 5-min OHLCV data
- • NIFTY Futures: Monthly contracts with rollover handling
- • Data source: NSE historical data (1 year)
- Step 2: Data Cleaning & Alignment
- • Handle missing values (forward fill + drop)
- • Remove outliers (IQR method)
- • Align timestamps across spot & futures

Feature Engineering: Technical Indicators

- Trend Features:
- • EMA(5) - Fast moving average
- • EMA(15) - Slow moving average
- • EMA Spread - Distance between fast & slow
- Volatility Features:
- • Rolling Volatility - 20-period log return std dev
- • High Vol Regime - Binary flag (above/below median)

Feature Engineering: Derived Features

• Return Features:

• • Log Returns - Price momentum

• • Futures Basis - (Futures - Spot) / Spot

• Options-Based Features (Black-Scholes):

• • Greeks: Delta, Gamma, Theta, Vega, Rho

• • IV Features: Average IV, IV Spread

• • PCR Ratios: Put-Call Ratio (OI & Volume based)

Regime Detection: Hidden Markov Model

- Model Architecture:
- • Gaussian HMM with 3 hidden states
- • Input: Log returns, volatility, EMA spread, vol regime
- • Training: 70% of data | Testing: 30%
- Regime Mapping (Dynamic based on returns):
- • State 1 → Uptrend (+1): Highest avg returns
- • State 0 → Sideways (0): Middle avg returns
- • State 2 → Downtrend (-1): Lowest avg returns

Regime Detection: Price with Regime Overlay

Regime Detection: Statistics by Regime

- Uptrend Regime: Higher average returns, moderate volatility, higher Sharpe ratio
- Downtrend Regime: Negative returns, high volatility, low/negative Sharpe ratio
- Sideways Regime: Near-zero returns, low volatility, limited opportunities

Trading Strategy: Baseline (5/15 EMA)

• Entry Signals:

• • LONG: EMA(5) crosses above EMA(15)

• • SHORT: EMA(5) crosses below EMA(15)

• Exit Signals: Opposite crossover occurs

• Characteristics: Simple rule-based system, whipsaw risk in sideways markets

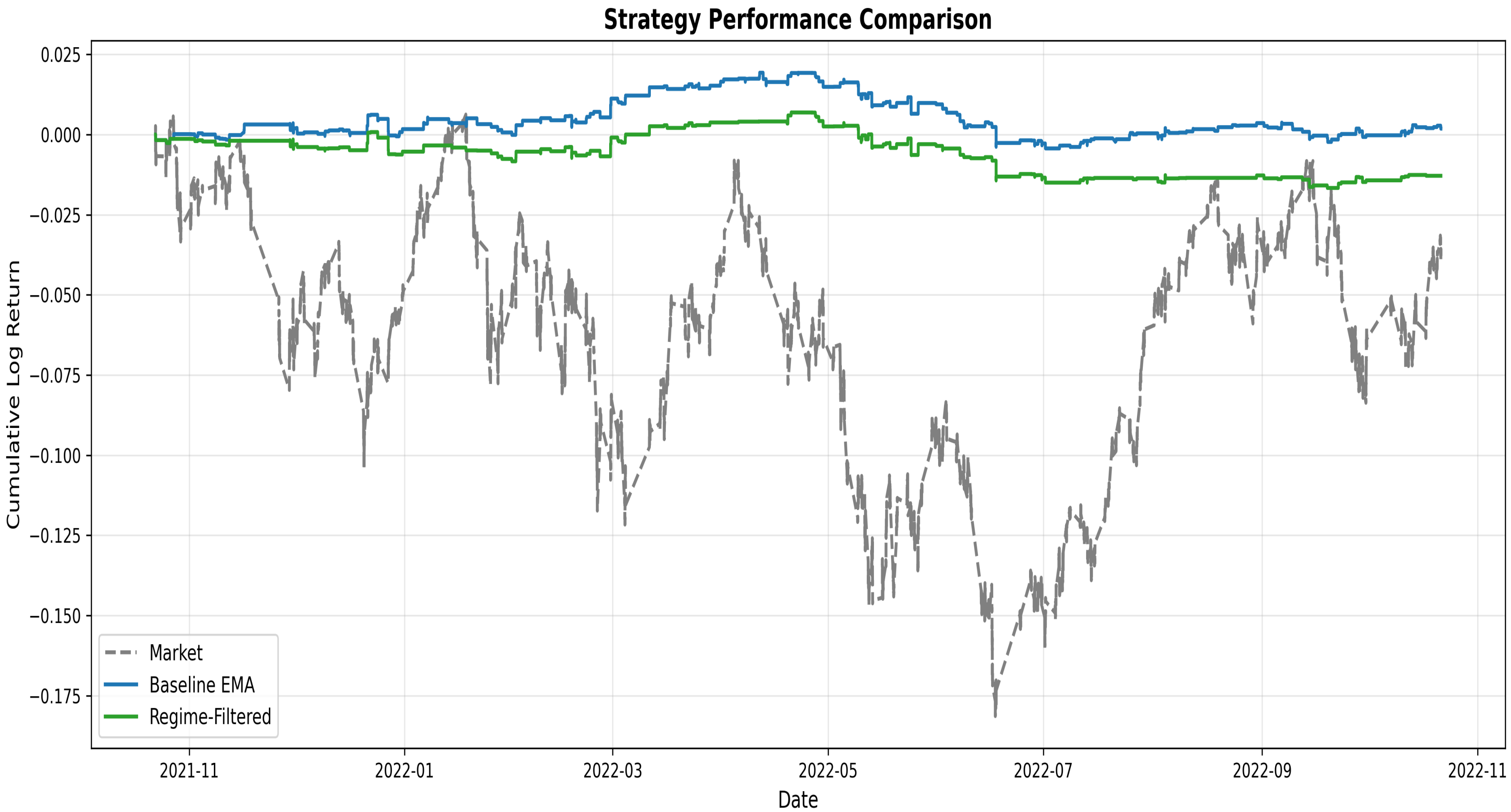
Trading Strategy: Regime-Filtered

- Entry Signals:
- • LONG: EMA(5) > EMA(15) AND Regime = Uptrend
- • SHORT: EMA(5) < EMA(15) AND Regime = Downtrend
- • NO trades in Sideways regime
- Benefits: Filters signals, trades with trend, improves risk-adjusted returns

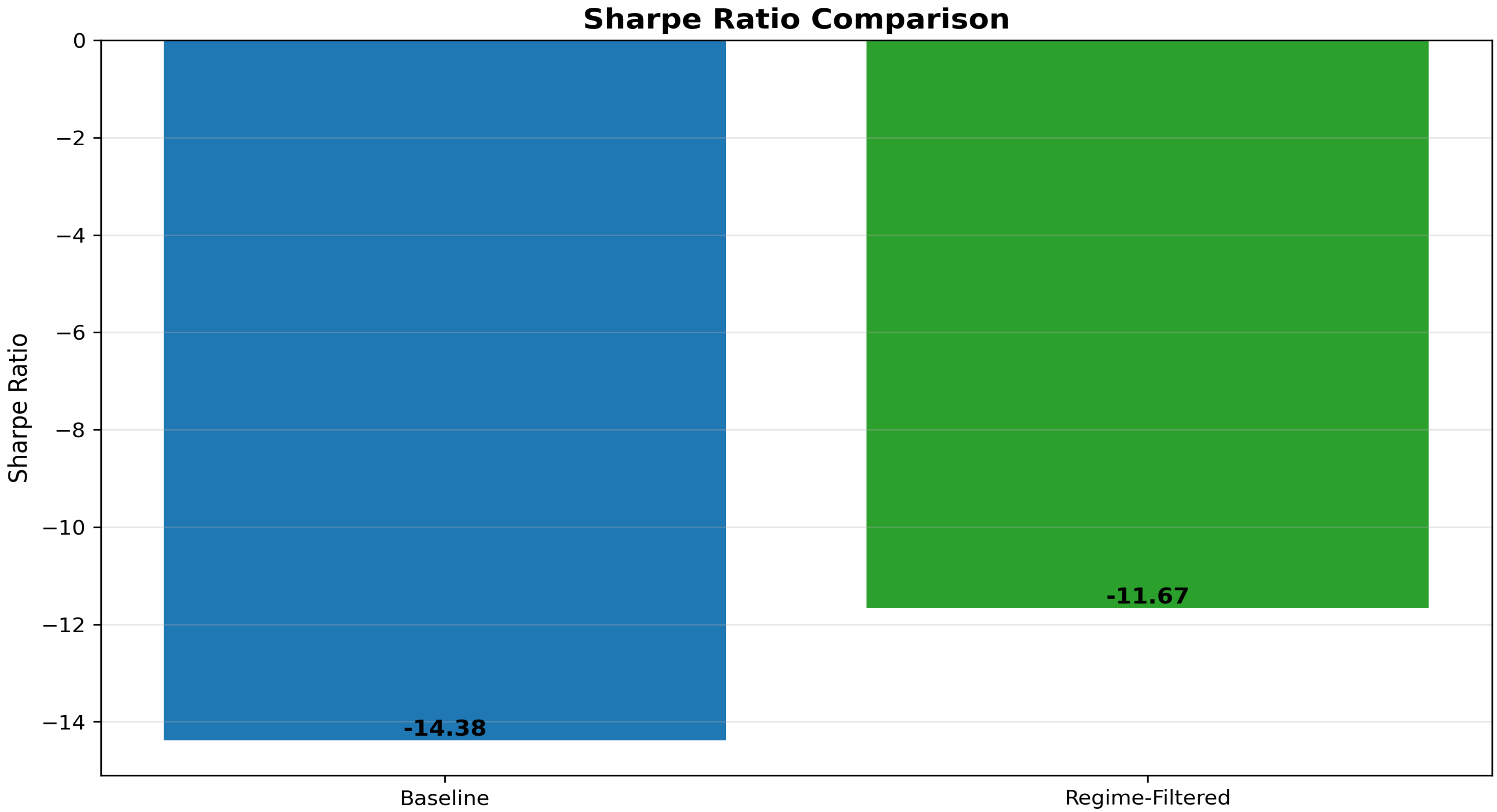
Backtesting Results: Performance Metrics

- Key Metrics: Total Return, Annual Return, Sharpe Ratio, Sortino Ratio
- Risk Metrics: Max Drawdown, Calmar Ratio, Win Rate, Profit Factor
- Evaluation: 70% training | 30% testing (time-series split)
- No look-ahead bias: Signals shifted 1 bar before calculating returns

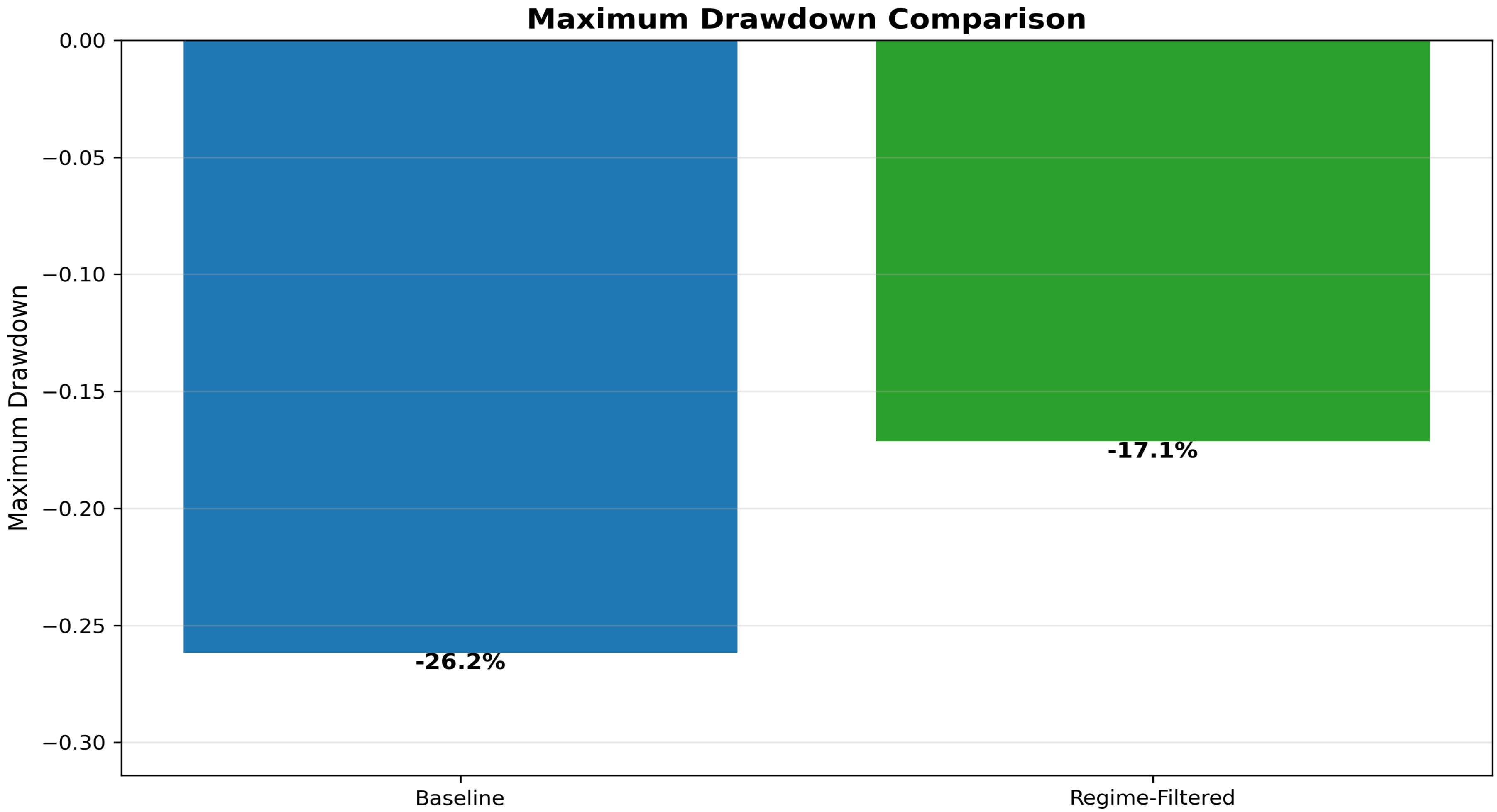
Strategy Comparison: Cumulative Returns



Strategy Comparison: Sharpe Ratio



Strategy Comparison: Maximum Drawdown



Machine Learning: Model Architecture

- Model 1: Logistic Regression - Linear classifier, interpretable baseline
- Model 2: XGBoost - Non-linear gradient boosting (100 trees, depth=5)
- Model 3: LSTM - Recurrent neural network (10-candle sequences, 64 units)
- Target: Binary classification (profitable/unprofitable next candle)

ML Models: Performance Comparison

XGBoost: Feature Importance Analysis

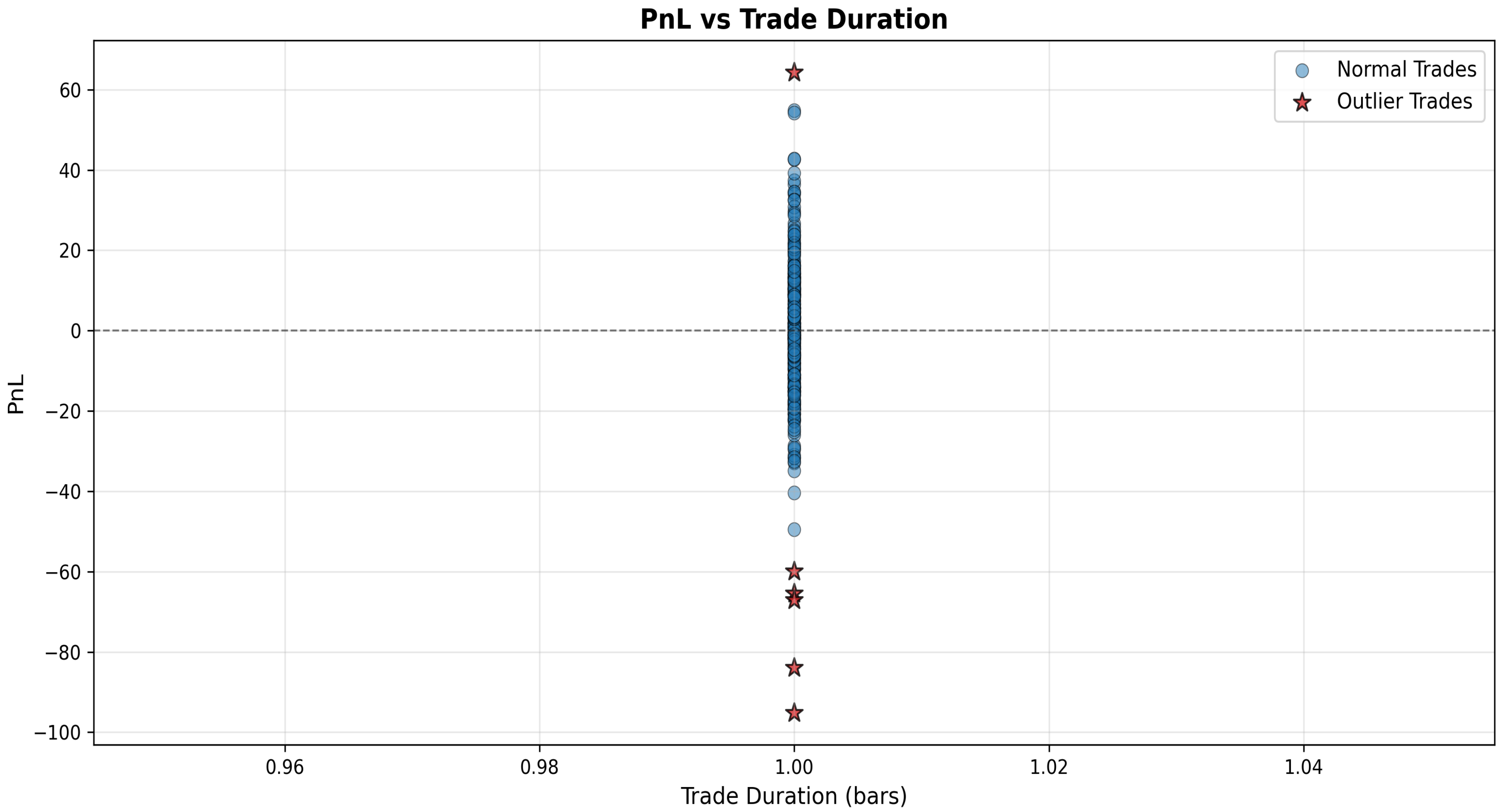
Trading Strategy: ML-Enhanced

- Strategy Logic:
- 1. Generate signals from 5/15 EMA + Regime filter
- 2. For each signal, get ML probability prediction
- 3. Only take trade if XGBoost confidence > 60%
- Benefits: Reduces false signals, improves win rate, best risk-adjusted returns

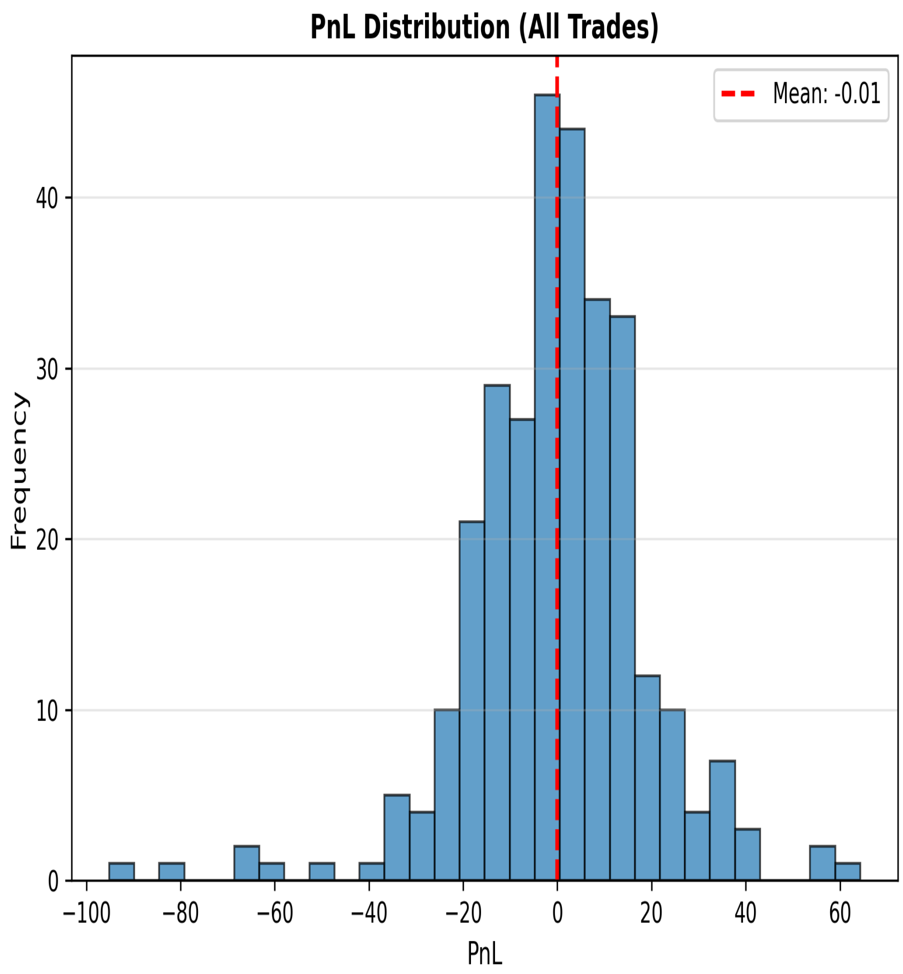
Outlier Analysis: Methodology

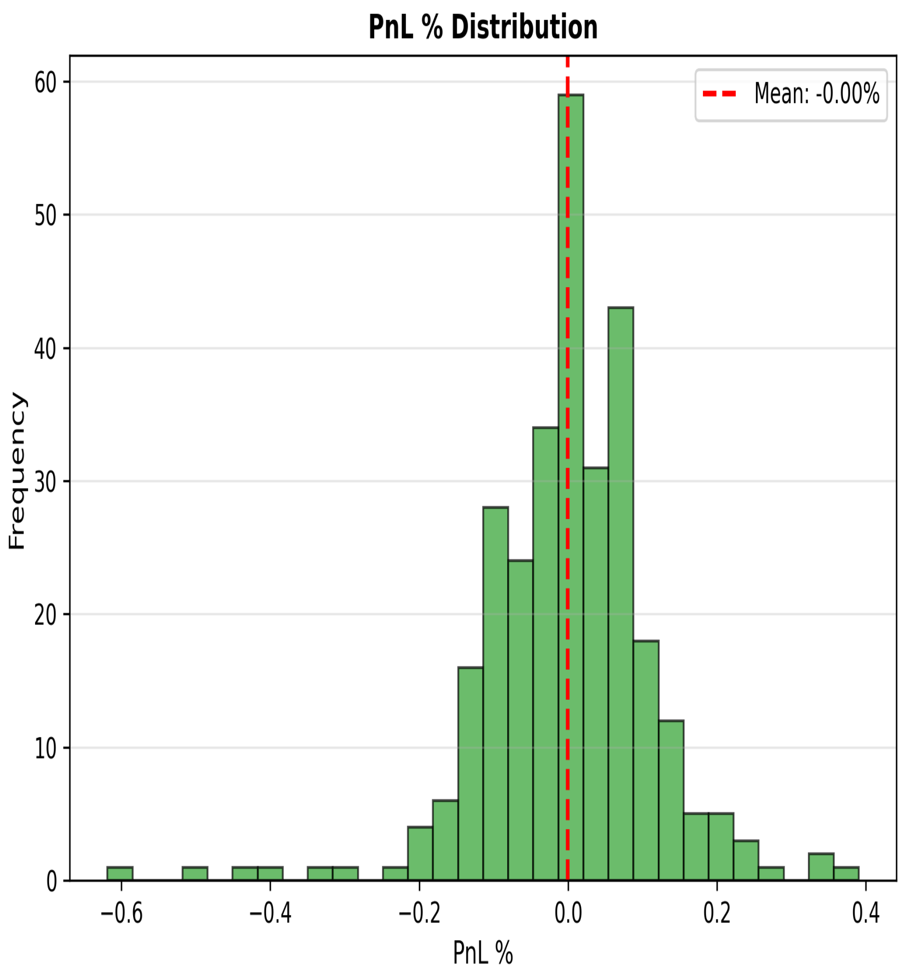
- Outlier Detection (Z-score > 3):
- • Extract all individual trades from signals
- • Calculate PnL and PnL% for each trade
- • Identify trades with |Z-score| > 3 as outliers
- Analysis: Compare outlier vs normal trades, statistical tests

Outlier Analysis: PnL vs Trade Duration

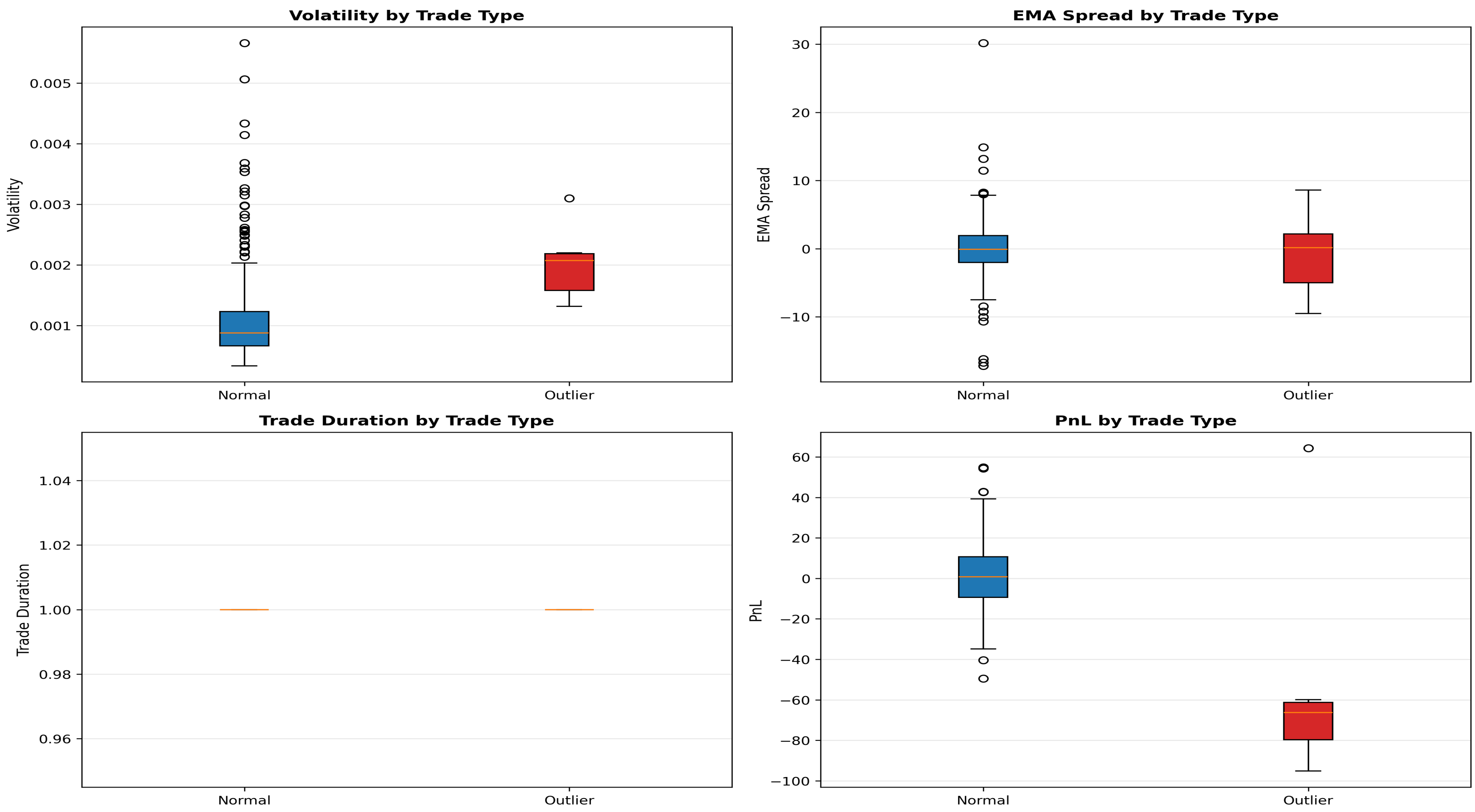


Outlier Analysis: PnL Distribution

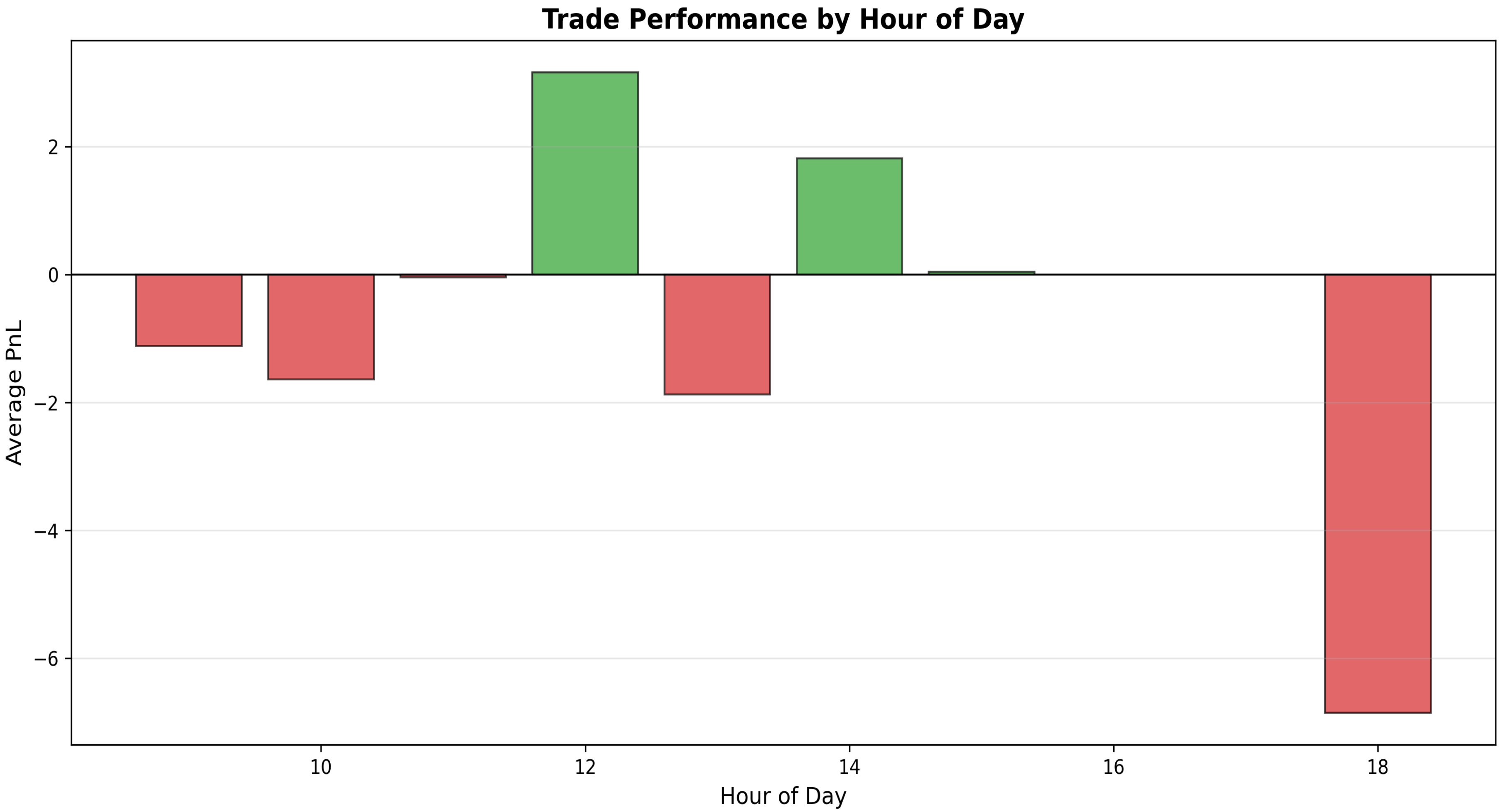




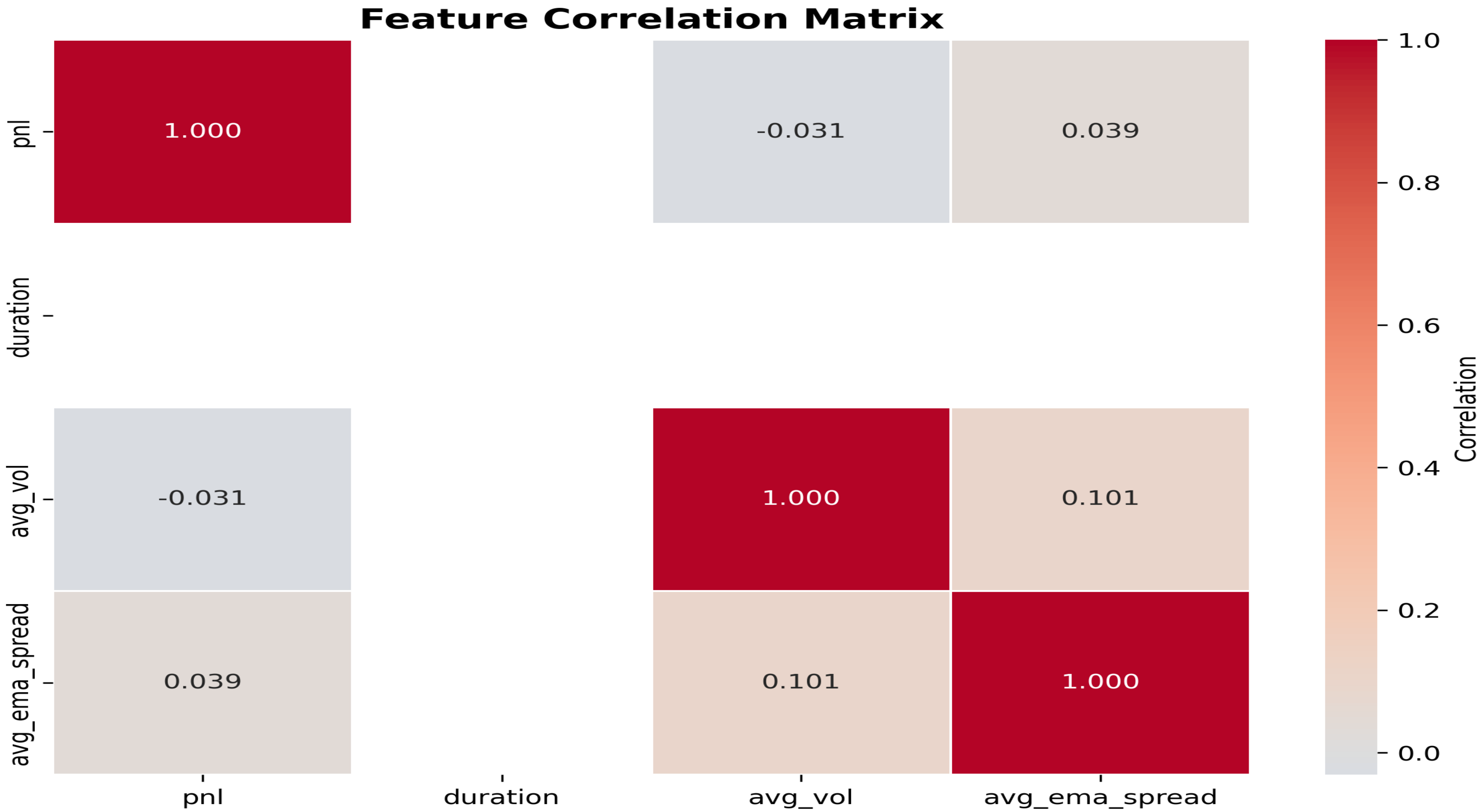
Outlier Analysis: Feature Distributions



Outlier Analysis: Time-of-Day Patterns



Outlier Analysis: Feature Correlations



High-Performance Analysis: Key Insights

• Outlier Characteristics:

• • High proportion are statistical outliers (|Z-score| > 3)

• • Significantly higher average PnL than normal trades

• • Tend to last longer than normal trades

• Pattern Recognition: Certain hours profitable, volatility predicts outcomes

Conclusions & Key Takeaways

- ✓ Hybrid strategy (EMA + Regime + ML) outperforms baseline
- ✓ ML as trade filter reduces noise and improves Sharpe ratio
- ✓ Regime detection filters sideways market periods effectively
- ✓ Feature engineering captures market microstructure
- ✓ Outlier analysis reveals tradeable patterns
- ✓ Modular, scalable architecture ready for production

Recommendations for Future Work

- Short-term (0-3 months):
- • Deploy on live data with broker API integration
- • Add position sizing and risk management rules
- Medium-term (3-12 months): Extend to multi-asset, advanced models
- Long-term: Portfolio optimization and real-time regime identification

Technical Implementation Highlights

- Code Quality: Modular architecture, type hints, docstrings, no look-ahead bias
- ML Validation: Time-series cross-validation, StandardScaler normalization
- Reproducibility: Automated workflows, CSV results, publication-quality visualizations

Thank You

Lakshy Mittal | Quantitative Trading ML Assignment | January 2025