Enhanced ML-Based Trading Bot Framework (Hybrid Rigid + Adaptive + ML)

1. Core Analyzer Tool: Parameters & Calculations

Base Parameters

- **Open Interest (OI)**
- **Implied Volatility (IV)**
- **Volume**
- **Price**
- **Bid-Ask Quantity** *(NEW)*
- **Bid-Ask Spread** *(NEW)*

Strike Selection & Calculations

- **Strikes considered:** 16 total = 5 ITM + 1 ATM + 10 OTM
- **Totals:** Σ(OI of 16 strikes)
- **Difference:** Current Totals – Baseline Totals
- **Growth Rate:** (Difference ÷ Totals) × 100
- **Baseline Totals:** First fetched point
- **CE & PE:** Calculated separately, then compared

NEW: Bid-Ask Integration

- **Bid-Ask Imbalance:** (Bid Qty - Ask Qty) / (Bid Qty + Ask Qty) per strike
- **Weighted Bid-Ask Score:** Σ(Imbalance × OI_weight) across 16 strikes
- **Spread Indicator:** Average spread across ATM ±2 strikes
- **Aggression Factor:** Volume executed at bid vs ask (when available)

Adaptive Upgrades

- **Dynamic dominance:** Rolling Z-scores → adapts to volatility regime
- **Volume delta filter:** OI buildup counts only if ΔOI backed by ΔVolume
- **Bid-Ask validation:** Significant OI changes must align with bid-ask imbalances

2. OI Participants & Interpretation

Participants

- Call Buyers/Writers
- Put Buyers/Writers

Positions

- Long, Short, Unwinding, Covering

Enhanced Market Style Table

Price	OI	IV	Bid-Ask	State	ML Confidence
↑ Rising	↑ Rising	↑ Rising	Call Heavy	Aggressive Long Build	ML Score %
↑ Rising	↑ Rising	↓ Falling	Balanced	Calm Long Build Up	ML Score %
↑ Rising	↓ Falling	↑ Rising	Put Heavy	Short Cover (Panic)	ML Score %
↑ Rising	↓ Falling	↓ Falling	Call Light	Smooth Short Cover	ML Score %
↓ Falling	↑ Rising	↑ Rising	Put Heavy	Aggressive Short Build	ML Score %
↓ Falling	↑ Rising	↓ Falling	Balanced	Calm Short Build Up	ML Score %
↓ Falling	↓ Falling	↑ Rising	Call Heavy	Long Exit (Panic)	ML Score %
↓ Falling	↓ Falling	↓ Falling	Put Light	Smooth Long Exit	ML Score %

NEW: ML Confidence Scoring

- **Input Features:** [OI_change, IV_change, Volume_ratio, BidAsk_imbalance, Spread_norm]
- **Model:** Lightweight Random Forest or XGBoost
- **Output:** Confidence probability (0-100%) for each market style
- **Training:** Historical patterns with labeled outcomes
- **Retraining:** Every 1000 data points or weekly

3. Enhanced Weighted Score & Dominance

Updated Formula

Weighted Score = (0.4 × %OI) + (0.25 × %IV) + (0.15 × %Volume) + (0.2 × %BidAsk_Imbalance)

Dominance Rules

- **CE Score > PE Score:** Call dominance
- **PE Score > CE Score:** Put dominance

Dynamic Thresholds (Adaptive)

- **Strong dominance:** CE-PE deviation > 1.2σ (rolling Z-score)
- **Mild dominance:** CE-PE deviation > 0.8σ
- **NEW: Expiry Day Filter:** Ignore OI unwinding on T+0, T+1 expiry days

NEW: Expiry Day Adjustments

- **Days to Expiry < 2:** Reduce OI weight by 50%, increase Volume weight by 25%
- **Expiry Day (T+0):** Focus on Volume and Bid-Ask, ignore OI changes
- **Post-Expiry Reset:** Recalibrate baseline totals for next series

4. Advanced Market Style Identifier

Core Parameters

- Enhanced Weighted Score (CE vs PE)
- Rolling 15-min Price Delta
- IV Expansion with Skew
- **NEW: Volatility Skew Tracker**

NEW: Volatility Skew & Smile Dynamics

- **Skew Calculation:** IV_call - IV_put at same strike distance from ATM
- **Smile Curvature:** Track IV across strike range (parabolic fit)
- **Skew Signal:**
  - Positive skew + Call dominance = Strong bullish
  - Negative skew + Put dominance = Strong bearish
  - High smile curvature = High uncertainty/volatility

Enhanced Rule Matrix

Condition	Style	Rationale	Skew Factor
Strong one-sided OI + dominance >1.2σ + IV expansion + Aligned skew	**Trending**	Smart money + vol expansion + skew alignment	High confidence
Both-side buildup + dominance 0.8σ + IV stable ±2% + Flat skew	**Sideways**	Balanced participants + stable volatility	Medium confidence

Condition	Style	Rationale	Skew Factor
CE OI ↑ + Price flat + IV expansion + Negative skew	**Volatile/Trap**	Trap attempts with fear skew	Low confidence
High smile curvature + Rapid skew changes	**Uncertainty**	Market confusion/event driven	Variable

5. ML-Enhanced Market Trend Rules

1. Trend Confirmation

- **Traditional:** Dominance ≥1.2σ + Price moves ≥20-29 pts + IV rising
- **ML Enhancement:** Confidence score ≥75% + Feature alignment score >0.8

2. Volatile/Trap Detection

- **Traditional Trap Flag:** Dominance flips 2x in 3 windows + IV expansion ≥3-5%
- **ML Enhancement:** Anomaly detection flags unusual pattern combinations
- **Promotion to Trend:** Price follow-through ≥20-29 pts + ML confidence >80%

3. Sideways Identification

- **Traditional:** Dominance 0.8σ for 2 bars + IV flat ±3% + Price range <15-20 pts
- **ML Enhancement:** Sideways pattern recognition with 70%+ confidence

NEW: Real-Time Anomaly Detection

- **Statistical Anomaly:** Z-score >2.5 for any parameter combination
- **Pattern Anomaly:** ML model flags deviation from learned patterns
- **Volume Spike:** Unusual volume without corresponding OI/Price movement
- **Response:** Flag for manual review or reduce automated signal weight

6. Enhanced Abnormal Days Filter

Base Rules (Enhanced)

- **IV Spike:** Today's IV ≥25-30% higher vs yesterday → Caution mode
- **Gap Check:** Open gap ≥350-400 pts → Caution mode
- **NEW: Bid-Ask Spread Alert:** Average spread >2x normal → Liquidity concern

NEW: Volatility Regime Classifier

- **Low Vol Regime:** ATM IV <15% → Higher sideways probability
- **Mid Vol Regime:** ATM IV 15-25% → Normal operation
- **High Vol Regime:** ATM IV >25% → Higher trap probability, reduce position sizes

Event-Driven Adjustments

- **Pre-Event:** 2 days before known events → Reduce signal sensitivity
- **Post-Event:** 1 day after events → Gradual sensitivity restoration
- **Intraday Events:** Real-time news sentiment integration (if available)

7. CRITICAL: Data Labeling Methodology

Overview

This section is mandatory for ML implementation. Labels define the target outcomes that ML models learn to predict.

7.1 Rule-Based Initial Labeling

Step 1: Extract Key Indicators Per Data Point

For each 10-second snapshot, calculate:

- **Price Direction:** Compare current price vs. previous 5 data points (Rising/Falling/Flat)
- **OI Direction:** Compare current OI vs. baseline (Rising/Falling/Stable)
- **IV Direction:** Compare current IV vs. previous 10 data points (Rising/Falling/Stable)
- **Bid-Ask State:** Classify as Call Heavy/Put Heavy/Balanced based on imbalance ratios

Step 2: Apply Rule Matrix for Market Style Labels

Using the Enhanced Market Style Table, assign labels:

```
IF Price=Rising AND OI=Rising AND IV=Rising AND BidAsk=Call_Heavy:
    Label = "Aggressive_Long_Build"

IF Price=Rising AND OI=Rising AND IV=Falling AND BidAsk=Balanced:
    Label = "Calm_Long_Build"

IF Price=Falling AND OI=Rising AND IV=Rising AND BidAsk=Put_Heavy:
    Label = "Aggressive_Short_Build"

... [continue for all 8 combinations from table]
```

Step 3: Assign Participant and Position Labels

- **Participant Labels:** Call_Buyers, Call_Writers, Put_Buyers, Put_Writers
- **Position Labels:** Long_Entry, Short_Entry, Long_Exit, Short_Exit, Covering, Unwinding

Logic:

```
IF Market_Style contains "Long_Build":
    Participant = "Call_Buyers"
    Position = "Long_Entry"

IF Market_Style contains "Short_Build":
    Participant = "Put_Buyers"
    Position = "Short_Entry"

IF Market_Style contains "Cover":
    Position = "Covering"

IF Market_Style contains "Exit":
    Position = "Long_Exit" or "Short_Exit" (based on direction)
```

7.2 Outcome-Based Validation Labeling

Future Price Movement Validation

For each data point, look forward 15 minutes and 30 minutes:

- **Strong Trend Validation:** Price moves >25 points in predicted direction
- **Weak Trend Validation:** Price moves 15-25 points in predicted direction
- **Sideways Validation:** Price moves <15 points in any direction
- **False Signal:** Price moves >20 points in opposite direction

Confidence Adjustment Based on Outcomes

```
IF Rule_Label = "Aggressive_Long_Build" AND Future_Price_Move > +25pts:
    Final_Label = "Aggressive_Long_Build"
    Confidence = 90%

IF Rule_Label = "Aggressive_Long_Build" AND Future_Price_Move < +15pts:
    Final_Label = "False_Signal"
    Confidence = 20%
```

7.3 Label Encoding for ML Models

Primary Target Labels (Market Style)

- 0: Sideways
- 1: Trending_Up
- 2: Trending_Down
- 3: Volatile_Trap
- 4: Uncertainty

Secondary Target Labels (Participant Type)

- 0: Call_Buyers
- 1: Call_Writers
- 2: Put_Buyers
- 3: Put_Writers

Tertiary Target Labels (Position Type)

- 0: Long_Entry
- 1: Short_Entry
- 2: Long_Exit
- 3: Short_Exit
- 4: Covering
- 5: Unwinding

7.4 Labeling Implementation Steps

Phase 1: Historical Rule-Based Labeling

1. **Load historical option chain data**
2. **Calculate all features** (OI changes, IV changes, price changes, bid-ask imbalances)
3. **Apply rule matrix** to assign initial labels
4. **Export labeled dataset** with features and rule-based labels

Phase 2: Outcome Validation

1. **For each historical data point, calculate future outcomes**
2. **Compare rule predictions with actual market movements**
3. **Adjust labels and confidence scores** based on validation results
4. **Create validated training dataset**

Phase 3: Continuous Learning Setup

1. **Set up pipeline to label new data automatically** using rules
2. **Implement outcome tracking** for continuous validation
3. **Update model training data** with new validated examples
4. **Retrain models weekly** or every 2000 new samples

7.5 Quality Control for Labels

Data Quality Checks

- **Completeness Check:** Ensure no missing values in key fields
- **Consistency Check:** Verify rule logic produces consistent labels
- **Outcome Alignment:** Minimum 60% accuracy between rule predictions and outcomes
- **Balance Check:** Ensure reasonable distribution across all label categories

Label Validation Process

1. **Manual Review:** Sample 100 random labeled examples weekly
2. **Statistical Validation:** Track label distribution and outcome accuracy
3. **Edge Case Handling:** Define specific rules for unusual market conditions
4. **Feedback Loop:** Use model prediction errors to improve labeling rules

This comprehensive labeling methodology ensures ML models can learn meaningful patterns while maintaining consistency with the trading framework's logic.

8. ML Model Architecture

Feature Engineering

Input feature vector (per 10-second snapshot):

- OI changes (CE & PE)
- IV means and skew
- Volume ratios
- Price changes (1min & 5min windows)
- Bid-ask imbalances
- Spread metrics
- Days to expiry
- Volatility regime classification

Model Types

1. **Primary Model:** XGBoost Classifier
    - Target: Market style (0=sideways, 1=trending_up, 2=trending_down, 3=volatile)
    - Output: Probability distribution across styles
2. **Anomaly Detector:** Isolation Forest
    - Input: Same feature vector
    - Output: Anomaly score (0-1)
3. **Confidence Estimator:** Random Forest Regressor
    - Input: Feature vector + primary model probabilities
    - Output: Confidence score (0-100%)

Training Strategy

- **Data Window:** 3 months historical data
- **Retraining:** Weekly or every 2000 new samples
- **Validation:** Walk-forward validation with 80-20 split
- **Labels:** Use comprehensive labeling methodology from Section 7

9. Implementation Pipeline

Data Flow

1. **Fetch** → Upstox API every 10 seconds
2. **Process** → Calculate all base features + ML features
3. **Predict** → Run through ML models
4. **Validate** → Apply filters and anomaly checks
5. **Signal** → Generate final trading signal with confidence

Real-Time Processing Architecture

- **Anomaly check** runs first as safety filter
- **Main ML prediction** generates market style probabilities
- **Confidence scoring** provides reliability measure
- **Traditional filters** apply final validation layers
- **Signal generation** combines all components with appropriate weights

Key Advantages of This Enhanced Framework

1. **Maintains original structure** while adding ML intelligence
2. **Bid-ask integration** provides microstructure insights
3. **Expiry day handling** prevents false signals
4. **Volatility skew** adds sophisticated options pricing insights
5. **ML confidence scoring** adapts to changing market conditions
6. **Real-time anomaly detection** protects against unusual market behaviors
7. **Simple implementation** with modular components
8. **Comprehensive labeling methodology** ensures reliable ML training

This framework preserves your core logic while adding sophisticated ML enhancements that make it more robust and market-worthy without overcomplicating the base structure.