1. Overview

To enhance the predictive performance of our model and capture meaningful market dynamics, we designed a comprehensive feature-engineering pipeline. This step transforms the raw Hull Tactical dataset—mainly consisting of M*, V*, and macroeconomic columns—into a more informative representation that reflects trend, volatility, momentum, and market regimes. All engineered features were created without introducing look-ahead bias.

2. Lag-Based Momentum Features

We generated lag features for key return-related variables, including `forward_returns`, `market_forward_excess_returns`, and `risk_free_rate`.

- Lags used: **1, 2, 3, 5, 10 days**
- Purpose: capture short-term momentum, autocorrelation, and mean-reversion signals

These lagged features help the model understand how recent price movements influence future returns.

3. Rolling Statistics (MA, Volatility, Momentum)

To capture medium-term market structure, we computed several rolling-window statistics over **5, 10, 20, and 60 days**:

- **Rolling mean** (`ret_mean_*d`) → measures local trend strength
- **Rolling volatility** (`ret_std_*d`) → captures risk and uncertainty
- **Rolling cumulative momentum** (`ret_mom_*d`) → aggregates directional persistence

These are widely used in quantitative trading strategies because they encode trend-following and risk-based signals.

4. Volatility Aggregation from V* Columns

The dataset provides dozens of volatility-related predictors (V*).
To summarize their combined information, we computed:

- `vol_mean` – average volatility signal
- `vol_std` – cross-volatility dispersion
- `vol_max` – maximum volatility across V* columns

This provides a stable volatility proxy rather than relying on a single noisy V* feature.

5. Drawdown Features

Using a synthetic cumulative price index:

*price_index=∏(1+rt)price_index = \prod (1 + r_t)*price_index=∏(1+rt)

we computed:

- **Cumulative max price** (`cum_max_price`)
- **Drawdown** (`drawdown = price_index / cum_max_price - 1`)

This measures market stress and investors' risk aversion.
Large drawdowns often lead to strong rebounds or trend continuation—both predictive signals.

6. Market Regime Indicator

We constructed a simple **bull/bear regime flag**:

*regime=1if 60-day mean return>0regime = 1 \quad \text{if } \text{60-day mean return} > 0*regime=1if 60-day mean return>0

This binary feature helps the model behave differently in uptrending vs. downtrending environments.

7. Interaction Features

To capture potential non-linear effects, we created:

- `M_interaction = M0 * M1` (first two major M* macro features)

Such cross-features often help tree-based models detect higher-order structure.

8. Final Feature Set

After removing ID columns, target variables, and intermediate calculation fields,
the final dataset contained:

- **127 engineered features**, including
  lagged returns, moving averages, volatility, drawdown, regime, and interaction signals.

This enriched feature set significantly improved model performance, especially in LightGBM and XGBoost, as shown in our feature-importance and backtesting results.