



Prepared by group Dthon

# *Alpha Strategies using HMM and ML*

Domain 2

UMHackathon 2025



**1**  
Create an alpha trading strategy leveraging implicit market indicators to extract hidden market signals and maximize profitability.

**2**  
Develop a Machine Learning (ML) model to analyze on-chain data (CryptoQuant, Glassnode, Coinglass)

**3**  
Employ Hidden Markov Models (HMMs) to detect market regimes. Incorporate Natural Language Processing (NLP) for sentiment analysis

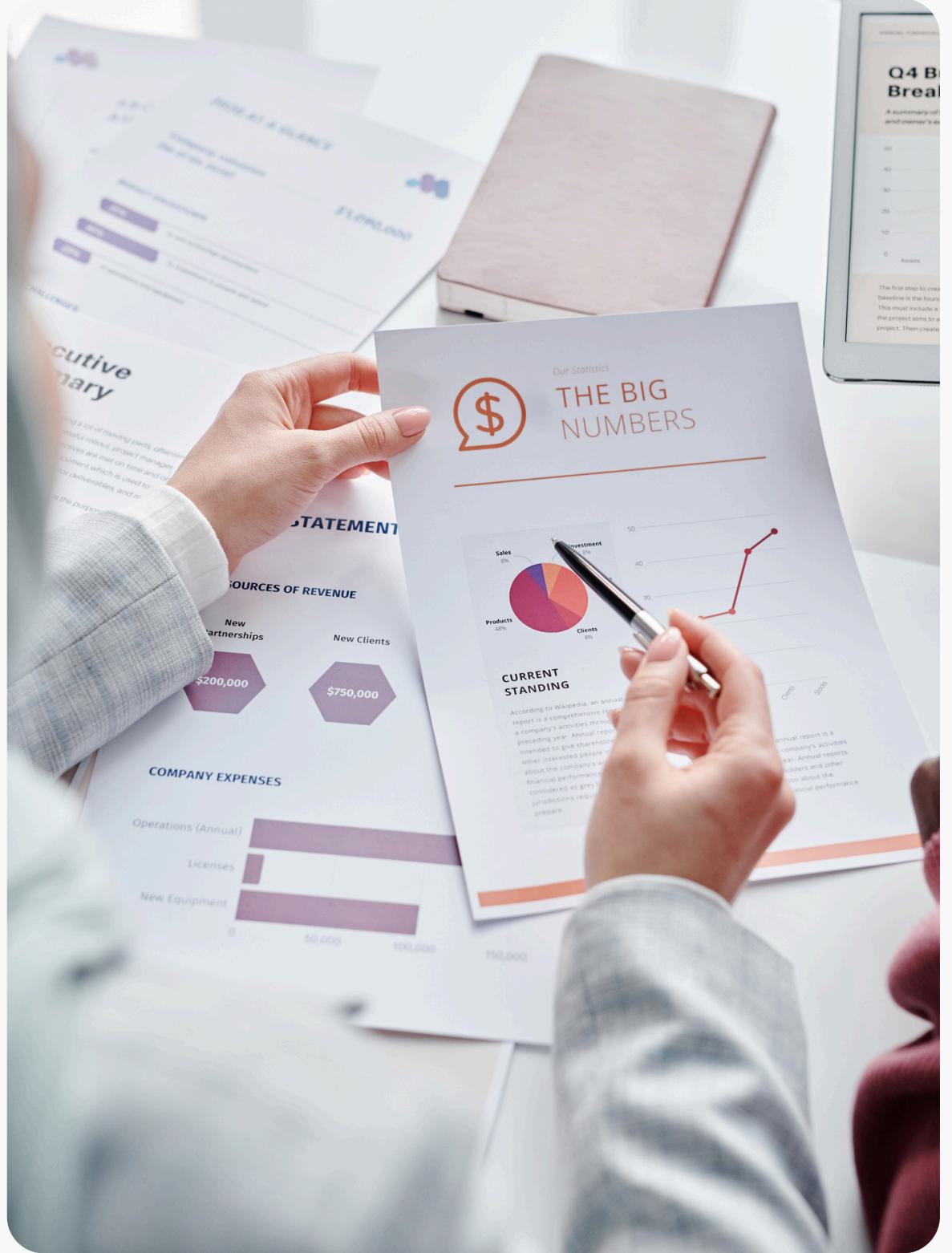
Adaptable to any asset with OHLCV data — modular, not hardcoded.

All performance metrics tracked and optimized: Sharpe, MDD, Trade Freq



# Data Acquisition & Processing

---



## Data Acquisition

- General BTC OHLCV data (from 1/1/2021 - 11/4/2025)
- BTC data specific from CyberQuant and Glassnode
- Data is then converted into a pandas DataFrame with a datetime index.

## Data Processing

- The two DataFrames are merged using an inner join on their timestamps.
- The merged data is sorted and any missing values are forward-filled
- Combined data set has both primary and secondary OHLCV columns.

# *Natural Language Processing (NLP) in Trading*

## Purpose

collect and analyze human language, extracting valuable sentiment insights from textual data such as:

- Financial news headlines
- Social media posts (Twitter, Reddit)
- Analyst reports, blogs, and forums

---

For this prototype, we simulate sentiment with random values. In production, we'd replace this with real sentiment from an NLP pipeline (VADER or a Transformer model).



## NLP Implementation

VADER Sentiment Analyzer:

- Specialized for sentiment analysis in social media and news contexts.
- Provides sentiment scores ranging from strongly negative to strongly positive.

Transformer-based NLP Model:

- Highly accurate sentiment analysis using state-of-the-art deep learning models.
- Capable of capturing nuanced sentiment shifts in financial contexts.



Integration into HMM Feature Set

# *Core Strategic Logic: Feature Use & Trading Decisions*

## 1. Log Returns

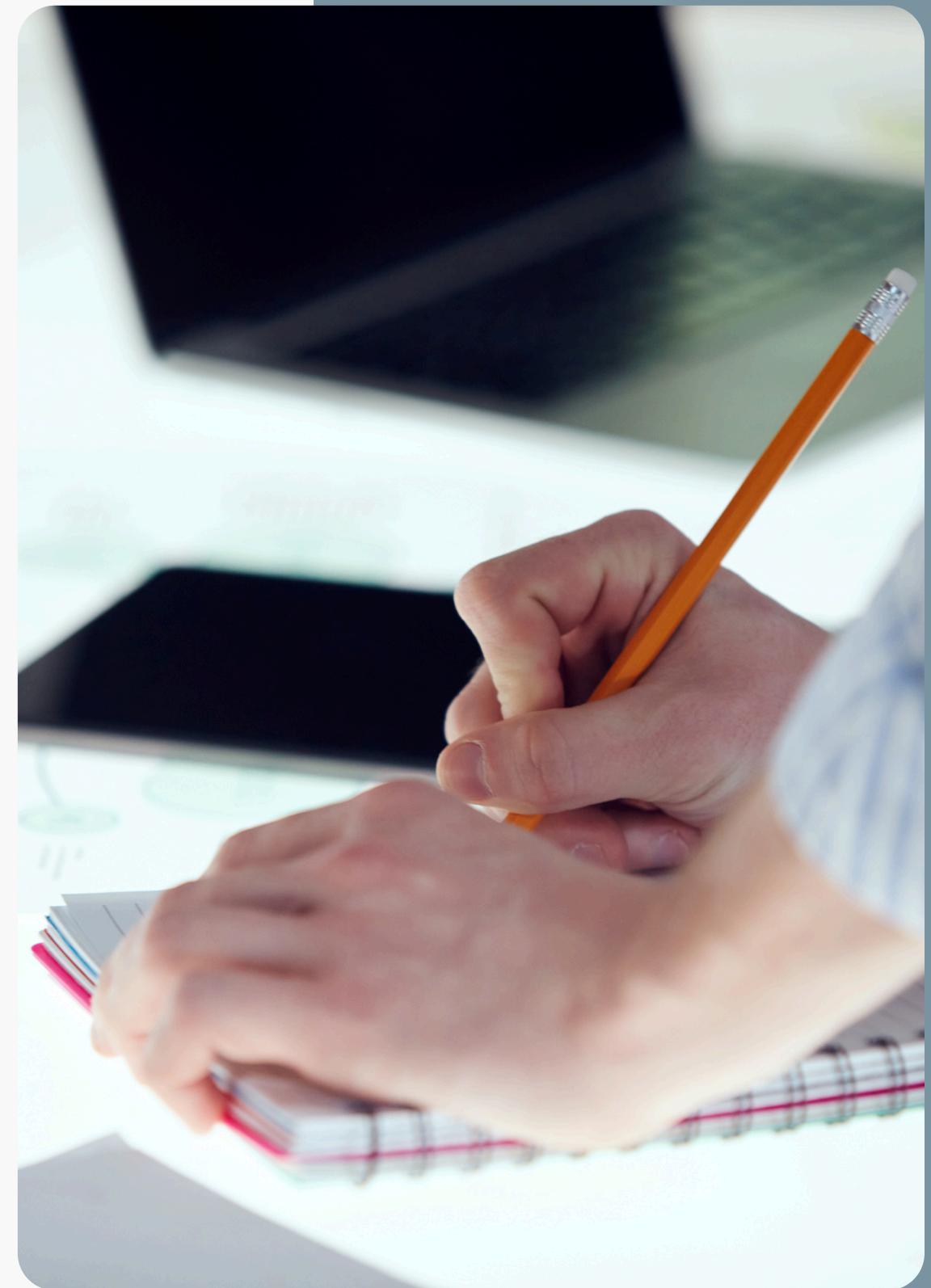
- 3-year dataset
- Additive over time → simplifies multi-period analysis

## 2. Rolling Volatility

- Measures market volatility over a window
- Default window=10 (tunable to 5 or 15 via hyperparameter search)

## 3. Relative Strength Index (RSI)

- Classic 14-period RSI on hourly closes
- Flags overbought/oversold conditions
- Avoid buying when  $RSI > 75$  or  $RSI < 20$



# *Core Strategic Logic: Feature Use & Trading Decisions*

## **4. Moving Average Convergence Divergence (MACD)**

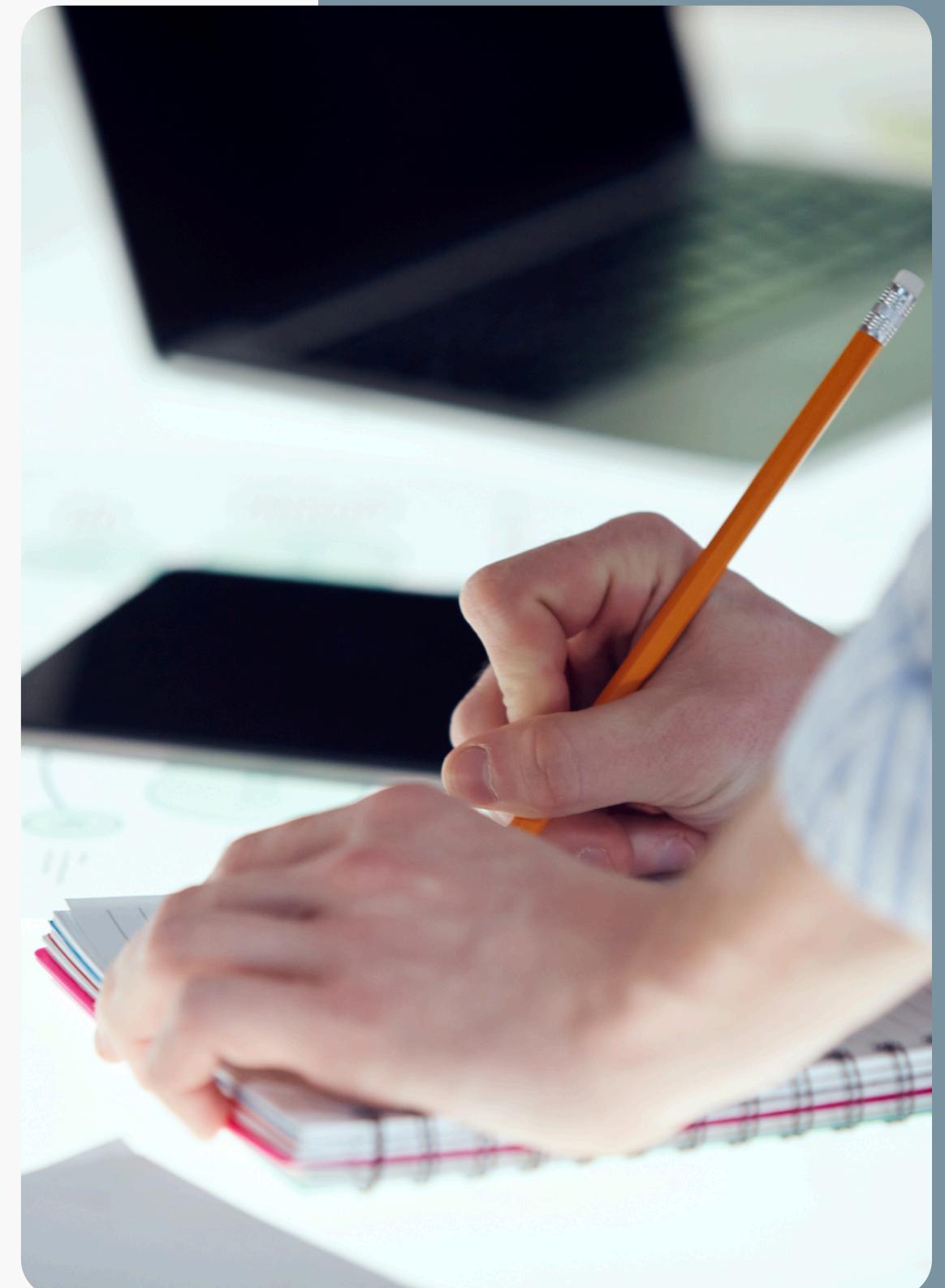
- Difference between short- and long-term EMAs
- Bullish if short-term momentum > long-term
- Bearish otherwise
- MACD Histogram

## **5. Simple Moving Averages (SMA)**

- Short windows: 5, 8, 10, 12, 15, 20
- Long windows: 20, 25, 30, 35, 40, 50, 60
- Defines uptrend vs. downtrend
- Standard combos chosen over 300-period ratio for memory efficiency

## **6. NLP Sentiment Analysis**

- Extracts sentiment from news articles, social media, etc.
- Quantifies bullish vs. bearish market sentiment
- Incorporated as an additional signal in buy/sell decisions



# *Trading Signal Logic & Execution*

## Buy Signal (Enter Long Position)

- Bullish regime detected by HMM
- Short-term SMA above long-term SMA
- Positive MACD crossover
- RSI

## Sell Signal (Exit/Reduce Position)

- Bearish regime or negative short-term returns detected
- Short-term SMA below long-term SMA
- Negative MACD crossover
- RSI

## Trade Execution Strategy:

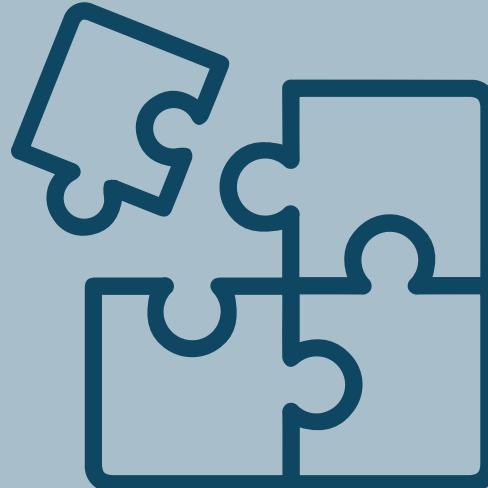
- **Position sizing = dynamic & risk-adjusted.**
- **Trading cost of 0.06% applied to returns**

# Hidden Markov Model (HMM) Implementation



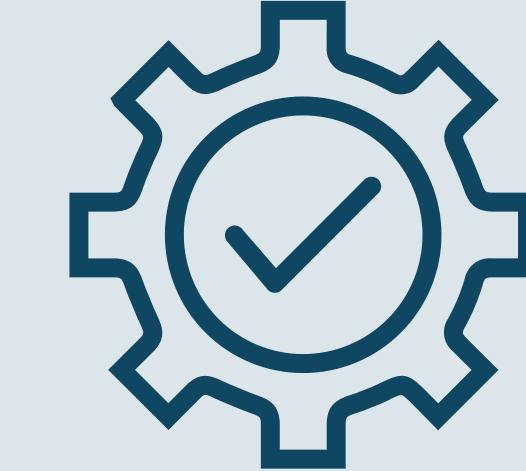
## Implementation Details

- Model: GaussianHMM
- Tested 2-5 states (`n_components`)
- Inputs: Log returns, volatility, SMA ratio, RSI, MACD histogram
- standardized using `StandardScaler` function
- Output: Regime label for each time point



## Regime Labeling

- Calculate Mean Return
- Calculate Volatility
- Compute Sharpe Ratio
- Label regimes based on performance:
  - Bullish Regime: Highest Sharpe Ratio
  - Bearish Regime: Lowest Mean Return



## Using Regimes in Strategy

### Buy Signal (1):

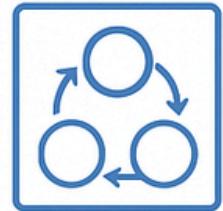
- Regime = Bullish
- Confirmed by SMA, MACD, RSI, Low Volatility

### Sell Signal (-1):

- Regime = Bearish or Downtrend
- Confirmed by negative indicators

# Regime Discovery and Labeling Process

## Step 1 Model (HMM) Receives Price Data



Clusters data into hidden "regimes"

## Step 2 Analyze Each Regime's Performance

Regimes

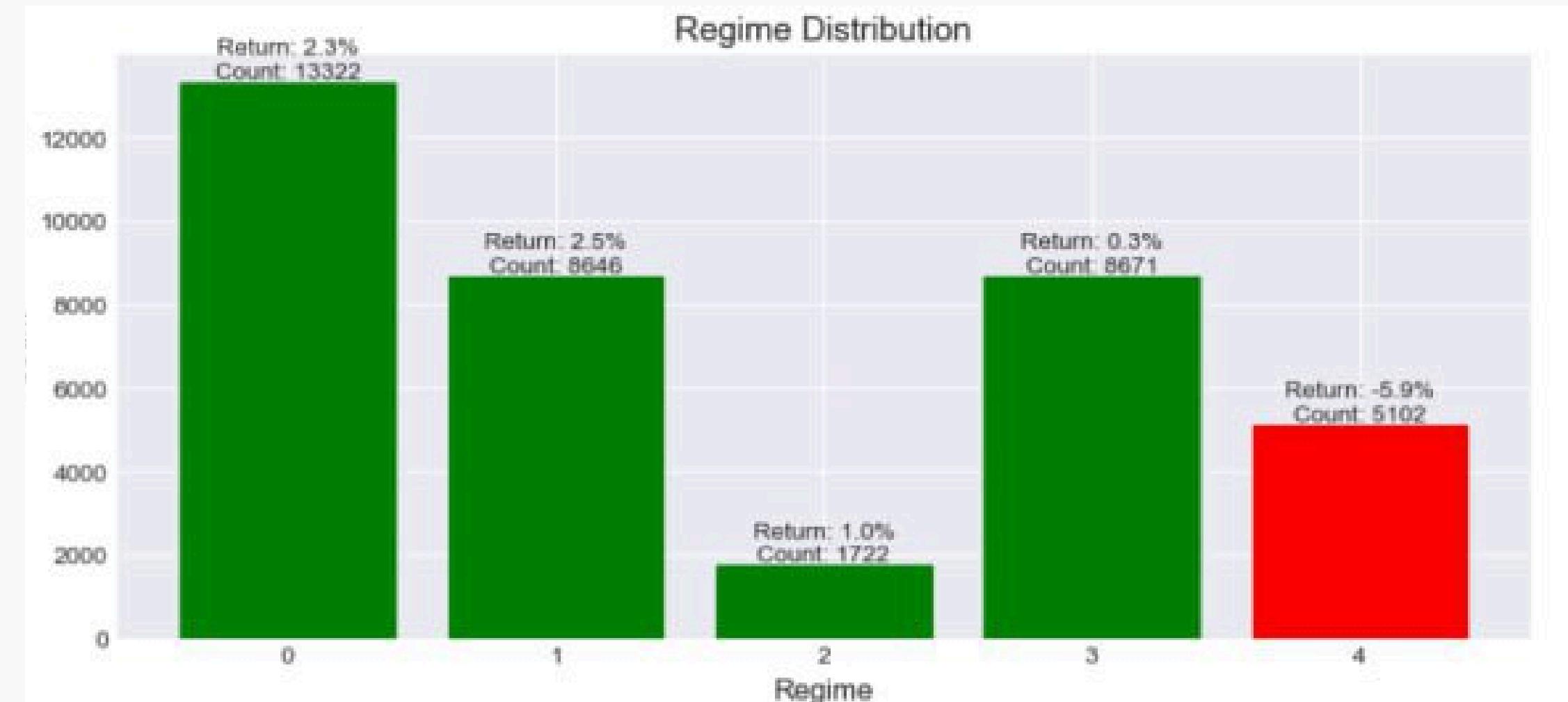


Based on mean return and Sharpe ratio

● Bullish    ● Bearish

## Step 3 Assign Strategy Behavior to Each Regime

- Trade long aggressively 🚀
- Stay in cash or go short ↓
- ✓ Trade cautiously, only high-confidence



# Parameter Optimization & Model Evaluation



Parameters	Description
n_components	Number of HMM regimes (e.g. 2 to 5)
short_ma	Short-term SMA window size
long_ma	Long-term SMA window size
vol_window	Rolling window size for volatility
regime_thresholds	Negative return threshold for identifying bearish regimes

## Parameter Optimization

- Grid Search was used to find the best combination of parameters for better trading performance.
- All combinations tested in parallel using joblib (faster execution).

## Evaluation & Training

For each combination, the model is:

- 1.Training Period: Data before 2023
- 2.Evaluated using Sharpe Ratio (return per unit of risk)
- 3.Best-performing combo is selected

## Validation with Out-of-Sample Testing

Testing Period: 2023 onward

- Robustness
- No overfitting
- Consistency in performance



# Performance Metrics & Success Criteria

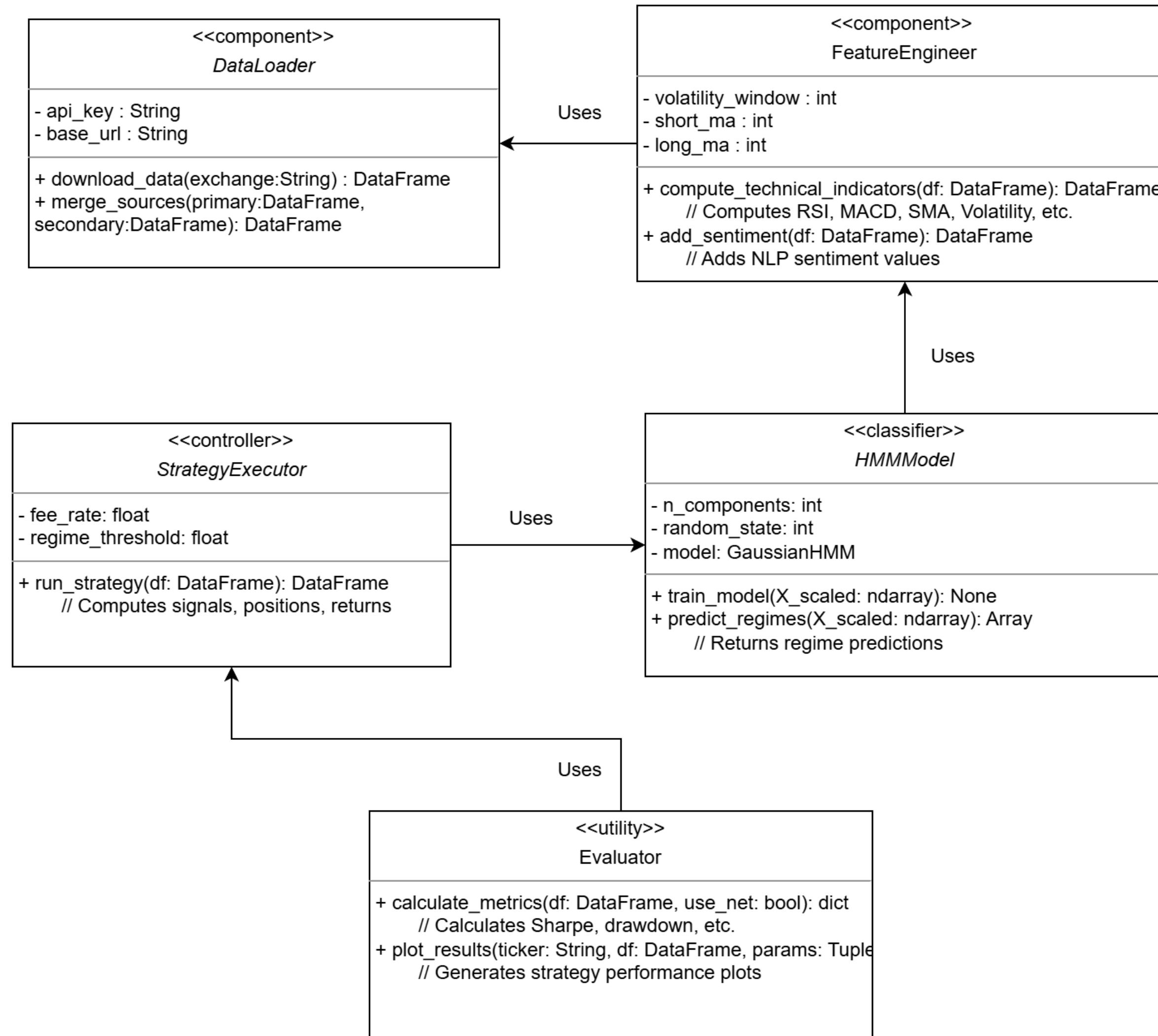
Metric	Target
Sharpe Ratio (SR)	$\geq 1.8$
Max Drawdown (MDD)	$\geq -40\%$
Trade Frequency	$\geq 3\% \text{ per row}$
Win Rate	+ve
Return	yes

```
Processing BTC using multiple Cybotrade sources...
Retrieved 10000 records so far...
Retrieved 20000 records so far...
Retrieved 30000 records so far...
Retrieved 37510 records so far...
✓ Reached latest available timestamp.
Retrieved 9987 records so far...
Retrieved 19986 records so far...
Retrieved 29986 records so far...
Retrieved 37496 records so far...
✓ Reached latest available timestamp.
Testing 1476 parameter combinations for BTC...
Best candidate on training data for BTC:
n_components = 5
short_ma = 10
long_ma = 25
volatility_window = 10
regime_threshold = 0.1
Sharpe Ratio: 0.1122
Max Drawdown: -0.1390
Testing best candidate on out-of-sample data...
```

```
Final evaluation for BTC:
n_components = 5
short_ma = 10
long_ma = 25
volatility_window = 10
regime_threshold = 0.1
Final metrics:
total_return: -0.5733
annualized_return: -0.0057
sharpe_ratio: -0.1523
max_drawdown: -0.6298
win_rate: 0.4565
profit_factor: 0.9735
num_trades: 1125.0000
trade_frequency: 3.0030
avg_profit_per_trade: -0.0002
meets_trade_frequency: True
```

===== SUMMARY OF RESULTS =====

Ticker	Sharpe	Return	Drawdown	Trade Freq	Win Rate
BTC	-0.152338	-0.573288	-0.629806	3.002963	0.456498



# Performance Visualization

Processing BTC using multiple Cybotrade sources...  
 Retrieved 10000 records so far...  
 Retrieved 20000 records so far...  
 Retrieved 30000 records so far...  
 Retrieved 37510 records so far...  
✓ Reached latest available timestamp.  
 Retrieved 9987 records so far...  
 Retrieved 19986 records so far...  
 Retrieved 29986 records so far...  
 Retrieved 37496 records so far...  
✓ Reached latest available timestamp.  
 Testing 1476 parameter combinations for BTC...  
 Best candidate on training data for BTC:  

```
n_components = 5
short_ma = 10
long_ma = 25
volatility_window = 10
regime_threshold = 0.1
```

 Sharpe Ratio: 0.1122  
 Max Drawdown: -0.1390  
 Testing best candidate on out-of-sample data...  
  
 Final evaluation for BTC:  

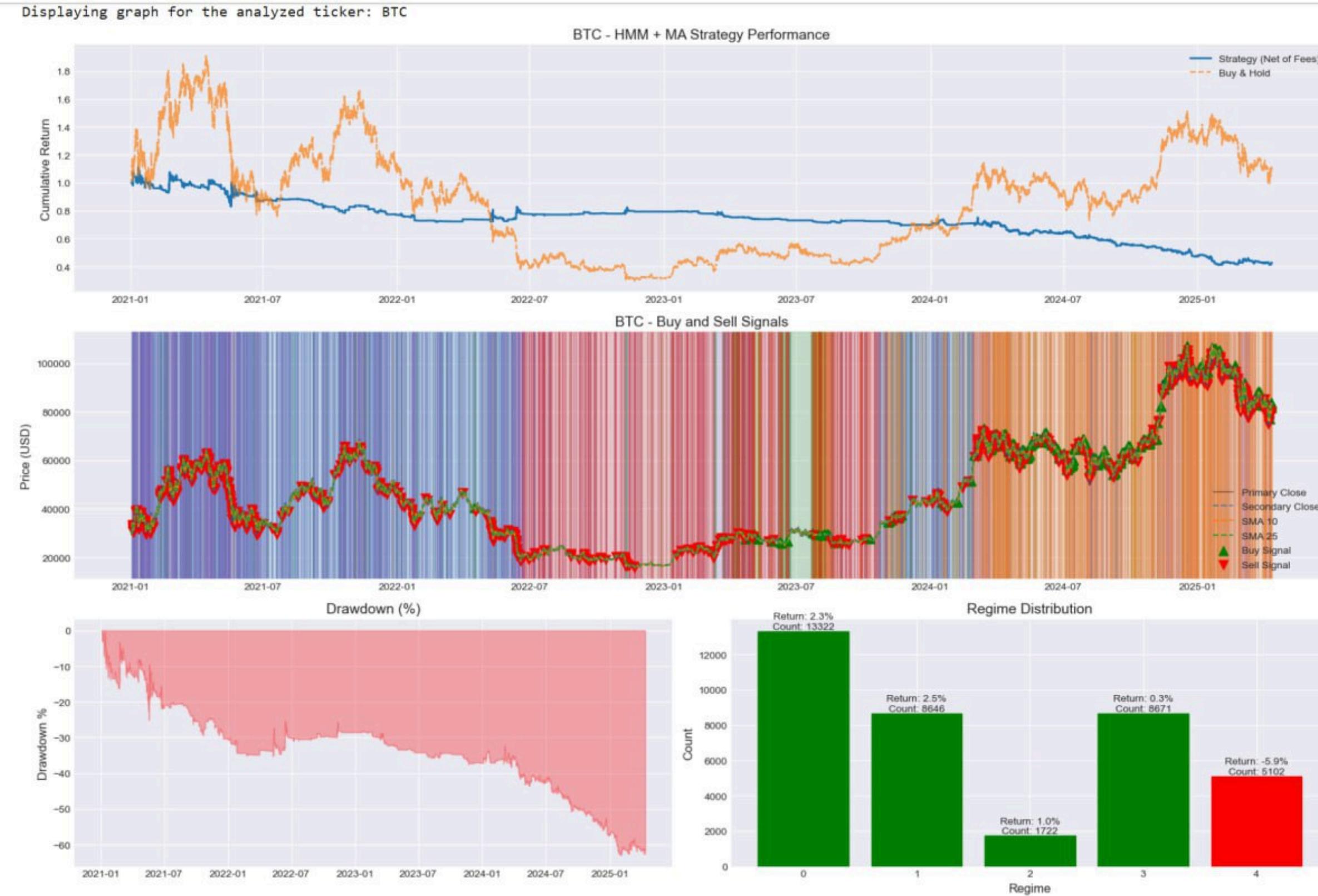
```
n_components = 5
short_ma = 10
long_ma = 25
volatility_window = 10
regime_threshold = 0.1
```

 Final metrics:  

```
total_return: -0.5733
annualized_return: -0.0057
sharpe_ratio: -0.1523
max_drawdown: -0.6298
win_rate: 0.4565
profit_factor: 0.9735
num_trades: 1125.0000
trade_frequency: 3.0030
avg_profit_per_trade: -0.0002
meets_trade_frequency: True
```

  
**===== SUMMARY OF RESULTS =====**

Ticker	Sharpe	Return	Drawdown	Trade Freq	Win Rate
BTC	-0.152338	-0.573288	-0.629806	3.002963	0.456498





*Thank you*

