Methodology

- **Our approach is a supervised learning, binary classification task: Will the price go up or down?**
- **Data Sourcing and Resampling**
  - Using 1-minute BTC/USDT data, resampled to 1min, 4H, and 1D intervals for multi-frequency analysis.
- **Feature Engineering**
  - Features include lagged returns, volatility, moving average ratios, MACD, and ATR.
- **Model & Backtesting**
  - A LightGBM classifier is trained and evaluated on a hold-out test set to simulate real-world trading.
 - Ratio of train and test data 9:1
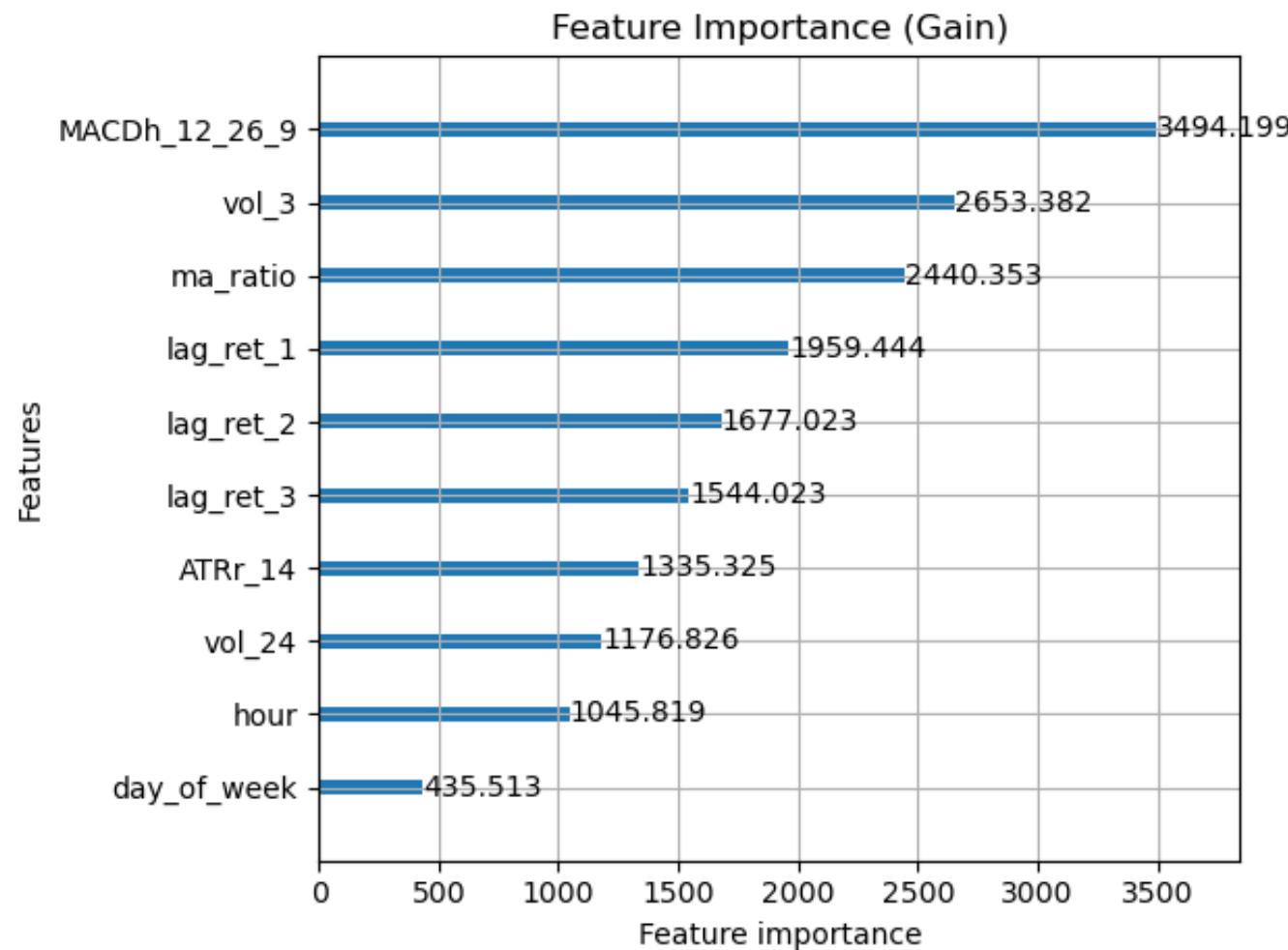
```
base_clf = lgb.LGBMClassifier(objective='binary', random_state=42, scale_pos_weight=scale_pos_weight_value)
```

Model Deep Dive: LightGBM
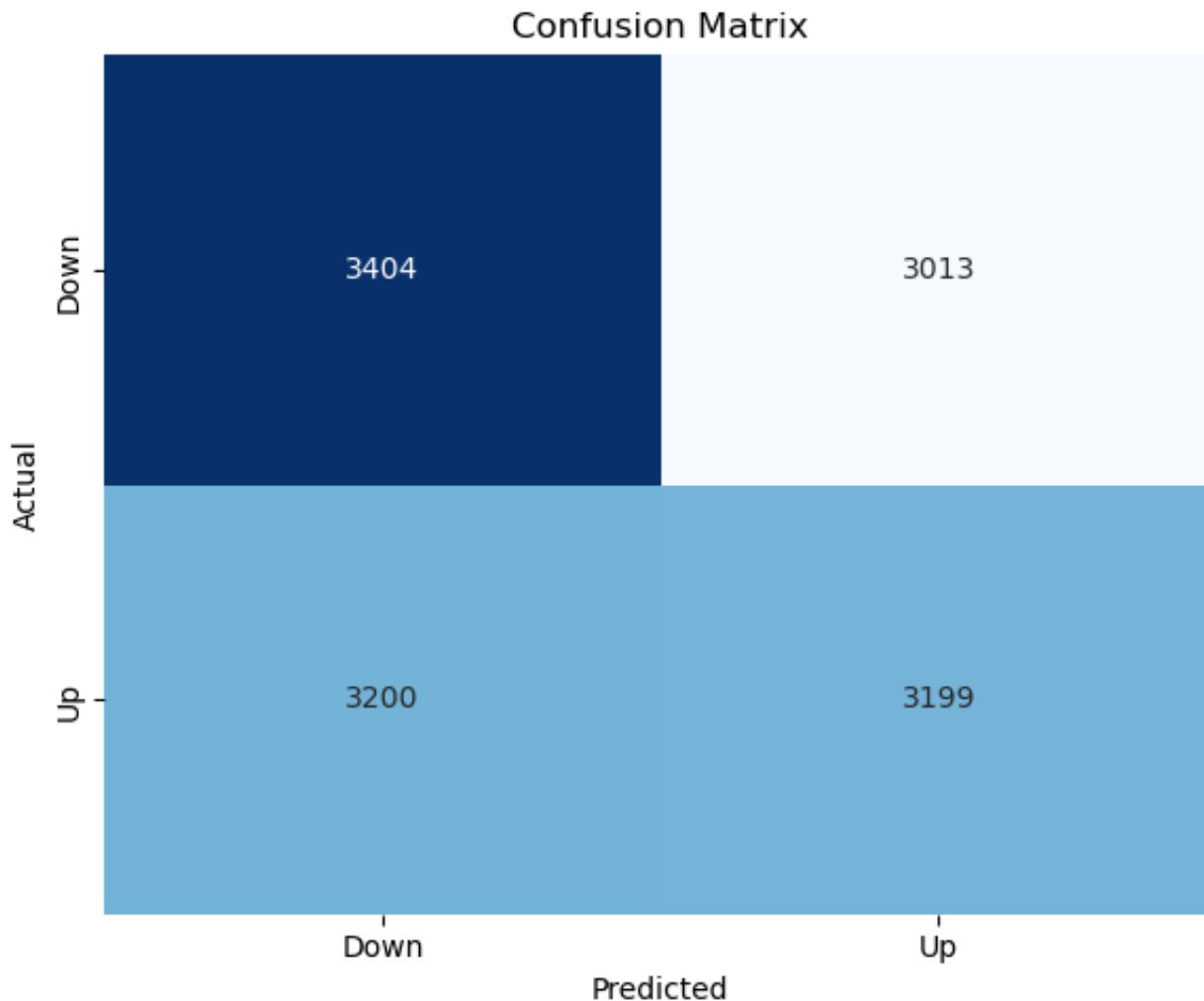
- **LightGBM (Light Gradient Boosting Machine)**
  - A fast, efficient gradient boosting framework that builds an ensemble of decision trees, with each tree correcting prior errors.
- **Handling Class Imbalance with `scale_pos_weight`**
  - Problem: Financial data is often imbalanced (fewer 'up' moves than 'down'/'flat').
  - Solution: `scale_pos_weight` gives more importance to the minority class (the 'up' moves).
  - Benefit: Prevents the model from just predicting the majority class, improving its ability to find valuable trade signals.
- We use GridSearchCV() and scoring='roc_auc' to find the best parameters.

- *We take 10mins-interval of BTC results as an example to illustrate the training process and results.*

During the training process, for each case, we identify the most important features

Feature Importance (Gain)
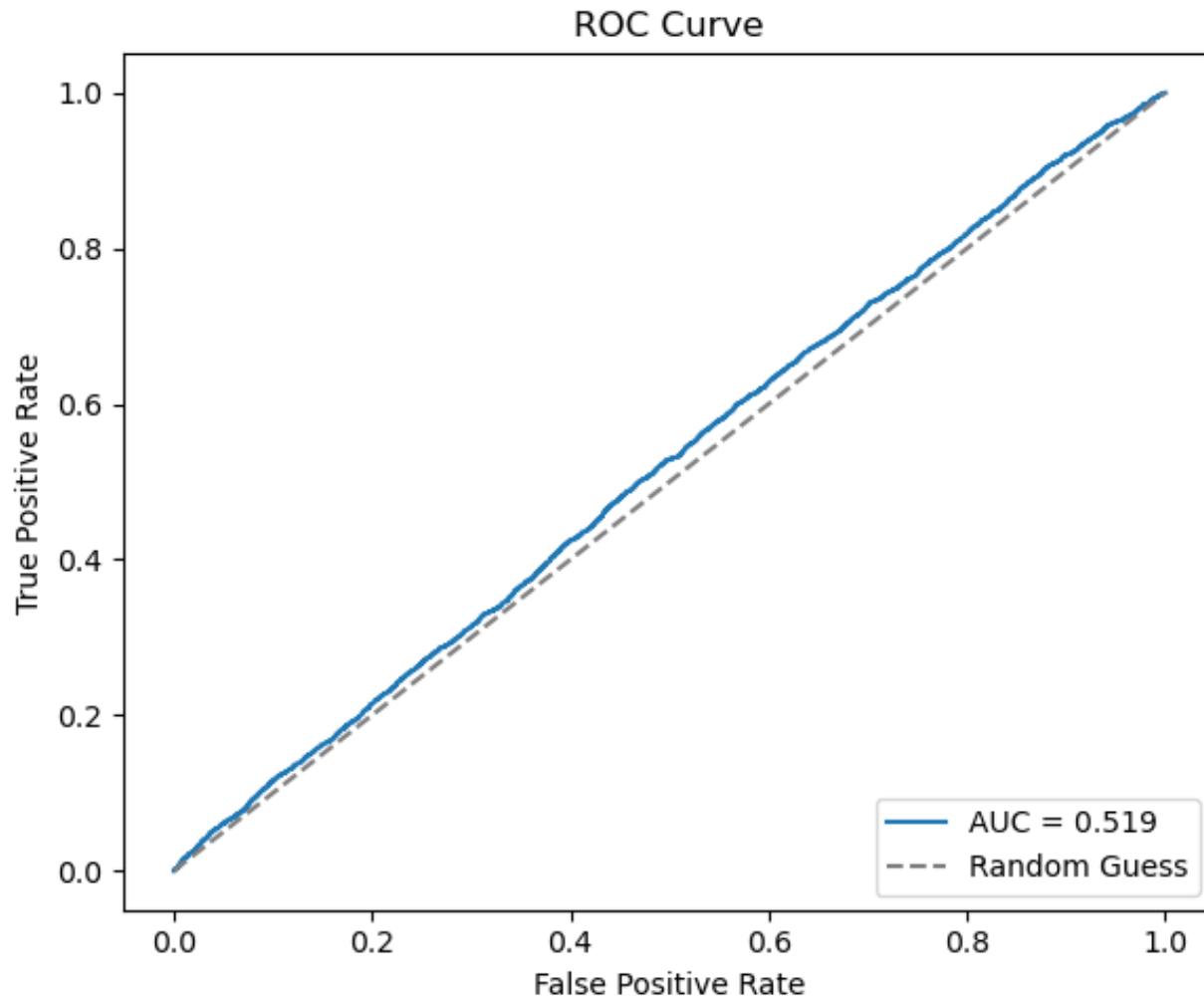
The other important signature in the classier is the confusion matrix
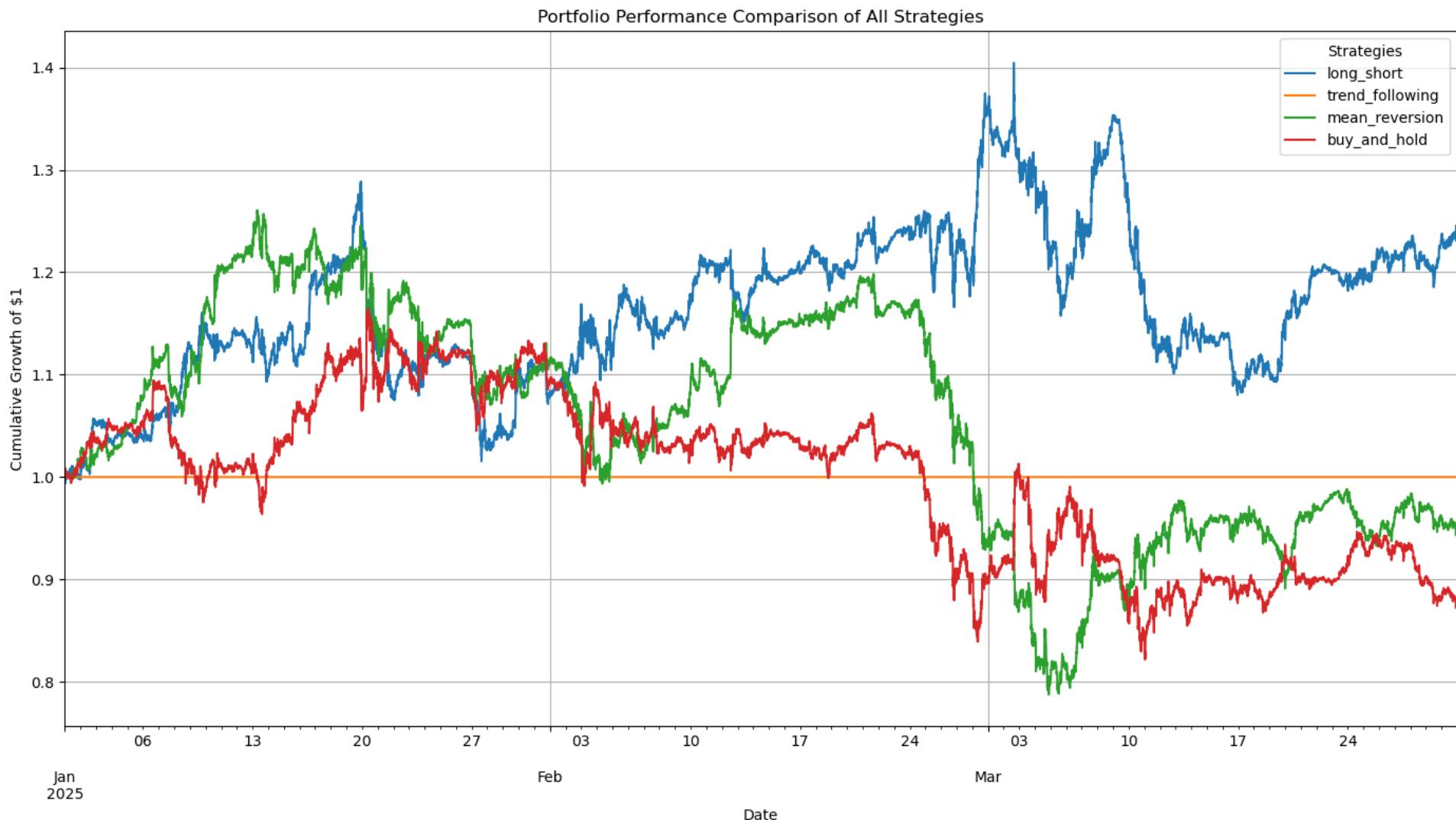
Accuracy of prediction is above 50%.

The other important signature in the classier is the confusion matrix


ROC is slightly above 50%.
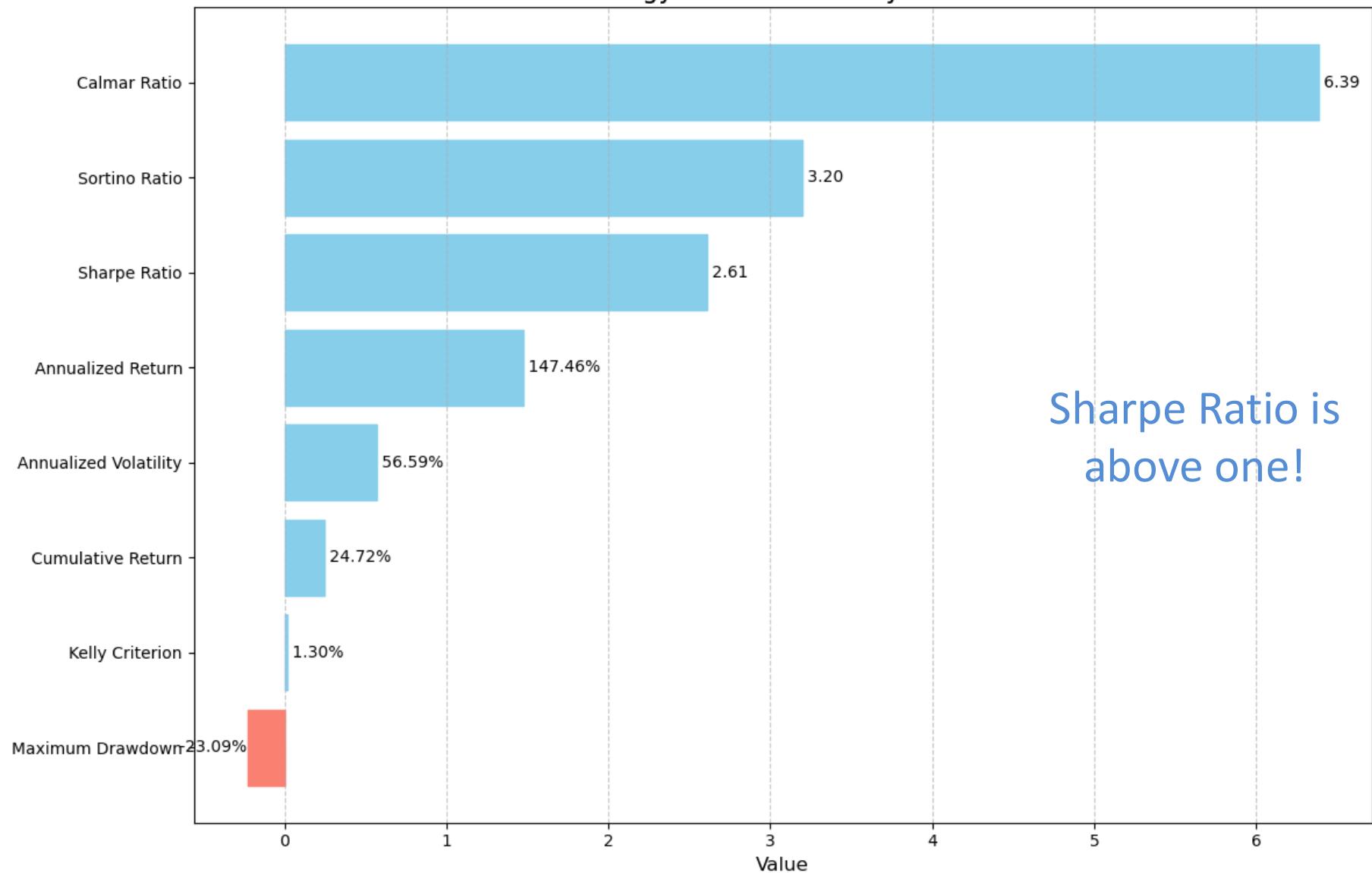
ML-based Strategy versus non-MLs

Portfolio Performance Comparison of All Strategies

ML strategies win over standard ones for BTCs!

ML-based Strategy KPI scores
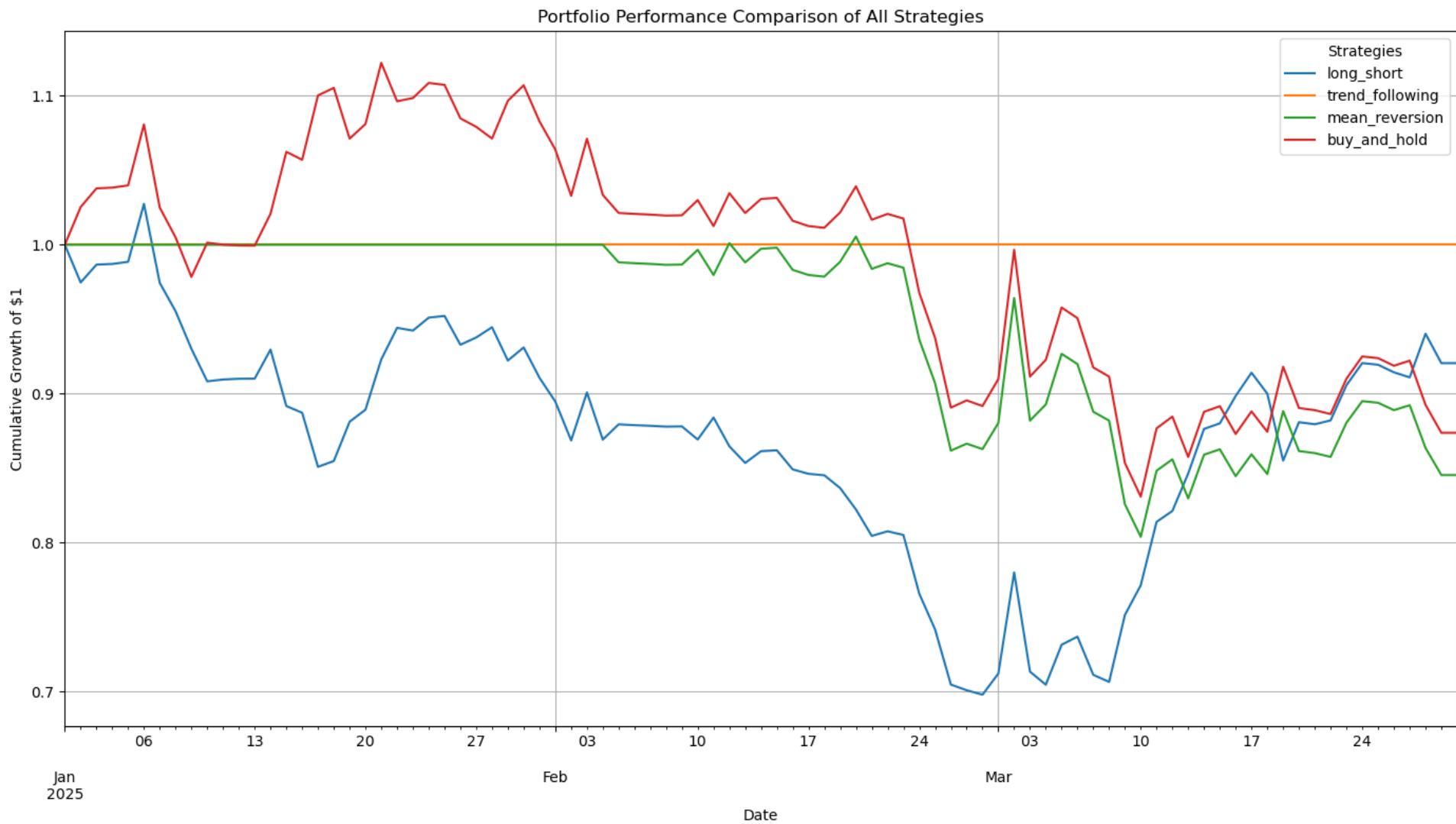
Strategy Performance Key Metrics

Metric	Value
Calmar Ratio	6.39
Sortino Ratio	3.20
Sharpe Ratio	2.61
Annualized Return	147.46%
Annualized Volatility	56.59%
Cumulative Return	24.72%
Kelly Criterion	1.30%
Maximum Drawdown	-23.09%

Sharpe Ratio is above one!

Now we also present the results of 1mins, 1h, 4h, 1d-interval to show that
the best interval for ML is below 1h.
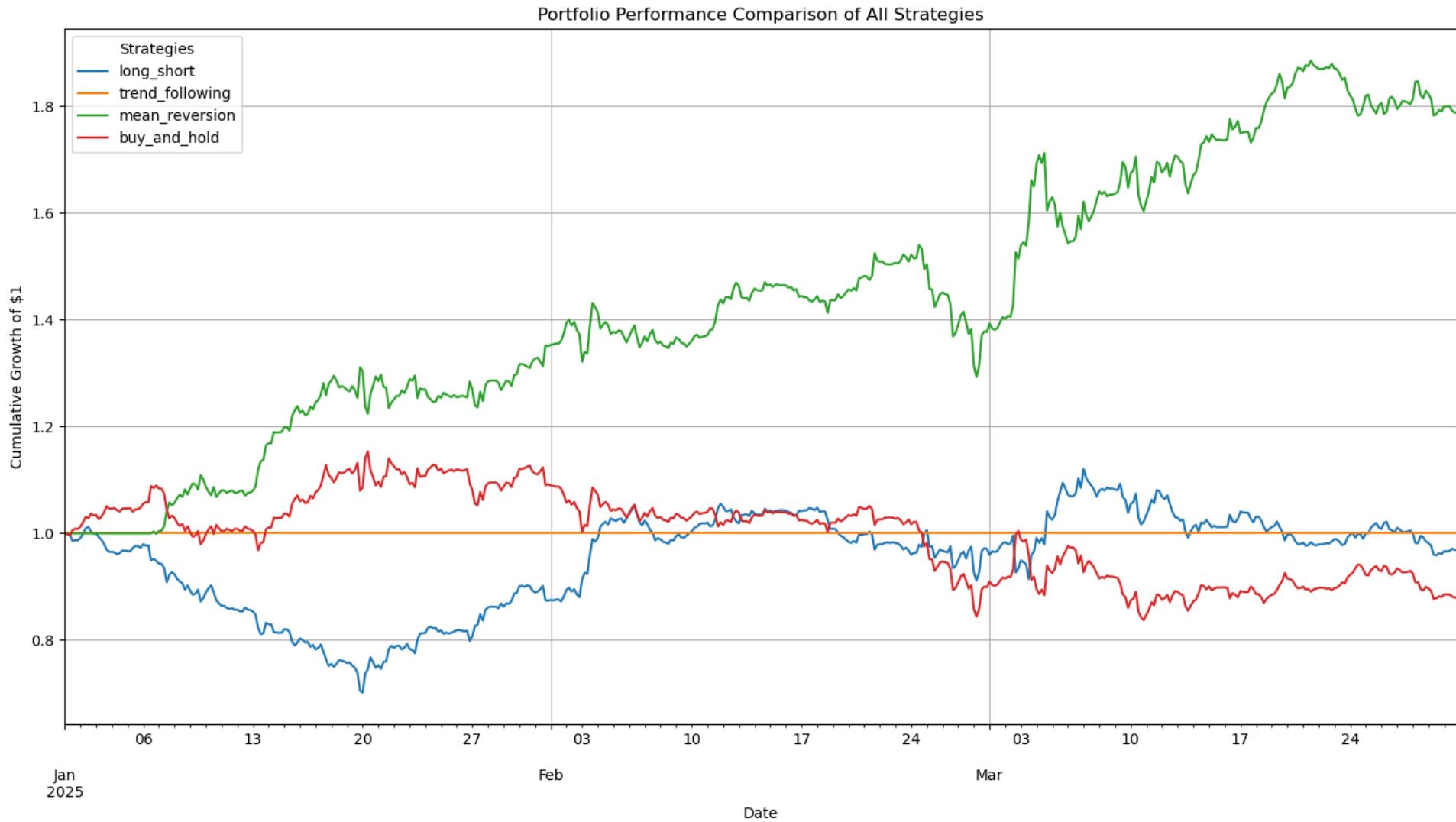
We start from 1d-interval results.

ML-result-based strategy is bad performance

Portfolio Performance Comparison of All Strategies
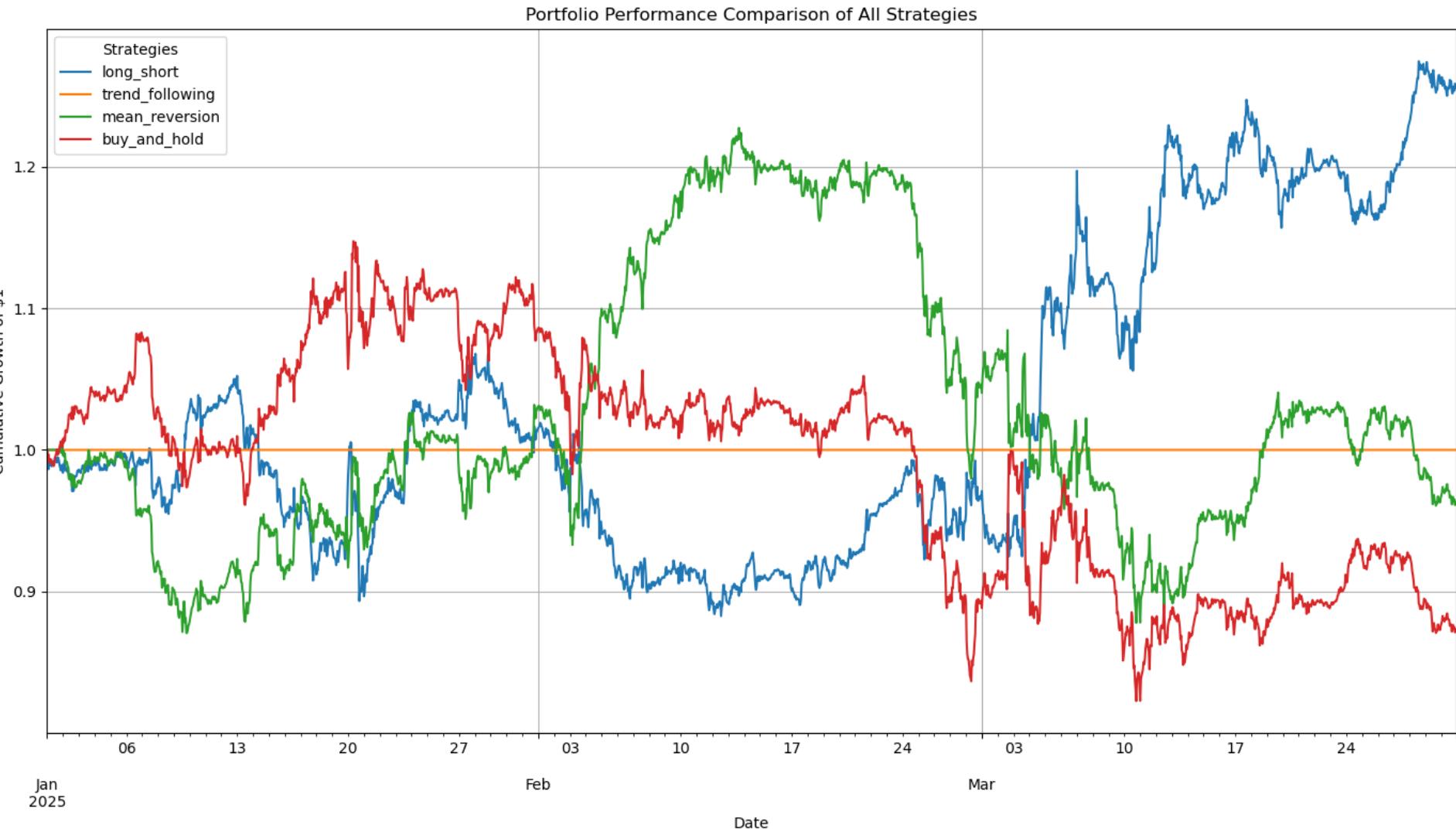
Now consider 4h-interval results.

ML-result-based strategy is bad again

Portfolio Performance Comparison of All Strategies
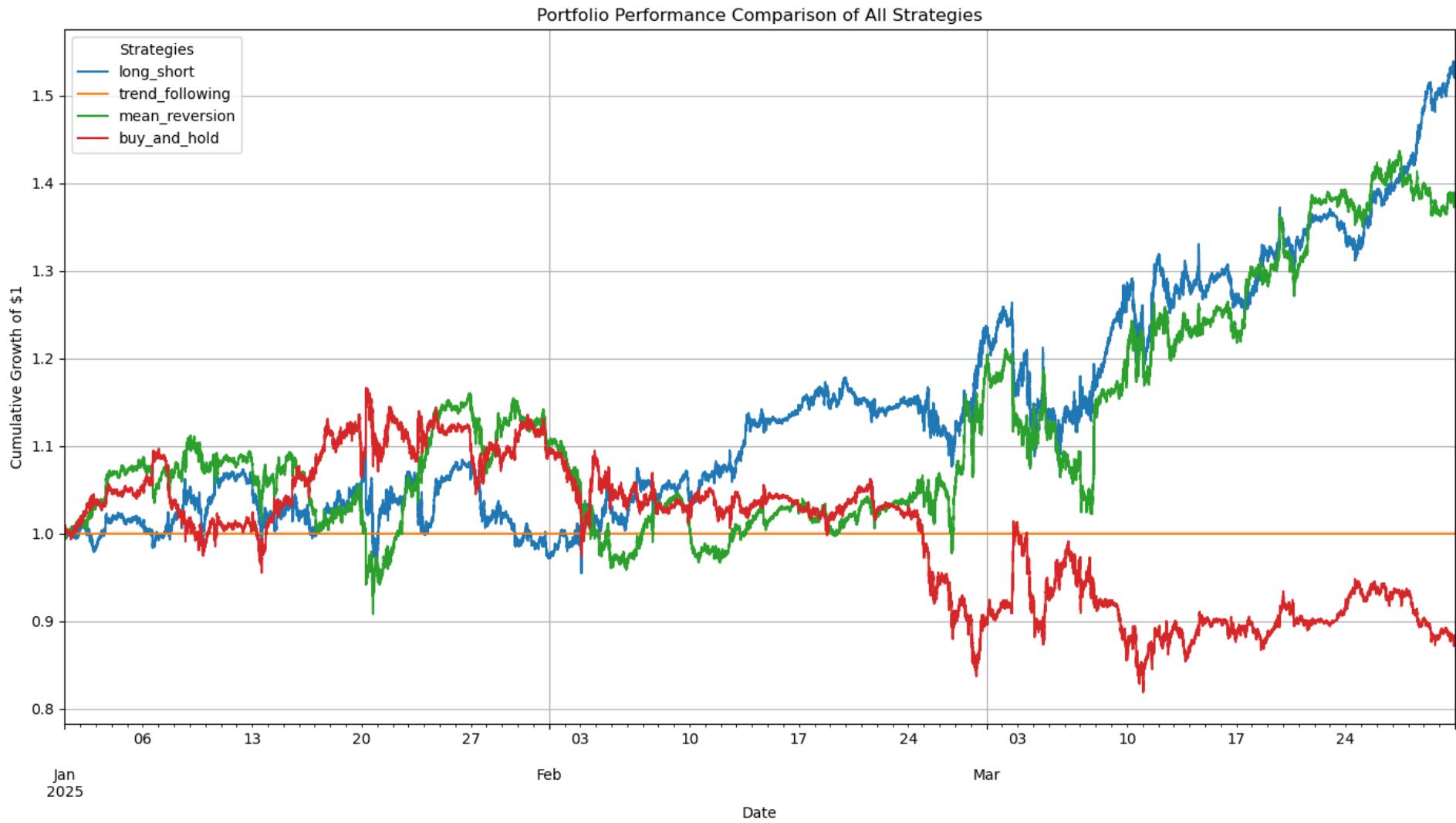
Now consider 1h-interval results.

ML-result-based strategy is better for 1h!

Portfolio Performance Comparison of All Strategies

Now consider 1mins

ML-result-based strategy is also good for 1mins!

Portfolio Performance Comparison of All Strategies
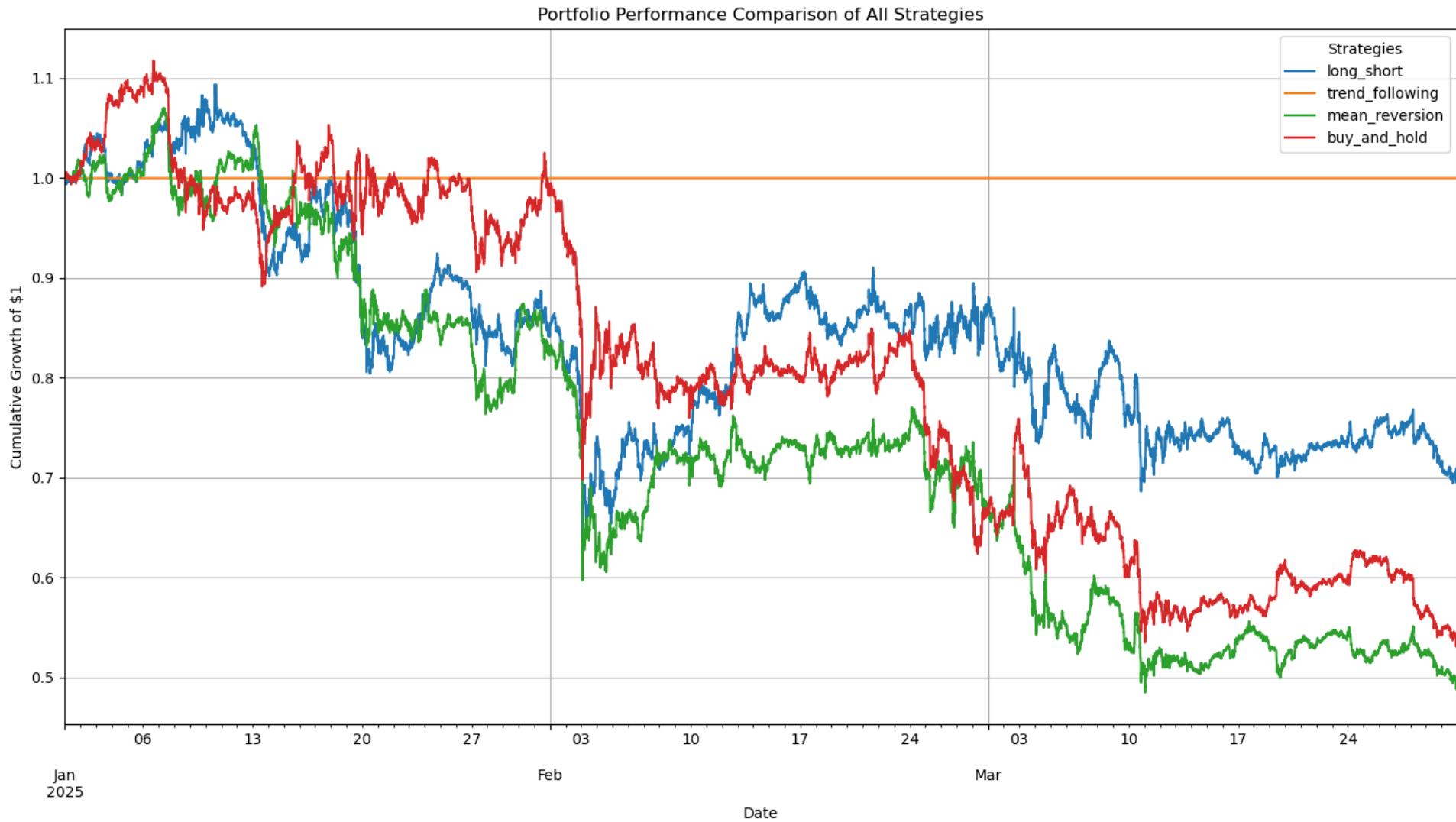
So we reach the first conclusion:
LightGBM classifier ML model is good when the time interval of training data is smaller than 1h.

Now we also present the results of 10mins results of the other four crypto-coins.
*It turns out that the model is good for some coins.*
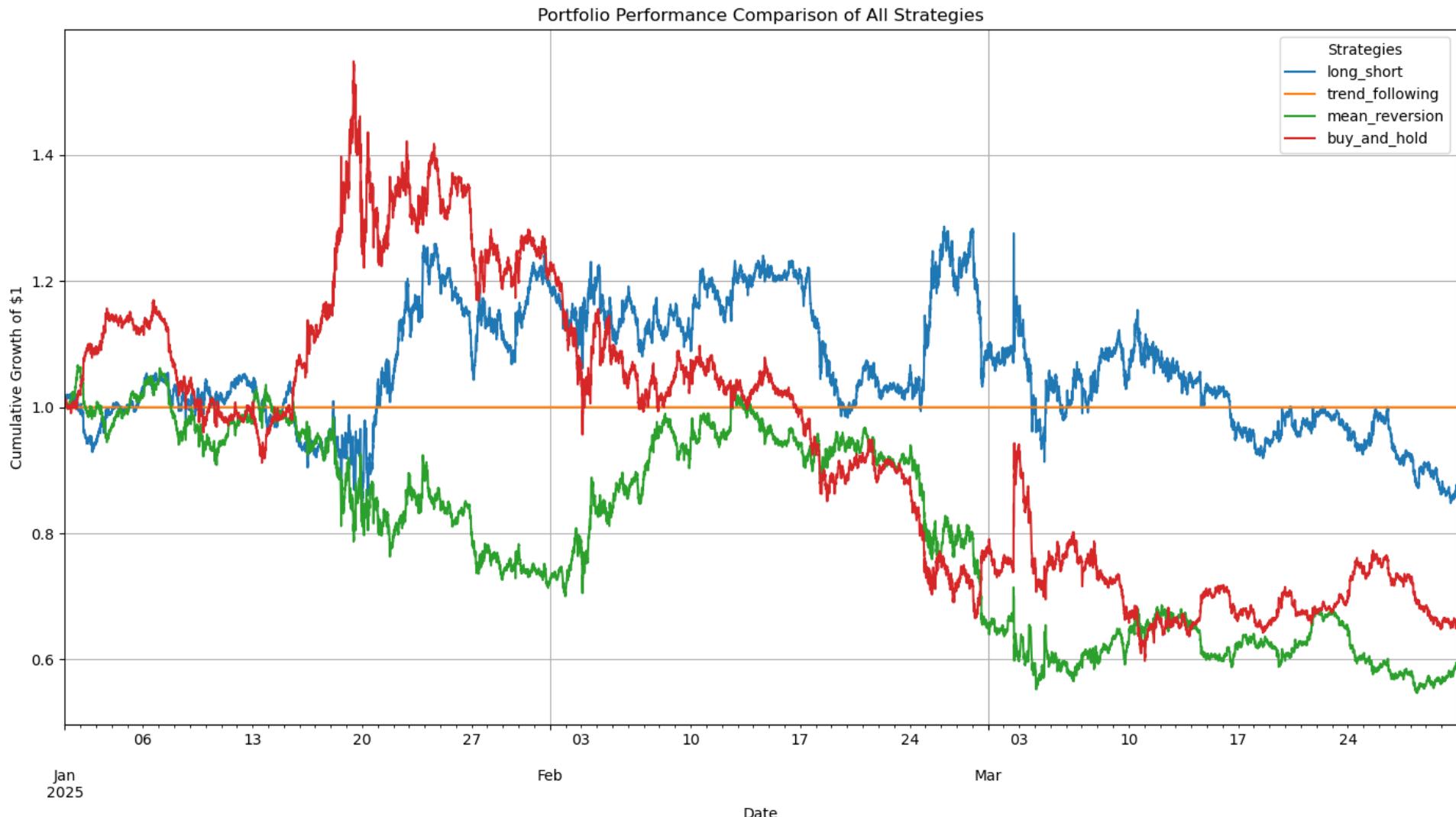
ETH

All strategies are not good but ML-based is best among them.

Portfolio Performance Comparison of All Strategies
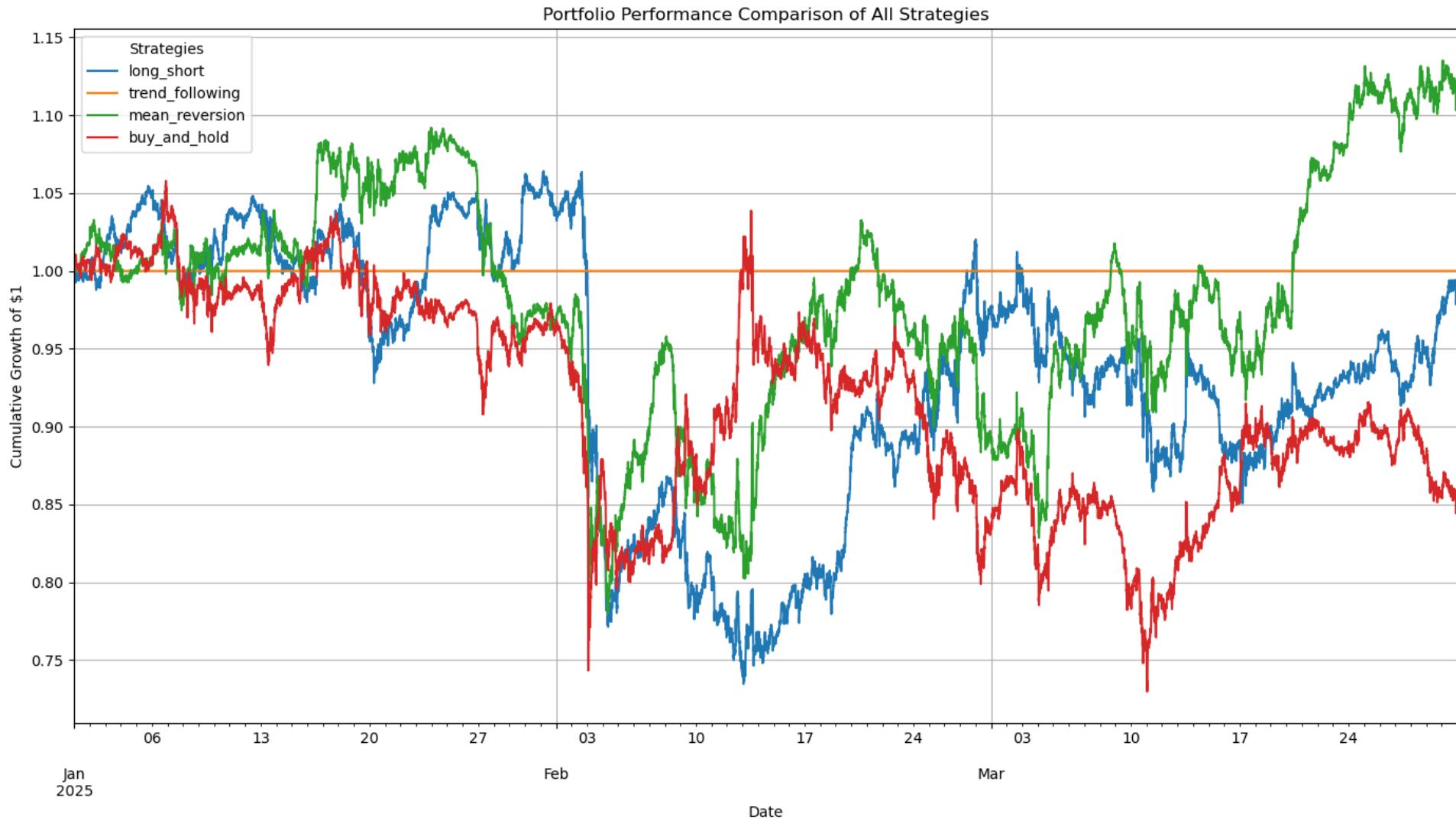
SOLU

ML-based is best among them and potentially profitable.

Portfolio Performance Comparison of All Strategies
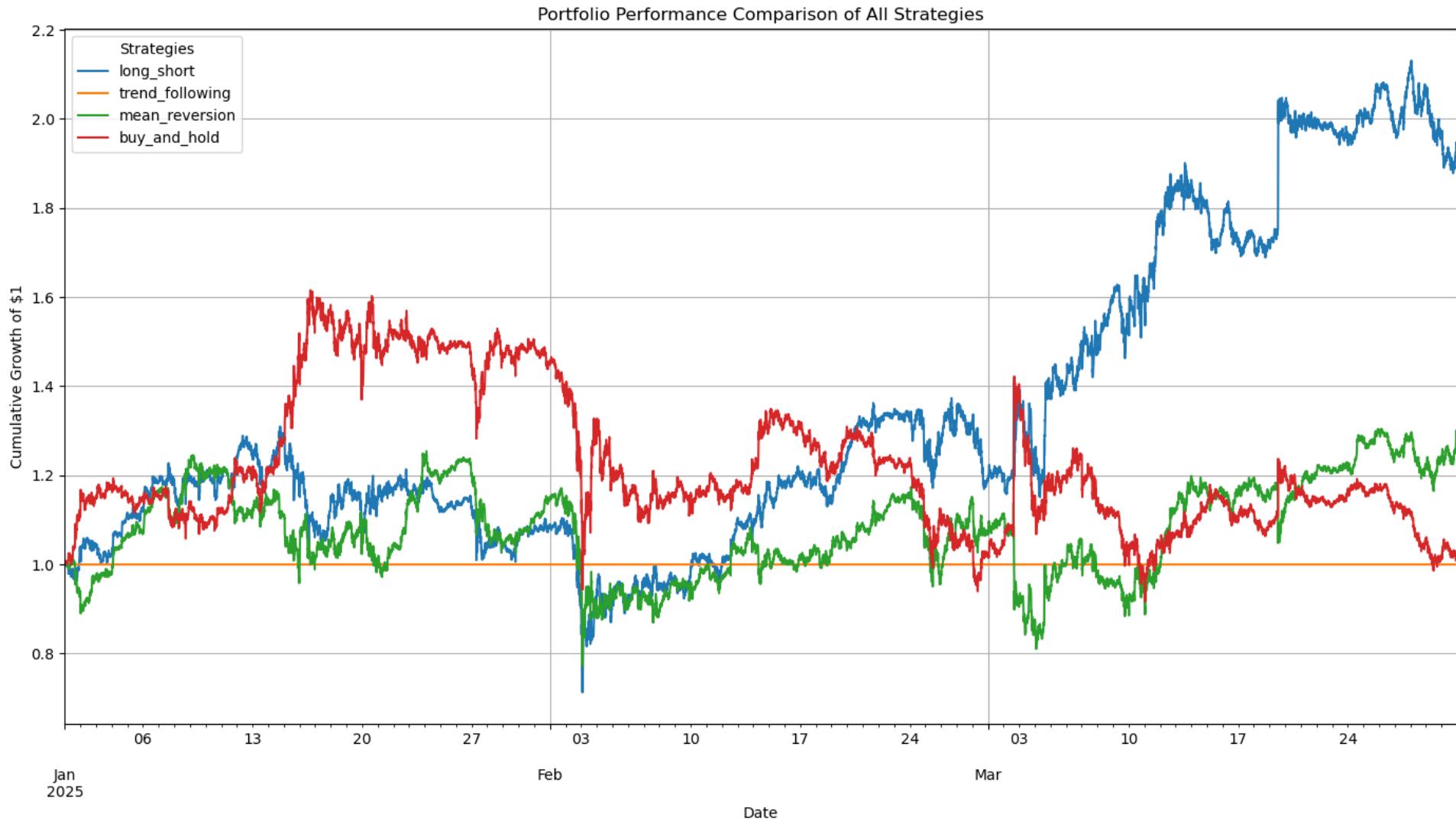
BNB

ML-result-based strategy is not good for BNB,

Portfolio Performance Comparison of All Strategies

XRPU

ML-based is best among them and can be profitable.

Portfolio Performance Comparison of All Strategies

So we reach the second conclusion:
LightGBM classifier ML model is better than conventional ones for almost all coins except BNB; Profitable for BTC, XRPU, SOLU.