Crypto Forecasting Model and Strategy

Christian Park

May 23, 2025

Overview

This repo trains a separate Lasso or Ridge regression model for each altcoin to predict the next day's closing price using only past data — no lookahead bias. Each model uses 2023 data to train and is tested on 2024 onward. For each coin, the model with the highest R^2 score after a time-based split is chosen.

Features

All features are computed using $t-1$ data:

- `log_return`: $\log\left(\frac{P_{t-1}}{P_{t-2}}\right)$

- `ema_20`: 20-day exponential moving average

- `macd_hist`: MACD histogram (momentum signal)

- `daily_range_pct`: $\frac{\text{High}_{t-1} - \text{Low}_{t-1}}{\text{Close}_{t-1}}$

- `volume_change`: Percent change in volume from $t-2$ to $t-1$

- `price_to_support`: Distance from prior close to a rolling support level

Data is sourced from Polygon.io and Yahoo Finance. Daily OHLCV is pulled and cleaned for each coin up to May 21, 2025. Models then predict May 22 prices using May 21 data.

Performance Summary

On backtest, average prediction error across all coins was roughly 4.15%. Coins like MATIC, ARB, and RUNE had accurate predictions. Coins like INJ and FET were off by 10–14%. Model R^2 scores ranged from 0.62 to 0.98.

Directionally, the model was correct about 60% of the time. However, accuracy improved when the predicted move was greater than 2% in either direction. This directional conviction serves as the foundation for the trading strategy.

Trading Strategy

Rather than directly trading on price predictions, the model is used as a signal generator:

- If predicted change $> 2\%$: go long

- If predicted change $< -2\%$: go short

- Else: do nothing (noise)

On May 22, 2025, this signal fired on six coins. Five long signals (ARB, RUNE, SOL, ADA, AVAX) and one short (MATIC). All predictions were directionally accurate. Average return: 7.2% in one day.

Hypothetical Strategy Backtest

Suppose we allocate $\$10,000$ per trade. Let r_i be the return of the i-th trade and N the total number of trades. Define portfolio return R and Sharpe ratio S as:

$$R = \frac{1}{N} \sum_{i=1}^{N} r_i, \quad S = \frac{\mathbb{E}[r_i - r_f]}{\sigma_r}$$

where $r_f = 0$ (assuming risk-free rate is negligible) and σ_r is the standard deviation of returns. Based on the May 22 signals, $R = 7.2\%$, $\sigma_r \approx 3.5\%$, so:

$$S \approx \frac{0.072}{0.035} \approx 2.06$$

A Sharpe ratio above 2 is considered excellent in quant finance.

Key Takeaway

This isn't a crystal ball. It's not psychic. But it's not random either. When the model is loud, it tends to be right. That signal—when strong—is tradable.