Signal Miner: Find Unique Alpha & Beat the Benchmark

Revolutionizing Staking:

Aligning users and the fund through unique models.

What is Signal Miner?

Signal Miner is a fully automated model mining framework

designed to generate models that outperform Numerai's benchmark models

in terms of correlation and Sharpe ratio. Instead of staking on pre-existing models

, this tool helps you discover your own unique alpha

, which has a better chance of producing positive MMC (Meta Model Contribution).

Why use Signal Miner?

- Unique Alpha:

Avoids the trap of staking on common, overused models.

- Better Payouts:

Unique signals increase your expected returns

compared to generic staking.

- Automated Discovery:

Efficiently scans a search space for high-performance models

using a scalable, asynchronous approach.

Quick Start: Install & Run

Clone the repo and set up your environment.Instructions available at Github project.

How It Works

The core workflow:

1. Define a Benchmark Model

: This is what your models will aim to outperform.

1. Launch Model Mining

: Explore a grid of hyperparameters asynchronously.

1. Monitor Performance

: Track model evaluations across cross-validation folds.

1. Compare to the Benchmark

: Identify models that exceed performance thresholds.

1. Export Winning Models

: Save the best models for staking or further tuning.

Defining a Benchmark Model

benchmark_cfg = { "colsample_bytree": 0.1, "max_bin": 5, "max_depth": 5, "num_leaves": 15, "min_child_samples": 20, "n_estimators": 2000, "reg_lambda": 0.0, "learning_rate": 0.01, "target": 'target' # Using the first target for simplicity }

Launch Mining

start_mining()

Once mining is started, models will be trained and evaluated

in the background.

Check Progress Anytime:

check_progress()

Progress: 122.0/2002 (6.09%)

Visualizing Cross-Validation Splits

To ensure proper evaluation

, the framework implements time-series cross-validation with an embargo period

:

[

output

1515×875 19.5 KB

](https://forum.numer.ai/uploads/default/original/2X/7/76db081941e77d5163ee7e55f9d476da073036a0.png)

Here, training and test sets are sequentially split

to mimic live trading conditions

—a crucial step for avoiding data leakage.

Mining Results: Past vs. Future Performance

Since yesterday, I've been running Signal Miner to evaluate 70+ models out of 1000

, and we already see many models outperforming the benchmark

on both validation and test

datasets.

Below is a scatter plot

showing how models that performed well in validation (past) also tended to do well in test (future).

Sharpe Ratio: Validation vs. Test

[

sharpe_scatter

1766×1304 154 KB

](https://forum.numer.ai/uploads/default/original/2X/c/ce6d7923e04eb5f46d971023f7565f93628d30c6.jpeg)

Key Insights:

- The red dot

represents the benchmark model.

- While the top validation model wasn't the best in test

, we found several models that outperformed the benchmark in both.

- Positive Correlation

: The best validation models tended

to be among the best in test as well.

- If the scatter plot looked random (a cloud of points), it would suggest the model selection process is noise

—but instead, we see a clear upward trend

.

Goal:

Find a model that beats the benchmark in both correlation & Sharpe ratio.

Still mining!

Scaling Behavior

This entire process can be viewed as a function of the number of trees in the search space

.

For this experiment, I set n_estimators=2000—but early results suggest that increasing this value improves overall performance

.

This hints at a scaling law

, an idea that has come up in community discussions before.

Join the Experiment!

This is an open-source

project, and everyone is welcome to:

Run their own mining experiments

Contribute improvements

(PRs welcome!)

Share results & insights

Ready to try?

Head over to Signal Miner on GitHub

and start mining unique alpha

today!

Let’s Make Staking Great Again!