

Hey everyone,

I've been thinking a lot about how we can improve the staking mechanism in Numerai to better align with the goals of both the fund and its users. The crux of my argument is this: staking should always benefit both the user and the fund.

This isn't a radical idea. In the broader crypto world, staking ETH or other tokens often serves two purposes: it rewards the user for locking their assets, and it directly contributes to the functionality of the system. Staking in Numerai should work the same way—not just as a tool for data scientists (DSs), but as a mechanism that improves the Meta Model by encouraging broader participation.

The Problem With Example Models

Right now, the system allows users to stake on example models, which creates a few issues:

1. No Unique Contribution

: When everyone stakes on the same example model, it provides little to no additional value to the fund.

1. No Real Incentive for Non-DS Stakers

: If staking example models neither benefits the Meta Model nor offers unique opportunities for users, it's hard to justify staking for non-DS participants.

The result? A system that doesn't take full advantage of staking's potential to improve the Meta Model while engaging a wider audience.

A Better Staking Paradigm: Randomized Grid Models

Here's an alternative: a system that generates unique, high-correlation models for every staker through a randomized grid search.

Here's how it could work:

- Generic Staking:

Instead of staking on example models, users stake their NMR to generate a random, well-performing model on the validation set.

- Unique Contributions:

Each staker's model would be unique, meaning every stake contributes a new signal to the Meta Model.

- Aligned Incentives:

The better the signal, the better the payout for the user. This directly aligns user rewards with the Meta Model's performance.

- Risk and Opportunity:

Users could hedge their per-model risk by staking on a variety of randomized models, much like a diversified portfolio. This benefits the fund by delivering more unique signals and benefits users by increasing their chances of higher payouts.

Staking as Signal Mining: The Pipeline for Randomized Models

I can provide the pipeline to implement this idea effectively, giving every staker a unique, randomized high-correlation model. Here's the hyperparameter grid I typically use for generating models:

python

CopyEdit

```
param_dict = { 'colsample_bytree': list(np.linspace(0.001, 1, 100)), 'reg_lambda': list(np.linspace(0, 100_000, 10_000)),
'learning_rate': list(np.linspace(.00001, 1.0, 1000)), 'max_bin': list(np.linspace(2, 5, 4, dtype='int')), 'max_depth':
list(np.linspace(2, 12, 11, dtype='int')), 'num_leaves': list(np.linspace(2, 24, 15, dtype='int')), 'min_child_samples':
list(np.linspace(1, 250, 250, dtype='int')), 'n_estimators': list(np.linspace(100, 25_000, 24_000, dtype='int')), 'target': targets,
# User-specified target values }
```

Using this grid, we can compute the total number of unique hyperparameter combinations:

- Total Combinations

$= 100 \times 10,000 \times 1,000 \times 4 \times 11 \times 15 \times 250 \times 24,000 = 39,600,000,000,000,000 \times 10,000 \times 1,000 \times 4 \times 11$

$$15 \times 250 \times 24,000 = 39,600,000,000,000,000 \times 10,000 \times 1,000 \times 4 \times 11 \times 15 \times 250 \times 24,000 = 39,600,000,000,000,000$$

Yes, you read that right: trillions of potential unique models! With this massive space, every staker could generate a completely unique model, even if we scale up participation dramatically.

Why This Is Better

1. For the Fund

: Instead of everyone staking on the same model, the Meta Model gains from a diverse set of unique signals. This improves the overall performance and robustness of the Meta Model.

1. For the Users

: Generic stakers—who may not be data scientists—can now contribute meaningfully to the system and have a fair chance to earn rewards based on their contribution.

1. For the Ecosystem

: It incentivizes participation from a broader audience, fostering growth and sustainability for Numerai.

This system represents a huge improvement over the current example model setup, where staking on one shared model offers no real benefit to either the staker or the fund.

Call for Feedback

I'd love to hear your thoughts on this idea. I can share code and examples to demonstrate how this system works, why scaling the number of random models consistently improves ensemble performance out-of-sample, and why models generated through this process are likely to maintain strong performance in the future.

How else can we maximize the potential of staking to benefit both users and the Meta Model? Let's collaborate and unlock the full potential of staking for Numerai.