

Hello! I was new a couple of weeks ago and had some of the same questions. For the data:

At the core of the Numerai Tournament is the free dataset. It is made of high quality financial data that has been cleaned and regularized and obfuscated. Each id

corresponds to a stock at a specific time era

. The features

describe the various quantitative attributes of the stock at the time. The target

represents an abstract measure of performance ~4 weeks into the future.

<https://docs.numer.ai/tournament/learn#data>

The data are heavily obfuscated but they are not imaginary. The problem, as I understand it, is less about applying any subject-matter knowledge to the features themselves and more about the general problem of low-signal-high-noise tabular data.