

Hi Richard,

[Umap](#) generates useful features from the dataset, while reducing dimensionality. Unlike PCA, Umap works with the Numerai dataset.

I got the idea, from [Marcos Lopez de Prado's lecture](#):

I've been using it for a while.

See here:

[Numerai](#) (a newer but apparently improved version)

[Numerai](#) (since round 275)

It won't hit the #1

spot on the leaderboard based on CORR, but the created model has ~10% correlation with the metamodel and MMC/CORR ratio is good. It will be interesting to see it's TC score.

Best part of it is, that it's very simple:

```
fit = umap.UMAP(n_components=100, min_dist=0)
```

```
transformed_data = fit.fit_transform(data)
```

Because it's learns embedding without the labels, I can use test AND live for the umap model as well.

Once the dataset is transformed you can train any model on it.