





















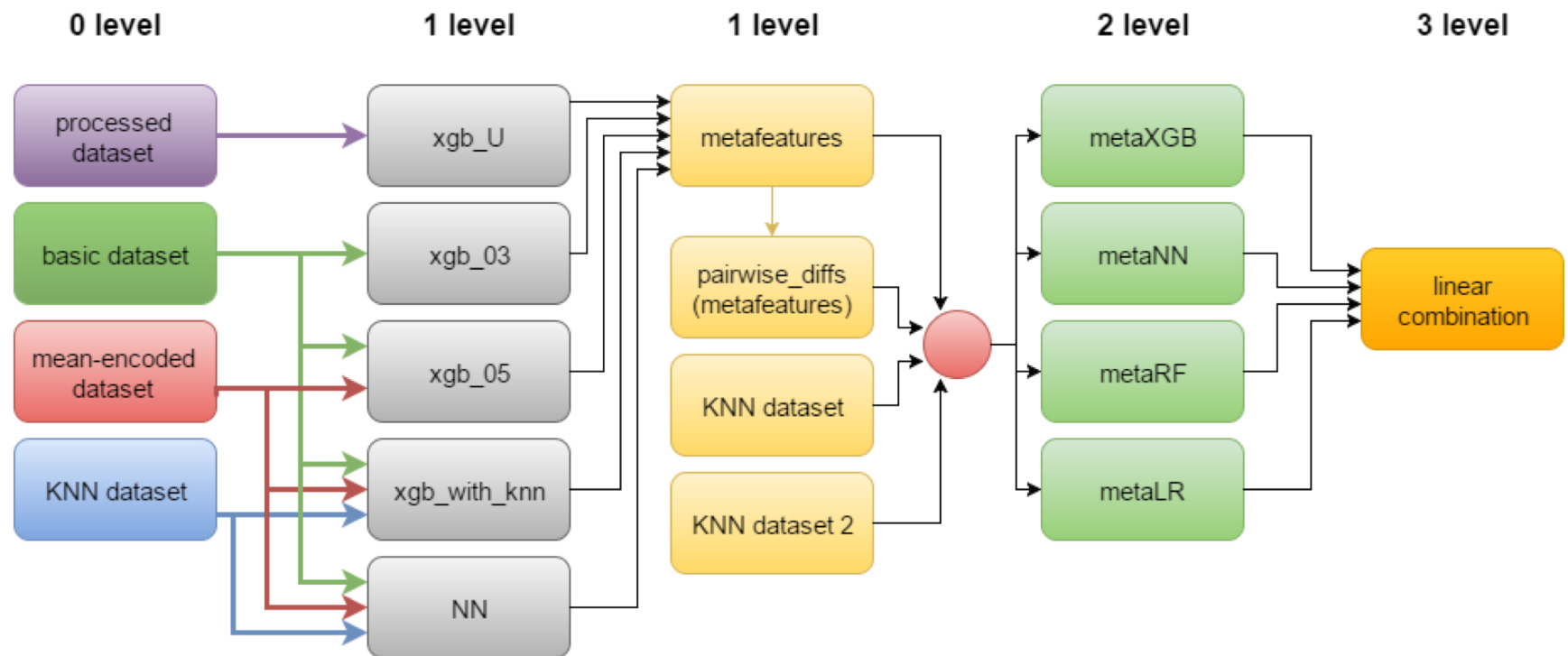


Lending made personal

■ In the money ■ Gold ■ Silver ■ Bronze

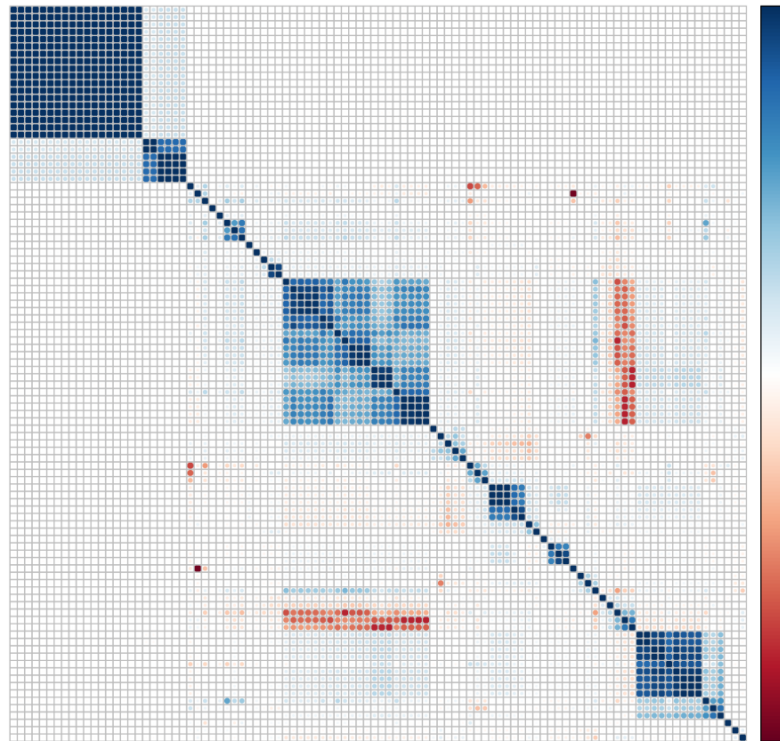
#	Δpub	Team Name	Kernel	Team Members	Score ?	Entries	Last
1	—	Asian Ensemble		   +3	0.80427	274	2y
2	▲ 1	.baGGaj.		   +3	0.80394	166	2y
3	▲ 1	Merging the Mundane and th...		  	0.80390	44	2y
4	▼ 2	ARG eMMSamble		   +3	0.80367	328	2y
5	—	n_m			0.80208	50	2y
6	▲ 1	KazAnova & clobber		 	0.80195	64	2y
7	▼ 1	Gzs_iceberg			0.80179	30	2y
8	—	DP&MJ&JL&SS		   	0.80158	202	2y

# Stacking scheme

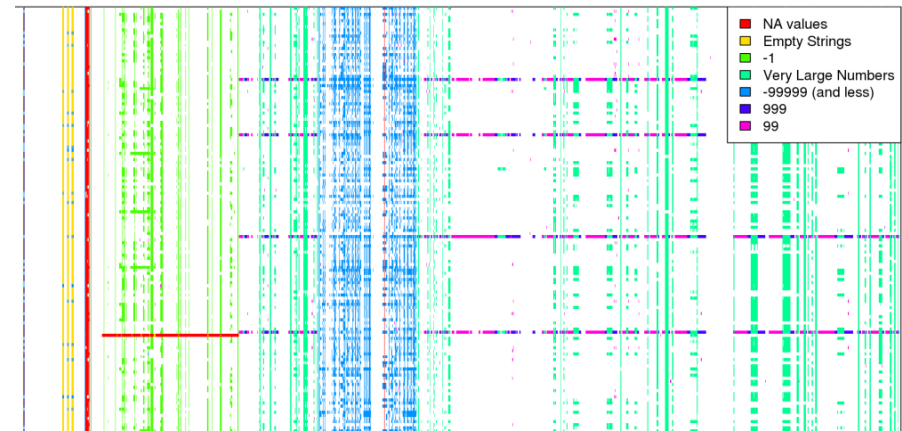


# Determine whether to send a direct mail piece to a customer

2 classes  
AUC



145k objects in train  
anonymized features  
(~2k)



# Feature packs

0 level

processed  
dataset

**processed dataset** - Advanced data cleaning and feature engineering

basic dataset

**basic dataset** - Basic data cleaning and feature engineering

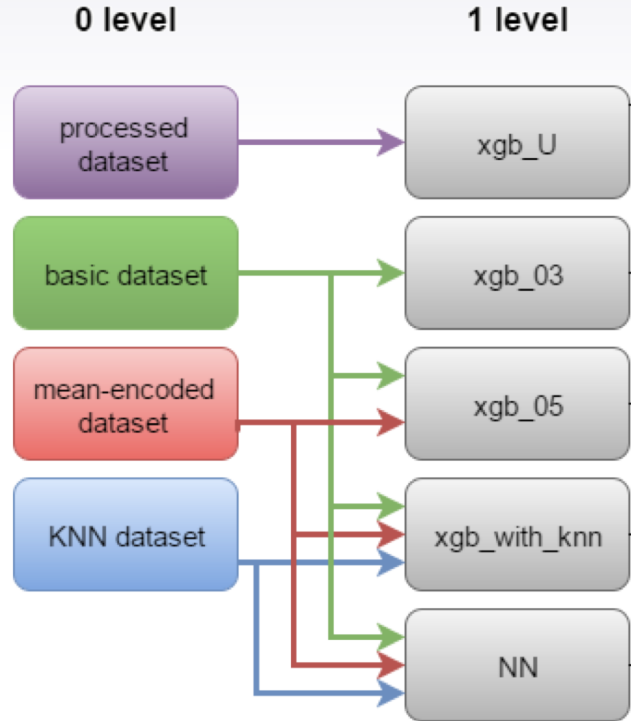
mean-encoded  
dataset

**mean-encoded dataset** - Projecting features into homogeneous space

KNN dataset

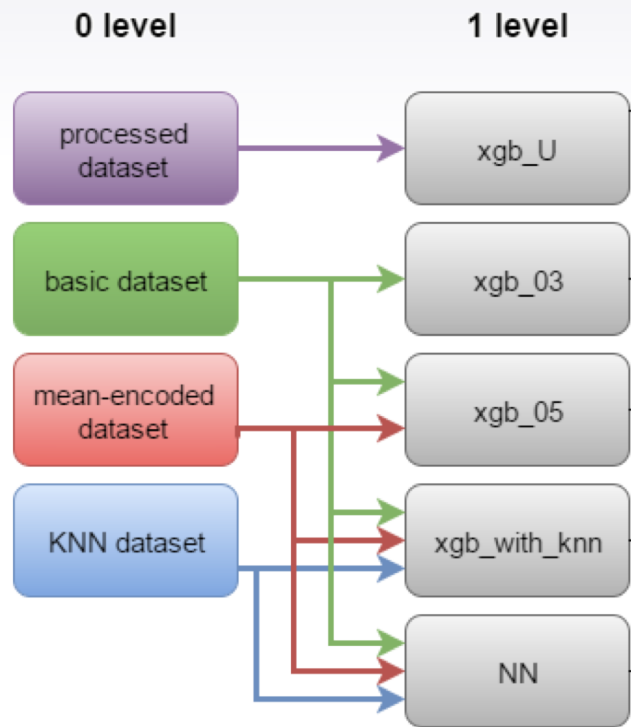
**KNN dataset** - distance features on mean-encoded

# 1 level: out-of-fold predictions - xgboost



- oof predictions (== metafeatures) should be diverse
- each metafeature should bring 'new' information about Y

# 1 level: out-of-fold predictions - NN



- StandardScaler:

$$x = \frac{x - \text{mean}(x)}{\text{std}(x - \text{mean}(x))}$$

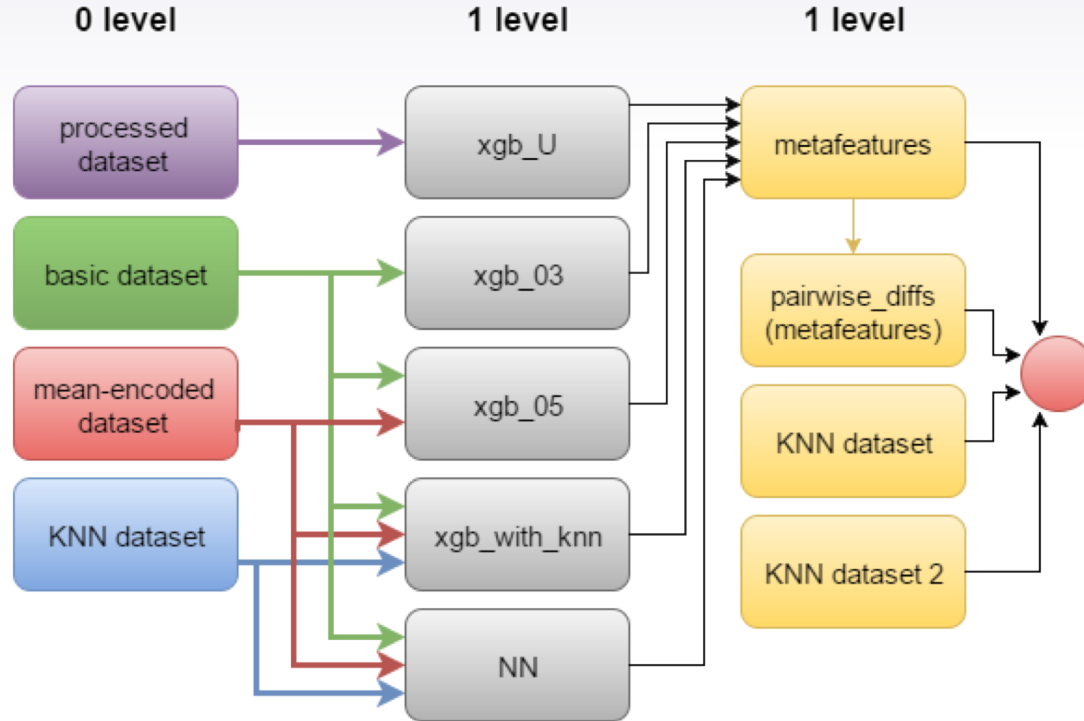
- Ranks:

$$\text{rank}([0.1, 4, -2]) = [2, 3, 1]$$

- Power:

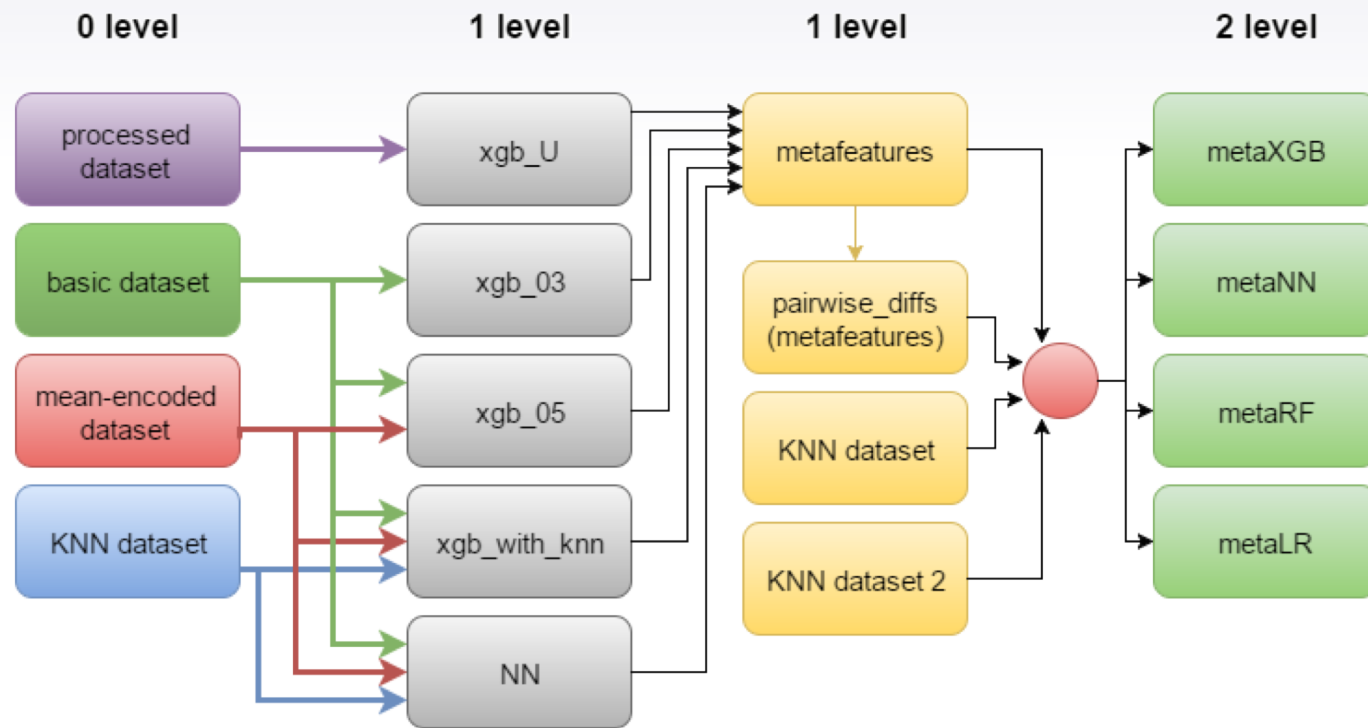
$$\text{feature} = \text{sign}(\text{feature}) * |\text{feature}| ** 0.5$$

# 1 level - features



- KNN dataset - distance features on mean-encoded
- KNN dataset 2 - distance features on (mean-encoded \*\* 0.5)
- Additional features should bring new information to 1 level

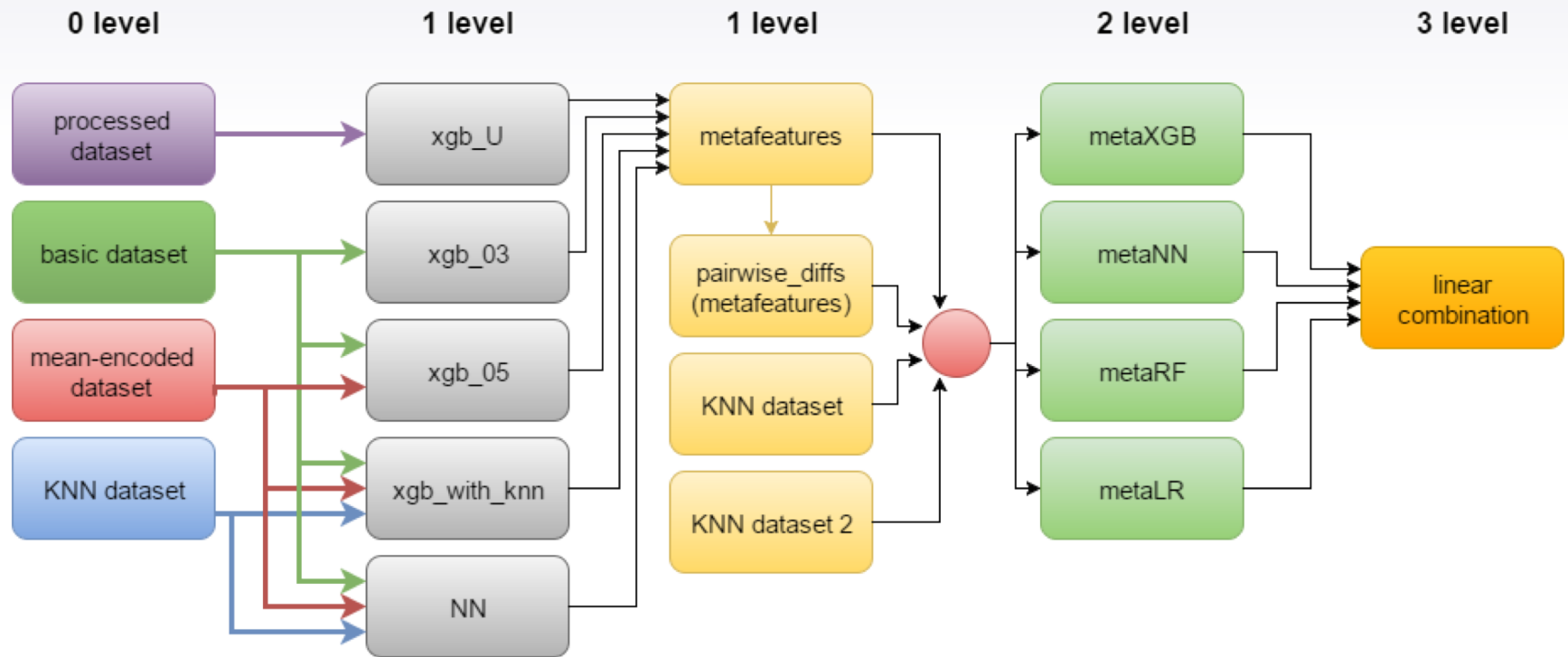
# 2 level - classifiers



- classifiers should be simple
- predictions should be diverse



# 3 level - final



- coefficients are estimated directly

# Scripts on kaggle:

- <https://www.kaggle.com/steves/springleaf-marketing-response/visualizing-na-values/notebook>
- <https://www.kaggle.com/darraghdog/springleaf-marketing-response/explore-springleaf>