

## A. Hyper Parameter Tuning

### A.1. Classification

Parameter Name	Search Space
Model	Catboost
n_estimators	qrandint(1, 1000, 10)
max_depth	randint(2, 10)
reg_lambda	randint(2, 10)
auto_class_weights	choice(['Balanced', 'SqrtBalanced', None])
Model	Xgboost
n_estimators	qrandint(1, 1000, 10)
max_depth	randint(2, 10)
reg_alpha	randint(2, 10)
reg_lambda	randint(2, 10)
min_child_weight	randint(2, 30)
colsample_bytree	quniform(0.1, 0.99, 0.01)
subsample	quniform(0.1, 0.99, 0.01)
scale_pos_weight	randint(2, 50)
Model	MLP
epochs	qrandint(1, 40, 5)
batch_size	pow(2, randint(5, 12))
learning_rate	qrandint(1, 100, 4)*1e-4
optimizer_name	choice(['adam', 'adamw'])
loss_criterion	choice(['cross_entropy', 'weighted_cross_entropy'])
num_layers	randint(1, 7)
layer_type	choice(['linear', 'highway'])
activation_fn	choice(['relu', 'tanh', 'gelu'])
dropout	qloguniform(0.05, 0.6, 0.05)
Model	Tab Transformer
epochs	qrandint(1, 40, 5)
batch_size	pow(2, randint(5, 12))
learning_rate	qrandint(1, 100, 4)*1e-4
optimizer_name	choice(['adam', 'adamw'])
loss_criterion	choice(['cross_entropy', 'weighted_cross_entropy'])
activation_fn	choice(['relu', 'tanh', 'gelu'])
dropout	qloguniform(0.05, 0.6, 0.05)
dim	randint(4, 9)
depth	randint(0, 6)
heads	qrandint(1, 9, 2)

Table 3. Hyperparameter tuning for classification models

## A.2. Ranking

Parameter Name	Search Space
Model	MLP
epochs	qrandint(1, 40, 5)
batch_size	pow(2,randint(5, 12))
learning_rate	qrandint(1, 100, 4)*1e-4
optimizer_name	choice(['adam', 'adamw'])
loss_criterion	choice(['cross_entropy', 'weighted_cross_entropy'])
num_layers	randint(1, 7)
layer_type	choice(['linear', 'highway'])
activation_fn	choice(['relu', 'tanh', 'gelu'])
dropout	qloguniform(0.05, 0.6, 0.05)

Table 4. Hyperparameter tuning for MLP with ranking objective.

## A.3. Unsupervised

Hyper Parameter Name	Value
Model	TabTransformer using ELECTRA
batch_size	256
learning_rate	0.0005
generator_dim	128
generator_depth	12
generator_heads	4
generator_dropout	0.1
discriminator_dropout	0.1
discriminator_dim	128
discriminator_depth	12
discriminator_heads	16
model_mask_prob	0.15
opt_warmup_steps	10000
opt_num_training_steps	1000000
generator_ff_mult	2
discriminator_ff_mult	4
discriminator_dim_head	64
generator_dim_head	64

Table 5. Hyperparameters used for training a TabTransformer on categorical features for RTD task.