



OptiML Cheat Sheet

Weights

Sampling



OptiML Configuration

OptiML Configuration Options

| Option | Description | Default | API Name |
|------------------------------|--|-----------------------------|-------------------|
| Objective field | The field you want to predict. It can be a categorical or numeric field. | Last valid field in dataset | objective_field |
| Maximum Training Time | The maximum time in minutes to limit the OptiML runtime. | 1,800 | max_training_time |
| Model candidates | Allows you to select the maximum number of different models (i.e., models using a unique configuration) to be trained and evaluated during the OptiML process. The default is 128 which is usually enough to find the best model, but you can set it from 4 up to 200. | 128 | model_candidates |
| Models to optimize | Allows you to select the type of models (or algorithms) that you want to optimize: decision trees, ensembles (up to 256 trees, including boosted trees and Random Decision Forests), logistic regressions (only for classification problems), and deepnets. | all | model_types |
| Evaluation | Allows you to select if you want to perform Monte Carlo cross-validation or select a specific test dataset to evaluate your models during the optimization process. | cross-validation | test_dataset |
| Optimization metric | Allows you to select the metric used for model selection during the optimization process. For classification problems, BigML uses the maximum phi coefficient by default, and for regression problems, the R squared. | max_phi and r_squared | metric |
| Positive class | Allows you to select one class to be optimized; otherwise, the average metric for all classes will be optimized. | all classes | metric_class |

| Option | Description | Default | API Name |
|--------------------------|--|---------|-------------------|
| Objective weights | Sets a specific weight for each class of the objective field. If a class is not listed, it is assumed to have a weight of 1. Weights of 0 are also valid. This option is only available for classification models. | False | objective_weights |
| Weight field | Sets instance weights using the values of the given field. The selected field must be numerical and it must not contain missing values. This is valid for both regression and classification models. | False | weight_field |

Missing values

| Option | Description | Default | API Name |
|------------------------------|---|---------|-----------------------|
| Missing splits | Tells whether to consider missing data as a split criterion. Only valid for decision trees and ensembles. | False | missing_splits |
| Default numeric value | Replaces missing numeric values in your dataset by the field's maximum, mean, median, minimum, or zero. If you do not activate this option or Missing numerics option, your instances with missing numeric values will be ignored for logistic regressions and deepnets. | Null | default_numeric_value |
| Missing numerics | Allows the logistic regression and deepnets to consider missing values for the numeric fields as valid values. If you do not activate this option or set a Default numeric value , your instances with missing numeric values will be ignored. Only valid for logistic regressions and deepnets. | True | missing_numerics |

| Option | Description | Default | API Name |
|-------------|--|---------|-------------|
| Rate | Sets the proportion of the dataset you want to consider between 0% and 100%. | 100% | sample_rate |