



# BigML Evaluation Cheat Sheet

## Sampling

## Ordering



## Evaluation Configuration

| Option             | Description  | Default                   | API Name    |
|--------------------|--|---------------------------|-------------|
| <b>Rate</b>        | Sets the proportion of the dataset you want to consider between 0% and 100%.   | 100%                      | sample_rate |
| <b>Range</b>       | Specifies a subset of instances from which to sample, e.g., from instance 5 to instance 1,000. The <b>Rate</b> you set will be computed over the <b>Range</b> configured.  | (1, max. rows in dataset) | range       |
| <b>Sampling</b>    | Allows you to choose between a random sampling or a deterministic sampling. When using deterministic sampling the random-number generator will always use the same seed, producing repeatable results.   | Random                    | seed        |
| <b>Replacement</b> | Allows a single instance to be selected multiple times. Sampling without replacement ensures that each instance cannot be selected more than once.   | False                     | replacement |
| <b>Out of bag</b>  | Selects only the out-of-bag instances for the currently defined sample. If an instance is not selected as part of a sample, it is considered out of bag. It is only selectable when a sample is deterministic and the sample rate is less than 100%. | False                     | out_of_bag  |

## Operating kinds for Ensembles

| Option             | Description   | Default | API Name                    |
|--------------------|---|---------|-----------------------------|
| <b>Probability</b> | Averages the per-class probability distributions for all trees in the ensemble and predicts the class with higher probability. For regression ensembles, the global prediction is the mean of the individual predictions.                               | True    | operating_kind: probability |
| <b>Confidence</b>  | Averages the per-class confidences distributions for all trees in the ensemble and predicts the class with higher confidence. For regression ensembles, the global prediction is the mean of the individual predictions weighted by the expected error. | False   | operating_kind: confidence  |
| <b>Votes</b>       | Gives one vote to each model in the ensemble. For classification models, the category with the majority of votes wins. For regression models, the global prediction is the mean of the individual predictions.  | False   | operating_kind: votes       |

## Missing Strategies

| Option                 | Description  | Default | API Name           |
|------------------------|--|---------|--------------------|
| <b>Last prediction</b> | Returns the parent node prediction when a missing value is found in the testing data for a decision node.  | True    | missing_strategy:0 |
| <b>Proportional</b>    | Recombines the subtrees predictions based on the proportion of data in each subtree when a missing value is found in the input data for a decision node. | False   | missing_strategy:1 |