



BigML Association Discovery Cheat Sheet



Association Discovery configuration

Minimum Levels for Measures

| Option | Description | Default | API Name |
|---------------------------|--|---------|--------------------|
| Minimum support | Sets a level of support between 0% and 100%. Associations below this support will be discarded. | 0% | min_support |
| Minimum confidence | Sets a confidence between 0% and 100%. Associations below this confidence will be discarded. | 0% | min_confidence |
| Minimum leverage | Sets a leverage between -100% and 100%. Associations below this leverage will be discarded. | 0% | min_leverage |
| Minimum lift | Sets any positive real number. Associations below this lift will be discarded. | 1 | min_lift |
| Significance level | Sets the maximum level of risk you are willing to take to discover a spurious association. Statistical tests are applied to control the risk of finding spurious associations. | 0.05 | significance_level |

Sampling

| Option | Description | Default | API Name |
|--------------------|--|-------------------|-------------|
| Rate | Sets the proportion of the dataset you want to consider between 0% and 100%. | 100% | sample_rate |
| Range | Specifies a subset of instances from which to sample, e.g., from instance 5 to instance 1,000. The Rate you set will be computed over the Range configured. | (1, max. dataset) | range |
| Sampling | 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 |
| Replacement | Allows a single instance to be selected multiple times. Sampling without replacement ensures that each instance cannot be selected more than once. | False | replacement |
| Out of bag | Selects only the out-of-bag instances for the currently defined sample. If an instance is not selected as part of a sampling, 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 |

Discretization

| Option | Description | Default | API Name |
|---------------|---|------------|----------|
| Pretty | Sets segment boundaries for numeric fields, so they are easy to read. For example, instead of <code>Pretty</code> <code>segment >= 20.678</code> you will get <code>Pretty</code> <code>segment >= 20</code> . If <code>Pretty</code> is enabled, the specified <code>Size</code> may act as a maximum. | True | pretty |
| Size | Sets the number of equal segments. You can set up to 50 segments. If <code>Pretty</code> is enabled this value acts as a maximum size. | 5 | size |
| Trim | Sets the portion of the overall population that may be removed from either tail of the distribution. You can set a number between 0% and 10%. A trim of 1% usually gives good results. | 0% | trim |
| Type | Sets whether the field is discretized using an equal width or equal population strategy for each segment. | Population | type |

Configuration Options

| Option | Description | Default | API Name |
|---------------------------------|--|----------|-----------------|
| Max. number (k) | Sets the maximum number of associations to be discovered. Higher numbers may take longer to calculate. You can set any value between 1 and 500. | 100 | max_k |
| Max. items in antecedent | Sets the maximum number of items to be considered within the antecedent itemset. You can set values between 1 and 10. The consequent itemset will always contain one item. | 4 | max_lhs |
| Search strategy | Selects the measure to prioritize the associations discovered. Leverage is one of the measures that gives relevant results in most cases. Two other measures frequently used are confidence and lift. The strategy chosen should be coherent with your application. | Leverage | search_strategy |
| Complementary items | Takes complementary items into account. For example, if there is an item <code>coffee</code> , its complement (<code>NOT coffee</code>) may also be included in some of the discovered associations. Complementary items will be represented with an exclamation point (<code>coffee</code> → <code>!coffee</code>). | False | complement |
| Missing items | Considers missing values to be valid items, which may appear in the discovered associations. | False | missing_items |