



BigML Fusion Cheatsheet



Fusion configuration

Fusion configuration options

| Option | Description | Default | API Name |
|----------------------|--|---------|----------------|
| Models | Select up to 1,000 models, ensembles, logistic regression, and/or deepnets to create your fusion. You can also use existing fusions to include them in a new fusion. | Null | models |
| Weights | Assigns different weights to the selected models before creating the fusion. | 1 | weights |
| Field mapping | Maps the same fields from two different models in case they have different names | Null | fields_mapping |



Prediction configuration

Operating kind options

| Option | Description | Default | API Name |
|--------------------|--|-------------|----------------|
| Probability | Averages the per-class probability distributions for all models in the fusion and predicts the class with higher probability. For regression fusions, the global prediction is the mean of the individual predictions. | probability | operating_kind |

Probability threshold

| Option | Description | Default | API Name |
|------------------------------|--|---------|-----------------|
| Probability threshold | A percentage between 0% and 100% that can be used with classification fusions so that they only return the positive class when the probability on the prediction is above the established threshold. | Null | operating_point |

Default Numeric Values

| Option | Description | Default | API Name | Option | Description | Default | API Name |
|------------------------------|---|---------|-----------------------|-----------------------|--|---------|----------------|
| Default numeric value | Replaces missing numeric values in your dataset by the field's maximum, mean, median, minimum, or zero. | Null | default_numeric_value | Output dataset | Defines whether a dataset with the results should be automatically created or not. | True | output_dataset |

Output file options

| Option | Description | Default | API Name |
|-------------------------|--|---------|---------------|
| Fields separator | Allows you to choose the best separator for your fields. | Comma | separator |
| New line | Sets the character to use as the line break in the generated csv file: "\n", "\r", "\r\n". | LF | newline |
| Show/hide fields | Shows or hides the rest of the fields in your output file. | True | output_fields |
| Headers | Shows or hides the names of your columns in the output file. | True | header |

| | | | |
|--------------------------------|---|----------------------|----------------------------------|
| Prediction column name | Sets the name for the objective field. By default BigML takes the name of the fusion's objective field. | Objective Field Name | prediction_name |
| Include probability | Includes an additional column with the probability for classification fusions or confidence (expected error) for regression fusions | False | probability_confidence |
| Probability column name | Sets the name you want for the probability or confidence (expected error) column. By default it is named "probability" or "confidence". | Probability | probability_name_confidence_name |

| | | | |
|---------------------------------|---|-------|-------|
| Single model predictions | Defines whether to include a column for each of the individual model predictions of the fusion. This will add a column per model, named <prediction_name>_n, where n is the position of the model in the model list in the fusion, starting at 1. | False | votes |
|---------------------------------|---|-------|-------|

| | | | |
|----------------------|---|-------|---------------|
| Probabilities | Includes a column for each of the objective field classes indicating their probabilities per instance predicted. This will add a column per field, named <objective_field_class> probability. | False | probabilities |
|----------------------|---|-------|---------------|

| | | | |
|--------------------|--|-------|------------|
| Importances | Defines whether the batch prediction includes a column for each of the field importances for the fusion predictions. There is a column per field, named <field_name> importance. | False | importance |
|--------------------|--|-------|------------|

Output Dataset