

Настройка назначений столбцов

- **Назначение столбца:**

- **Не выбрано** — поле не будет участвовать в обработке, но будет включено в результирующую выборку без изменений.
- **Входное** — значения поля будут использоваться как входные.
- **Входное (внешнее разбиение)** — появляется, если задан порт *Внешние диапазоны квантования*. Для значений входного поля применяются параметры настроенного внешнего разбиения.
- **Входное (замороженное)** — появляется после первого выполнения и обучения узла Конечные классы. При перенастройке или переобучении столбца будет пересчитана только статистика, а разбиение на конечные классы останется прежним для всех столбцов. Рекомендуется в случае, если пользователь не намерен изменять разбиение на конечные классы с учетом новых данных.
- **Выходное** — значения поля будут использоваться как выходные (целевые).

- **Настройки входного непрерывного поля:**

- **Предварительное квантование** — задает начальное число интервалов квантования входного признака (начальные классы) из которых, в соответствии с настроенными параметрами, будут сформированы конечные классы. Рекомендуется использовать ее при большом числе уникальных значений в непрерывном поле.
- **Включить верхние границы интервала** — верхнее значение границы будет включено в текущий интервал. Если флаг неактивен, то значение верхней границы интервала относится к следующему интервалу (т.е. включается в него в качестве нижней границы). Например, флаг активен: $10 < \dots \leq 20$ и флаг неактивен: $10 \leq \dots < 20$.
- **Количество интервалов** — количество интервалов предварительного квантования.

- **Настройки входного дискретного поля:**

- **Начальные классы в качестве разбиения** — уникальные значения (начальные классы) используются в качестве конечных классов. То есть количество конечных классов будет соответствовать количеству уникальных значений на входе, максимально возможное число — 1000.

- **Настройки выходного поля:**

- **Настраиваемое значение "события"** — требуется выбрать значение бинарной целевой переменной, которое является событием. Определяется целью и логикой решаемой задачи. Установка данного параметра влияет на интерпретацию результатов WoE-анализа. Рекомендуется в качестве события всегда назначать редкий класс (такой вариант предлагается по умолчанию).

- **Настройки внешнего разбиения:**

- **Идентификатор внешнего разбиения** — появляется, если выбрано назначение *Входное (внешнее разбиение)*, где разбиение берется из таблицы, а не рассчитывается алгоритмом на

основе текущих данных (см. Настройка внешнего разбиения).

- **Настройки алгоритма:**

- **Минимальная доля класса, %** — это отношение числа наблюдений, для которых значение входного признака попадает в данный класс, к общему числу наблюдений исходного набора данных (вес класса). По умолчанию устанавливается 5%. Формирование классов с весом меньше заданного не допускается. Малый вес класса указывает на его низкую значимость и целесообразность объединения с некоторым другим классом.
- **Максимальное количество классов** — наибольшее допустимое количество классов, создаваемое узлом для столбца. По умолчанию устанавливается значение 5. Это значение можно изменить: большое число классов приводит к уменьшению их веса, а малое — к уменьшению информационного индекса. Число сформированных классов может оказаться меньше заданного значения из-за ограничения по весу классов.
- **Равномерность** — определяет способ разбиения диапазона изменения входного признака на классы (интервалы). По умолчанию параметр равен 0, и может изменяться от 0 до 1. Значение, равное 0, обеспечивает такое разбиение на классы, которое максимизирует информационный индекс. При значении параметра, равном 1, алгоритм формирует классы таким образом, чтобы в каждый из них попало примерно одинаковое число наблюдений. Таким образом, в первом случае повышается значимость признака, а во втором — интерпретируемость конечных классов.