



Конечные классы

Визуализатор Конечные классы предназначен для следующих задач:

- преобразование непрерывных и дискретных входных полей, используемых для построения моделей бинарной классификации, путем квантования на основе метода совокупности доказательств или WoE-анализа (weights of evidence, WoE);
- сокращение размерности данных за счет исключения признаков с низкой значимостью, снижения разнообразия значений признаков;
- восстановление пропусков, когда пропуски образуют отдельную метку интервала квантования или объединяются с соседним, близким по значению WoE-индекса;
- борьба с выбросами и экстремальными значениями — формирование меток интервала квантования при дискретизации непрерывного поля или объединение редких уникальных значений в одну категорию;
- упрощение описания исследуемых объектов.

Визуализатор включает в себя:

- Список входных столбцов;
- Область начальных классов;
- Область конечных классов.

Примечание: визуализатор *Конечные классы* может использоваться только с узлом Конечные классы.

Интерфейс

Операции

Для *Списка входных столбцов* доступна операция  **Скрыть боковую панель**. Нажатие на кнопку скрывает панель со списком входных столбцов, повторное нажатие возвращает панель.

Область конечных классов можно представить в виде таблицы  **Показать таблицу конечных классов** (*Alt+One*) или в виде диаграмм  **Показать диаграммы конечных классов** (*Alt+Two*).

Также имеется переключатель  **Доли событий** /  **Количество событий** — для выбора отображения абсолютных и относительных значений.

Список входных столбцов

- Статус поля:
 - **Разморожено** — показывает, что данное поле может быть использовано в процессе формирования конечных классов при переобучении модели.
 - **Заморожено** — показывает, что при переобучении данное поле не будет использоваться.

Примечание: статус поля изменить нельзя, он только отражает настройки в соответствующем узле Конечные классы.

- Метка столбца.
- Оцененный уровень значимости IV (см. рисунок 1).

	Столбец	IV
	9.0 Сумма кредита	 2,26
	9.0 Стоимость кредита	 2,26
	12 Срок кредита	 2,38
	31 Дата кредитования	 0,26
	ab Цель кредитован...	 0,13
	12 Возраст	 0,29
	ab Пол	 0,02
	ab Образование	 0,10

Рисунок 1. Список входных столбцов.

Область начальных классов

Начальные классы представлены в виде одного варианта отображения — диаграммы.

При наведении курсора мыши на диаграмму в правом верхнем углу появляется кнопка , нажав на которую, диаграмма развернется на всю область визуализатора.

Диаграмма начальных классов может представлять:

- Долю;
- WoE (вес доказательства);
- IV (информационный индекс).

Выбор отображения переключается кнопками (см. рисунок 2) в нижней части диаграммы.

■ Доля ■ WoE ■ IV

Рисунок 2. Вариант отображения диаграммы.

При увеличении диаграммы во всю область визуализатора, появляется выпадающий список, в нем можно выбрать поле, для которого будет построена диаграмма (см. рисунок 3).

При наведении на столбец диаграммы появляется всплывающее окно, которое отображает значение поля и значение WoE/IV/Доли (см. рисунок 3).

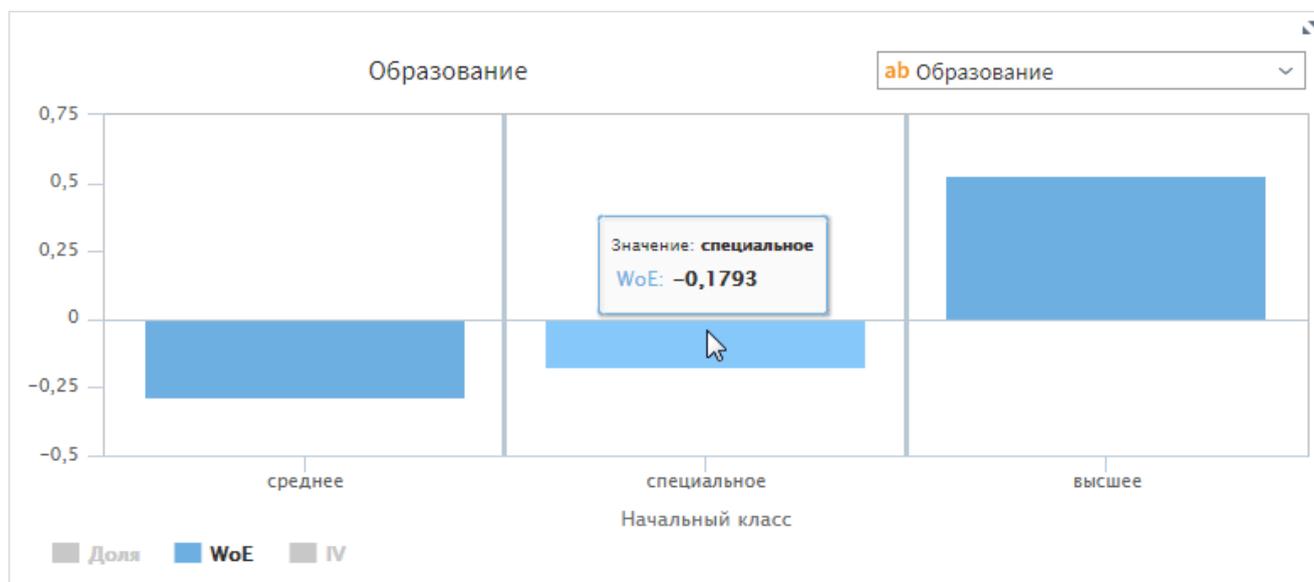


Рисунок 3. Диаграмма начальных классов.

Область конечных классов

Конечные классы отображаются в виде таблицы или диаграммы WoE (см. рисунок 4) / диаграммы IV.

Диаграмма IV может отображать:

- Количество (см. рисунок 5);
- Доля (см. рисунок 6).

Таблица

При установке переключателя в положение *Количество событий*, таблица будет состоять из следующих полей:

Поле	Описание
№	Номер класса

Поле	Описание
Метка	Совокупность названий начальных классов, которые входят в конечный класс
Нижняя	Нижняя граница интервала
Верхняя	Верхняя граница интервала
События	Одно из состояний бинарной целевой переменной класса
Не-события	Противоположное состояние бинарной целевой переменной класса
Всего	Сумма событий и не-событий класса
Доля	Доля конечного класса от общего объема записей
Вес доказательства	Коэффициент WoE
<u>Инф. индекс</u>	Величина, определяющая значимость признака в модели <u>бинарной классификации</u>

В нижней части таблицы отображается сумма:

- Событий всех классов;
- Не-событий всех классов;
- Событий и Не-событий всех классов;
- Информационных индексов всех классов.

При установке переключателя в положение *Доли событий*, вместо полей *События* и *Не-события* будут отображаться поля *Доля событий* и *Доля не-событий*.

При этом в нижней части таблицы будет отображаться:

- Доля суммы Событий всех классов ;
- Доля суммы Не-событий всех классов;
- Сумма Событий и Не-событий всех классов;
- Сумма информационных индексов всех классов.

Диаграмма

При наведении курсора мыши на диаграмму в правом верхнем углу появляется кнопка  , позволяющая развернуть диаграмму на всю область визуализатора.

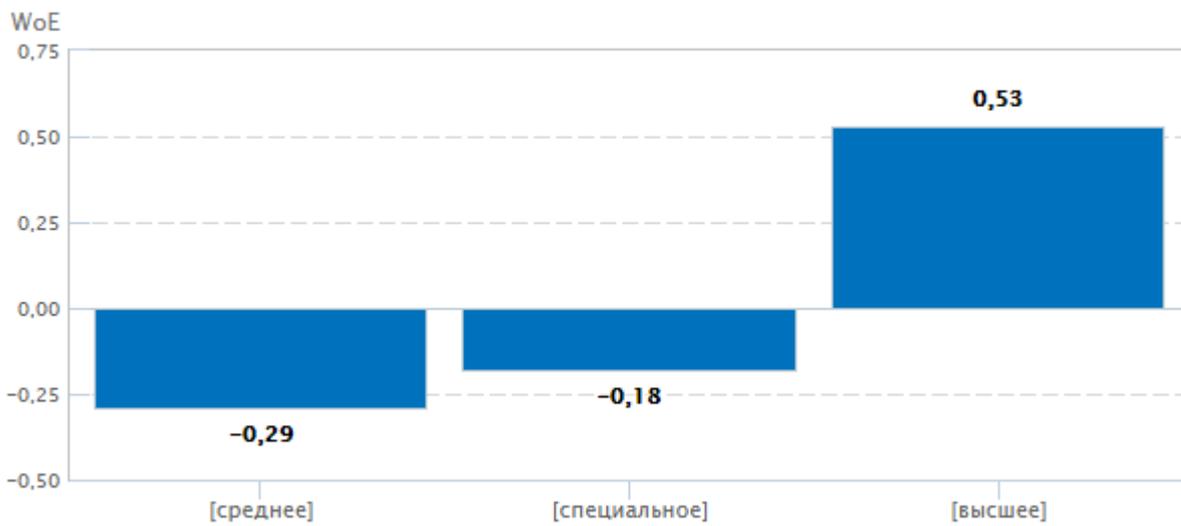


Рисунок 4. Диаграмма WoE.

Диаграммы Количество и Доля переключаются между собой кнопками .

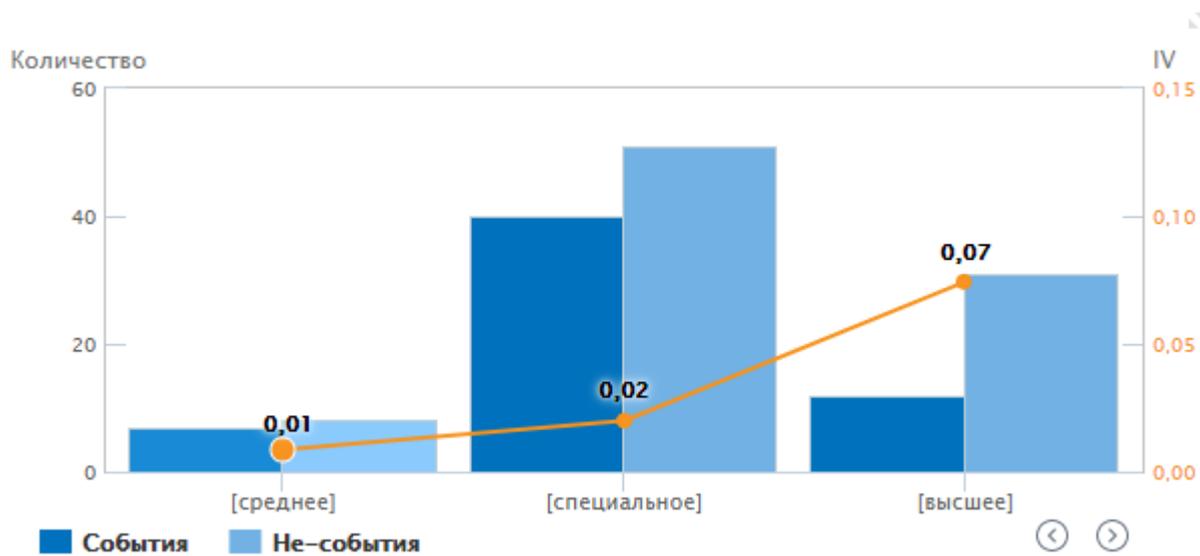


Рисунок 5. Диаграмма Доля.

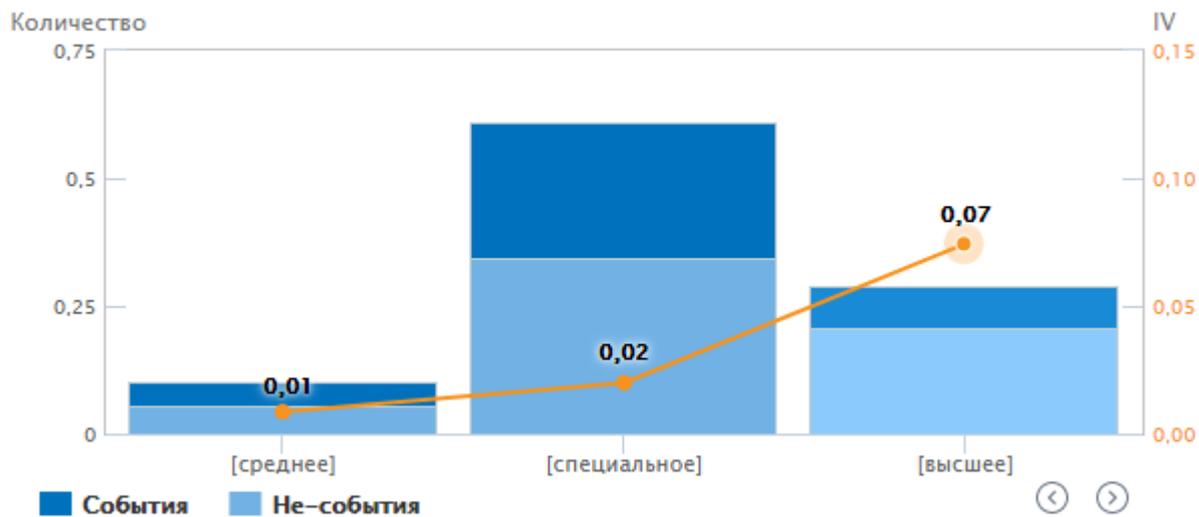


Рисунок 6. Диаграмма Количество.

При наведении курсора мыши на столбец диаграммы всплывает окно (см. рисунок 7), в котором отображается:

- Для диаграммы WoE:
 - значение WoE.
- Для диаграммы IV:
 - значение IV.
 - количество Событий.
 - количество Не-событий.

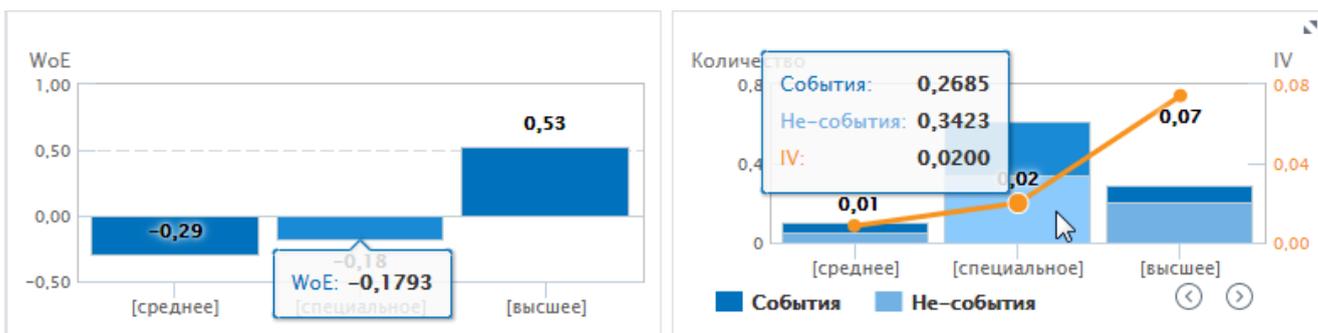


Рисунок 7. Диаграммы WoE и IV.