

Loginom Silver Kit

- Версия: 3.1.0
- Проверено: Все редакции Loginom 7.1.0

Библиотека **Loginom Silver Kit** поддерживается и развивается **Loginom Company**. В ней представлены простые и полезные компоненты, облегчающие рутинный труд аналитика по разработке сценариев обработки данных. Компоненты библиотеки широко используются в электронных учебных курсах компании и библиотеках компонентов к бизнес-курсам.

Установка

1. Определите рабочий каталог, где будут расположены ваши библиотеки:
 - Для серверных редакций — в рабочем каталоге Loginom Server (в папке пользователя или в общей папке пользователей);
 - Для настольных редакций — в любой папке на локальном диске.
2. Создайте в нем подкаталог **libs**.
3. Разместите папку **silver_kit** в каталоге **libs**.
4. Добавьте ссылку на пакет **loginom_silver_kit.lgp** в своем пакете и используйте компоненты библиотеки.



Требования

Для работы библиотеки **Loginom Silver Kit** необходимо:

- Установленное ПО Loginom. Версия не ниже 7.1.0

ABC-анализ

[Назад к списку компонентов](#)



Назначение

В компоненте реализован метод, который позволяет классифицировать ресурсы по степени их важности на основе принципа Парето. Происходит разделение объектов на три группы (A, B, C) согласно границам групп. Границы групп задаются в переменных.

Входные порты

Название	Тип
Входной набор данных	Таблица
Груницы групп	Переменные

Структура таблицы "Входной набор данных"

Метка	Тип	Описание
Объект	 Строковый	Идентификатор или наименование объекта, по которому проводится анализ. Например, артикул товара
Показатель	 Вещественный	Значение показателя по объекту. Например, сумма продаж

Переменные в порте "Границы групп"

№	Метка	Тип	Значение
1	Граница группы А, %	90 Вещественный	80
2	Граница группы В, %	90 Вещественный	95

Границы групп устанавливаются пользователем. По умолчанию указаны границы, которые используются чаще всего.

Выходные порты

Название	Тип
ABC-группы и вклады объектов	Таблица

Структура таблицы "ABC-группы и вклады объектов"

Метка	Тип	Описание
Объект	ab Строковый	Идентификатор или наименование объекта
Показатель	90 Вещественный	Значение показателя по объекту
Вклад нарастающим итогом	90 Вещественный	Накопительный вклад объектов
Группа	ab Строковый	Идентификатор группы объекта

Алгоритмы

- ABC-анализ

АВС-анализ (метод касательных)

[Назад к списку компонентов](#)



Назначение

Аналогичен компоненту **АВС-анализ**, но границы групп определяются автоматически методом касательных. Далее объекты распределяются на три группы (А, В, С) согласно рассчитанным границам групп.

Входные порты

Название	Тип
Входной набор данных	Таблица

Структура таблицы "Входной набор данных"

Метка	Тип	Описание
Объект	 Строковый	Идентификатор или наименование объекта, по которому проводится анализ. Например, артикул товара
Показатель	 Вещественный	Значение показателя по объекту. Например, сумма продаж

Выходные порты

Название	Тип
АВС-группы и вклады объектов	Таблица
Границы групп	Переменные

Структура таблицы "ABC-группы и вклады объектов"

Метка	Тип	Описание
Объект	ab Строковый	Идентификатор или наименование объекта
Показатель	90 Вещественный	Значение показателя по объекту
Вклад нарастающим итогом	90 Вещественный	Накопительный вклад объектов
Группа	ab Строковый	Идентификатор группы объекта

Переменные в порте "Границы групп"

№	Метка	Тип	Описание
1	Граница группы A, %	90 Вещественный	Рассчитанное значение границы группы A
2	Граница группы B, %	90 Вещественный	Рассчитанное значение границы группы B

Алгоритмы

- Метод касательных в ABC-анализе

XYZ-анализ

[Назад к списку компонентов](#)



Назначение

XYZ-анализ позволяет классифицировать ресурсы в зависимости от стабильности их потребления на основе оценки значений коэффициентов вариации. В компоненте происходит разделение объектов на три группы (X, Y, Z) согласно границам коэффициента вариации. Границы групп задаются в переменных.

Входные порты

Название	Тип
Входной набор данных	Таблица
Переменные	Переменные

Структура таблицы "Входной набор данных"

Метка	Тип	Описание
Объект	 Строковый	Идентификатор или наименование объекта, по которому проводится анализ. Например, артикул товара
Показатель	 Вещественный	Значение показателя по объекту. Например, сумма продаж

Переменные в порте "Границы групп"

№	Метка	Тип	Значение
1	Граница группы X, %	90 Вещественный	10
2	Граница группы Y, %	90 Вещественный	25

Границы групп устанавливаются пользователем. По умолчанию указаны границы, которые используются чаще всего.

Выходные порты

Название	Тип
XYZ-группы	Таблица

Структура таблицы "XYZ-группы"

Метка	Тип	Описание
Объект	ab Строковый	Идентификатор или наименование объекта
Группа	ab Строковый	Идентификатор группы объекта
Коэффициент вариации, %	90 Вещественный	Значение коэффициента вариации, %
Среднее значение	90 Вещественный	Среднее значение показателя для объекта

Алгоритмы

- XYZ-анализ — материал из Wikipedia

Выделение периода

[Назад к списку компонентов](#)

Назначение

Позволяет преобразовать значения в поле типа **Дата/время** в период. Период может обозначаться первой или последней его датой, возможно выделение периодов по ISO (месяц, неделя).

Как правило, компонент используется с целью последующей группировки по периоду с различными опциями.

Входные порты

Название	Тип
Входной набор данных	Таблица
Переменные	Переменные

Структура таблицы "Входной набор данных"

Метка	Тип	Описание
ID объекта	<div>ab</div> Строковый	Идентификатор объекта: клиента, товара, события и т.п.
Дата	<div>31</div> Дата/время	Дата, которую необходимо преобразовать

Переменные в порте "Переменные"

№	Метка	Тип	Значение
1	Тип периода	ab Строковый	m
2	Первый день	0/1 Логический	true

1. Допустимые типы периодов:

2. y — год;

3. q — квартал;

4. m — месяц (по умолчанию);

5. im — месяц по ISO;

6. w — неделя;

7. iw — неделя по ISO;

8. d — день.

9. **Первый день** — при значении **true** (по умолчанию) период в выходной таблице будет обозначаться его первой датой, иначе — последней датой.

Выходные порты

Название	Тип
Набор данных	Таблица

Структура таблицы "Набор данных"

Метка	Тип	Описание
ID объекта	ab Строковый	Идентификатор объекта
Дата	31 Дата/время	Исходная дата
Период	31 Дата/время	Период, к которому относится исходная дата

Заполнение периодов

[Назад к списку компонентов](#)

Назначение

Предназначен для добавления по объекту отсутствующих периодов (между первым и последним периодами) для создания полного временного ряда.

Использует компоненты [Генератор календаря](#) и [Выделение периода](#).

Входные порты

Название	Тип
Входной набор данных	Таблица
Переменные	Переменные

Структура таблицы "Входной набор данных"

Метка	Тип	Описание
ID объекта	ab Строковый	Идентификатор объекта: клиента, товара, события и т.п
Период	31 Дата/время	Список периодов по каждому объекту

Переменные в порте "Переменные"

№	Метка	Тип	Значение
1	Тип периода	ab Строковый	m
2	Первый день	0/1 Логический	true

1. Допустимые типы периодов:
2. `y` — год;
3. `q` — квартал;
4. `m` — месяц (по умолчанию);
5. `im` — месяц по ISO;
6. `w` — неделя;
7. `iw` — неделя по ISO;
8. `d` — день.
9. **Первый день** — при значении **true** (по умолчанию) период в выходной таблице будет обозначаться его первой датой, иначе — последней датой.

Выходные порты

Название	Тип
Набор данных	Таблица

Структура таблицы "Набор данных"

Метка	Тип	Описание
ID объекта	ab Строковый	Идентификатор объекта
Период	31 Дата/время	Список периодов по объекту (полный временной ряд)

Редкие значения

[Назад к списку компонентов](#)

Назначение

Поиск редких значений полей, формирование таблицы замен с учетом пороговой частоты для редких значений и условия замены единичных значений.

Входные порты

Название	Тип
Входной набор данных	Таблица
Переменные	Переменные

Структура таблицы "Входной набор данных"

Метка	Тип	Описание
Идентификатор	 Переменный	Идентификатор объекта
Поле	 Строковый	Наименование поля, к которому относится значение
Значение	 Переменный	Значение поля

Переменные в порте "Переменные"



№	Метка	Тип	Значение
1	Пороговое значение	90 Вещественный	0,0125
2	Замена единичных значений	0/1 Логический	false
3	Заменить на	ab Строковый	Прочее

1. **Пороговое значение** — пороговое значение частоты встречаемости, ниже которого значение считается редким. Задается пользователем.
2. **Замена единичных значений** — определяет, заменить ли значение в поле, если в нем найдено только одно редкое значение (**true**), или сохранить список уникальных значений в поле (**false**, по умолчанию).
3. **Заменить на** — значение, на которое будут заменены редкие значения. Задается пользователем.


Выходные порты

Название	Тип
Набор данных	Таблица
Замененные значения	Таблица
Таблица замен	Таблица

Структура таблицы "Набор данных"

Метка	Тип	Описание
Идентификатор	 Переменный	Идентификатор объекта
Поле	ab Строковый	Наименование поля, к которому относится значение
Исходное значение	 Переменный	Значение, которое было на входе
Новое значение	 Переменный	Если исходное значение признано редким, указывается значение переменной Заменять на
Заменено	Логический	Флаг, показывающий, была ли проведена замена

Структура таблицы "Замененные значения"

Метка	Тип	Описание
Поле	ab Строковый	Поле, значение которого было заменено
Исходное значение	 Переменный	Значение, которое было заменено
Количество замен	 Целый	Количество раз, котрое значение было заменено во входном наборе

Структура таблицы "Таблица замен"

Метка	Тип	Описание
Поле	 Строковый	Поле, значение которого было заменено
Исходное значение	 Переменный	Значение, которое было заменено
Новое значение	 Переменный	Значение, на которое оно было заменено

Алгоритмы

Для каждого уникального значения в конкретном поле рассчитывается доля объектов. Все значения объектов, доля которых ниже порога, заменяются на строковое значение, указанное в переменной **Заменять на** (по умолчанию **Прочее**).

Разбиение на пачки

[Назад к списку компонентов](#)

Назначение

Компонент предназначен для разбиения набора данных на пачки заданного размера. Применяется, когда требуется организовать выполнение сценария или фрагмента сценария по частям.

Применяется простое разбиение — подсчет количества строк по каждому идентификатору и их суммирование, пока не достигнут заданный в переменной размер пачки.

Входные порты

Название	Тип
Идентификатор объекта	Таблица
Переменные	Переменные

Структура таблицы "Идентификатор объекта"

Метка	Тип	Описание
Идентификатор объекта	ab Строковый	Поле, по которому происходит разбиение

Обратите внимание: для поля **Идентификатор объекта** необходимо подавать на вход реальное количество записей, в котором присутствует каждый идентификатор (НЕ список уникальных значений), т.к. разбиение будет производиться на основе расчета количества записей по каждому объекту.

Переменные в порте "Переменные"

№	Метка	Тип	Значение
1	Кол-во строк в пачке	12 Целый	100

Кол-во строк в пачке — максимально допустимый размер пачки, задается пользователем.

Выходные порты

Название	Тип
Набор данных	Таблица
Границы пачек	Таблица

Структура таблицы "Набор данных"

Метка	Тип	Описание
Идентификатор объекта	ab Строковый	Уникальный идентификатор объекта
Номер пачки	12 Целый	Номер пачки, в которую попал объект

Структура таблицы "Границы пачек"

Метка	Тип	Описание
Номер пачки	12 Целый	Уникальный номер пачки
Идентификатор объекта: первый	ab Строковый	ID первого объекта, попавшего в пачку (сортировка по возрастанию по строковому полю)
Идентификатор объекта: последний (ObjectIDLast)	ab Строковый	ID последнего объекта, попавшего в пачку (сортировка по возрастанию по строковому полю)

Разделитель строк

[Назад к списку компонентов](#)

Назначение

Позволяет разбить на несколько полей значение строки, записанное с разделителями. Учитывается только один разделитель (задается в переменной). На выходе получается динамическое количество полей (в зависимости от максимального количества частей входных строк).

Входные порты

Название	Тип
Входной набор данных	Таблица
Переменные	Переменные

Структура таблицы "Входной набор данных"

Метка	Тип	Описание
Входная строка	ab Строковый	Строки с разделителями, значения которых нужно разбить по разным полям

Переменные в порте "Переменные"

Метка	Тип	Значение
Разделитель	ab Строковый	,

Разделитель — указывается разделитель, который присутствует во входной строке. На его основе определяются части строки и их количество.

Выходные порты

Название	Тип
Набор данных	Таблица

Структура таблицы "Набор данных"

Метка	Тип	Описание
Входная строка	ab Строковый	Исходная строка
n	ab Строковый	Остальные поля содержат части исходной строки, n — номер части, начиная с 1 (записывается в метку поля)

Разделитель строк (переменные)

[Назад к списку компонентов](#)

Назначение

Позволяет разбить на несколько строк значение переменной, записанное с разделителями. Учитывается только один разделитель (задается в переменной). На выходе получается динамическое количество строк. Компонент по алгоритму разбора строки аналогичен компоненту **Разделитель строк**.

Входные порты

Название	Тип
Переменные	Переменные

Структура таблицы "Входной набор данных"

Метка	Тип	Описание
Входная строка	ab Строковый	Строка с разделителями, значения которой нужно разбить по разным строкам

Переменные в порте "Переменные"

Метка	Тип	Значение
Входная строка	ab Строковый	null
Разделитель	ab Строковый	,

Разделитель — указывается разделитель, который присутствует во входной строке. На его основе определяются части строки и их количество.

Выходные порты

Название	Тип
Набор данных	Таблица

Структура таблицы "Набор данных"

Метка	Тип	Описание
ID части строки	ab Строковый	Номер части исходной строки
Часть строки	ab Строковый	Содержит части исходной строки

Разметка событий

[Назад к списку компонентов](#)

Назначение

Разметка событий для списка объектов во временном окне транзакций. Это полезно, когда, например, нужно разметить бинарное поле с визитами клиентов в магазин в течение месяца.

Для входных объектов и временного окна, заданного датами, формируется поле **Событие**, которое равно `0` (`FALSE`), если в транзакциях не встретилось ни одной транзакции по данному объекту и `1` (`TRUE`), если встретилась хотя бы одна транзакция. Временное окно задается двумя датами, **Дата начала** и **Дата окончания**, попадание в окно проверяется условием `>= Дата начала` и `< Дата окончания` .

Входные порты

Название	Тип
Список объектов	Таблица
Транзакции	Таблица
Переменные	Переменные

Структура таблицы "Список объектов"

Метка	Тип	Описание
Объект	ab Строковый	Уникальный идентификатор объекта

Структура таблицы "Транзакции"

Метка	Тип	Описание
Объект	ab Строковый	Уникальный идентификатор объекта
Дата транзакции	31 Дата/Время	Дата и время транзакции

Переменные в порте "Переменные"

№	Метка	Тип	Значение
1	Дата начала	31 Дата/Время	null
2	Дата окончания	31 Дата/Время	null

1. **Дата начала** — Первая дата, начиная с которой начинаем наблюдать событие (включая эту дату).
2. **Дата окончания** — Вторая дата, на которой заканчиваем наблюдать событие (не включая эту дату).

Если даты не заданы (по умолчанию), то наблюдение события ведется во всех транзакциях.

Выходные порты

Название	Тип
События	Таблица

Структура таблицы "События"

Метка	Тип	Описание
Объект	ab Строковый	Уникальный идентификатор объекта
Событие	12 Целый	Событие 0/1
Событие	0/1 Логический	Событие TRUE/FALSE

Сэмплинг объектов

[Назад к списку компонентов](#)


Назначение

Отбирает подмножество объектов заданного размера из множества транзакций.

Входные порты

Название	Тип
События объектов	Таблица
Переменные	Переменные

Структура таблицы "События объектов"

Метка	Тип	Описание
Объект	 Строковый	Идентификатор или наименование объекта

Переменные в порте "Переменные"

№	Метка	Тип	Значение
1	Размер в процентах	 Вещественный	10,00

Размер в процентах — процент объектов, который необходимо вывести на выходе.

Выходные порты

Название	Тип
Выборка объектов	Таблица

Структура таблицы "События объектов"

Метка	Тип	Описание
Объект	ab Строковый	Список объектов (заданный процент от всех объектов)

Популярные объекты

[Назад к списку компонентов](#)

Назначение

Формирование списка популярных объектов с объединением остальных в одну категорию.

Входные порты

Название	Тип
Входной набор данных	Таблица
Переменные	Переменные

Структура таблицы "Входной набор данных"

Метка	Тип	Описание
Объект	ab Строковый	Наименование объекта
Показатель	90 Вещественный	Показатель объекта

Переменные в порте "Переменные"

№	Метка	Тип	Значение
1	Граница популярных объектов	12 Целый	10
2	Тип расчета	ab Строковый	top-k
3	Заменить на	ab Строковый	Прочее

- Граница популярных объектов** — число или процент объектов, которые будут в списке популярных.

2. **Тип расчета** — может принимать значения:
- *top-k* (по умолчанию) — выделить число популярных объектов от общего количества;
 - *top-%* — выделить процент популярных объектов от общего количества.
3. **Заменить на** — новое наименование, на которое будут заменены объекты не попавшие в список популярных.

Выходные порты

Название	Тип
Набор данных	Таблица
Популярные объекты	Таблица
Замененные объекты	Таблица

Структура таблицы "Набор данных"

Метка	Тип	Описание
Исходный объект	 Строковый	Наименование объекта из входного набора данных
Новый объект	 Строковый	Новое наименование замененного объекта
Показатель	 Вещественный	Показатель объекта
Заменено	 Логический	true — наименование объекта заменено, false — замена не осуществлялась

Структура таблицы "Популярные объекты"

Метка	Тип	Описание
Объект	ab Строковый	Наименование объекта, включая замененные
Показатель	90 Вещественный	Показатель объекта
Заменено	0/1 Логический	true — наименование объекта заменено, false — замена не осуществлялась
Количество объектов	12 Целый	Количество замененных и не замененных объектов

Структура таблицы "Замененные объекты"

Метка	Тип	Описание
Исходный объект	ab Строковый	Наименование замененного объекта из входного набора данных
Новый объект	ab Строковый	Новое наименование замененного объекта
Показатель	90 Вещественный	Показатель объекта

AUC

[Назад к списку компонентов](#)

Назначение



Предназначен для расчета индекса AUC, стандартной ошибки, доверительных интервалов, Z-оценки для оценки качества модели.

Присутствует словесная оценка качества модели. Результаты выводятся в переменных.

Входные порты

Название	Тип
Фактические и оцененные значения	Таблица
Параметры алгоритма	Переменные

Структура таблицы "Фактические и оцененные значения"

Метка	Тип	Описание
Событие	 Логический	true соответствует событию, false — не-событию
Оценка	 Вещественный	Оценка, которую выдал алгоритм (вероятность, балл, рейтинг, достоверность и т.д.)

Переменные в порте "Параметры алгоритма"

№	Метка	Тип	Значение
1	Точность	12 Целый	4

Точность — точность вычисления выходных значений (знаков после запятой).
Влияет на скорость расчетов: чем выше точность, тем дольше вычисления.

Выходные порты

Название	Тип
Результаты	Переменные

Переменные в порте "Результаты"

№	Метка	Тип	Описание
1	Индекс AUC	9.0 Вещественный	Расчитанное значение индекса AUC для модели
2	Качество	ab Строковый	Словесное описание качества модели (см. Алгоритмы)
3	Стандартная ошибка	9.0 Вещественный	Стандартное отклонение по набору
4	Доверительный интервал нижний	9.0 Вещественный	Нижняя граница доверительного интервала
5	Доверительный интервал верхний	9.0 Вещественный	Верхняя граница доверительного интервала
6	Z-оценка	9.0 Вещественный	Расчитанное значение Z-оценки

Алгоритмы

Индекс AUC

Используется стандартный алгоритм расчета площади под кривой. Кумулятивные доли событий и не-событий рассчитываются для каждого изменения выходной переменной **Оценка**, предварительно округленной до (**Accuracy + 1**) знаков после запятой.

Качество

Текстовое поле формируется на основе следующих правил.

Интервал AUC	Качество модели
> 0,9	Отличное
[0,8 - 0,9]	Очень хорошее
(0,7 - 0,8]	Хорошее
(0,6 - 0,7]	Среднее
<= 0,6	Неудовлетворительное

Стандартное отклонение

Рассчитывается методом [Hanley & McNeil, 1982]¹.

Расчеты имеют вид:

$$Q_1 = \frac{AUC}{2 - AUC}$$

$$Q_2 = \frac{2AUC^2}{1 + AUC}$$

$$SE = \sqrt{\frac{AUC(1 - AUC) + (na - 1)(Q_1 - AUC^2) + (nn - 1)(Q_2 - AUC^2)}{na \cdot nn}}$$

Доверительный интервал

Границы доверительного интервала для AUC рассчитываются следующим образом:

$$CI = AUC \pm 1,96SE$$

Z-оценка

Расчеты имеют вид:

$$Z = \frac{AUC - 0,5}{SE}$$

Критическое значение z-оценки для уровня значимости 95% принимается равным 1,96. Если $|z| > 1,96$, то гипотеза о случайности оценки площади под кривой ($AUC = 0,5$) отвергается.

Дополнительная литература

1. The meaning and use of the area under a receiver operating characteristic (ROC) curve
2. The magnificent ROC(Receiver Operating Characteristic curve)
3. Details of calculations for ROC curves

ROC-анализ

[Назад к списку компонентов](#)

Назначение

Оценка качества модели оценки вероятности некоторого события и расчет различных метрик, принятых в ROC-анализе:

- расчет чувствительности, специфичности, и других показателей ROC-анализа для каждого значения порога отсечения, заданного с указанной точностью;
- расчет коэффициента KS (статистика Колмогорова-Смирнова);
- выбор порога отсечения с помощью различных методов.

Входные порты

Название	Тип
Входной набор данных	Таблица
Параметры алгоритма	Переменные

Структура таблицы "Входной набор данных"

Метка	Тип	Описание
Событие	0/1 Логический	true — соответствует событию, false — соответствует не событию
Порог	9.0 Вещественный	Значение вероятности события

Переменные в порте "Параметры алгоритма"

№	Метка	Тип	Значение
1	Точность	12 Целый	4
2	Уровень чувствительности	90 Вещественный	90
3	Цена ошибки - ложноположительные	90 Вещественный	1
4	Цена ошибки - ложноотрицательные	90 Вещественный	1
5	Метод выбора порога отсека	ab Строковый	3

1. **Точность** — количество знаков после запятой, до которого требуется округлить пороговое значение.
2. **Уровень чувствительности** — уровень чувствительности для выбора порога отсека по выбранному методу.
3. **Цена ошибки - ложноположительные/Цена ошибки - ложноотрицательные** — веса ошибок.
4. **Метод выбора порога отсека:**
5. Заданная чувствительность;
6. Чувствительность, равная проценту точности события;
7. Максимальное значение KS;
8. Минимум издержек ошибок классификации;
9. Максимум точности события.

Выходные порты

Название	Тип
Набор данных	Таблица
KS, %	Переменные
Порог отсека	Таблица

Структура таблицы "Набор данных"

Метка	Тип	Описание
Порог округлённый	9.0 Вещественный	Пороговое значение, округленное до заданной точности
Истинноположительные случаи	12 Целый	Количество истинноположительных (TP)
Ложноположительные случаи	12 Целый	Количество ложноположительных (FP)
Предсказанные положительные	12 Целый	TP+FP
Истинноотрицательные случаи	12 Целый	Количество истинноотрицательных (TN)
Ложноотрицательные случаи	12 Целый	Количество ложноотрицательных (FN)
Предсказанные отрицательные	12 Целый	TN+FN
Чувствительность, %	9.0 Вещественный	$\frac{TP}{(TP + FN)} \cdot 100$
Специфичность, %	9.0 Вещественный	$\frac{TN}{(TN + FP)} \cdot 100$
(Процент событий) - (Процент не-событий)	9.0 Вещественный	$\left(\frac{TP}{(TP + FN)} - \frac{FP}{(TN + FP)} \right) \cdot 100$
Изменение количества истинноположительных	12 Целый	Разница между текущим и следующим значением TP
Изменение количества ложноположительных	12 Целый	Разница между текущим и следующим значением FP
% ошибок	9.0 Вещественный	$\frac{FP + FN}{(TP + FP + TN + FN)} \cdot 100$
% ложноположительных	9.0 Вещественный	$\frac{FP}{(TN + FP)} \cdot 100$

Метка	Тип	Описание
% ложноотрицательных	9.0 Вещественный	$\frac{FN}{(TP + FN) \cdot 100}$
Издержки классификации	9.0 Вещественный	Ошибки, взвешенные через их цену
% точности классификации	9.0 Вещественный	$\frac{TP + TN}{(TP + TN + FP + FN) \cdot 100}$
% точности события	9.0 Вещественный	$\frac{TP}{(TP + FP) \cdot 100}$
% точности не-события	9.0 Вещественный	$\frac{TN}{(TN + FN) \cdot 100}$

Переменные в порте "KS, %"

№	Метка	Тип	Описание
1	KS, %	9.0 Вещественный	Значение статистики Колмогорова-Смирнова

Структура таблицы "Порог отсечения"

Метка	Тип	Описание
Метод выбора порога	ab Строковый	Наименование метода
Порог округлённый	9.0 Вещественный	Выбранное значение порога отсечения, округленное до заданной точности
Истинноположительные случаи	12 Целый	Количество истинноположительных (TP)
Ложноположительные случаи	12 Целый	Количество ложноположительных (FP)

Метка	Тип	Описание
Предсказанные положительные	12 Целый	TP+FP
Истинноотрицательные случаи	12 Целый	Количество истинноотрицательных (TN)
Ложноотрицательные случаи	12 Целый	Количество ложноотрицательных (FN)
Предсказанные отрицательные	12 Целый	TN+FN
Чувствительность, %	9.0 Вещественный	$\frac{TP}{(TP + FN)} \cdot 100$
Специфичность, %	9.0 Вещественный	$\frac{TN}{(TN + FP)} \cdot 100$
(Процент событий) - (Процент не-событий)	9.0 Вещественный	$\left(\frac{TP}{(TP + FN)} - \frac{FP}{(TN + FP)} \right) \cdot 100$
Изменение количества истинноположительных	12 Целый	Разница между текущим и следующим значением TP
Изменение количества ложноположительных	12 Целый	Разница между текущим и следующим значением FP
% ошибок	9.0 Вещественный	$\frac{FP + FN}{(TP + FP + TN + FN)} \cdot 100$
% ложноположительных	9.0 Вещественный	$\frac{FP}{(TN + FP)} \cdot 100$
% ложноотрицательных	9.0 Вещественный	$\frac{FN}{(TP + FN)} \cdot 100$
Издержки классификации	9.0 Вещественный	Ошибки, взвешенные через их цену
% точности классификации	9.0 Вещественный	$\frac{TP + TN}{(TP + TN + FP + FN)} \cdot 100$

Метка	Тип	Описание
% точности события	9.0 Вещественный	$\frac{TP}{(TP + FP) \cdot 100}$
% точности не-события	9.0 Вещественный	$\frac{TN}{(TN + FN) \cdot 100}$

Алгоритмы

Псевдокоды алгоритма ROC-анализа доступны в статье по [ссылке](#).

Дополнительная литература

1. ROC-кривая - материал из Wikipedia (рус.)
2. ROC-анализ - материал из Wikipedia (англ.)
3. Bradley, A. The use of the area under the ROC curve in the evaluation of machine learning algorithms — статья в журнале Pattern Recognition, июль 1997.

Динамика события

[Назад к списку компонентов](#)

Назначение

Вычисление накопительной доли событий по когортам, общей доли событий, определение точки стабилизации. Когорты формируются на основе периода регистрации.

Для компонента настроены отчеты:

- Доля событий по когортам;
- Доля событий.

Входные порты

Название	Тип
Входной набор данных	Таблица
Переменные	Переменные

Структура таблицы "Входной набор данных"

Метка	Тип	Описание
Идентификатор объекта	ab Строковый	Уникальный идентификатор объекта (например, код клиента)
Дата регистрации	31 Дата/Время	Дата появления объекта (например, заключения договора с клиентом)
Дата первого события	31 Дата/Время	Дата первого события с объектом (например, дата первой покупки). Может совпадать с датой регистрации
Дата второго события	31 Дата/Время	Дата второго события с объектом (например, дата второй покупки)

Переменные в порте "Переменные"

№	Метка	Тип	Значение
1	Тип когорты	ab Строковый	m
2	Тип периода	ab Строковый	m
3	Порог, %	90 Вещественный	80

1. **Тип когорты/периода** — определяет величину периода для формирования когорт и расчета периодов жизни когорты:
2. **d** — день;
3. **w** — неделя;
4. **m** — месяц (по умолчанию);
5. **q** — квартал.
6. **Порог, %** — граница порога, % объектов, при котором мы считаем, что событие произошло для большинства.



Выходные порты

Название	Тип
Доля событий по когортам	Таблица
Доля событий	Таблица
Точка стабилизации	Переменные

Структура таблицы "Доля событий по когортам"

Метка	Тип	Описание
Когорта	 Дата/Время	Название когорты (первая дата периода, к которому относится Дата регистрации)
Период	 Целый	Порядковый номер периода
Количество объектов	 Целый	Количество объектов по каждой когорте, для которых событие произошло в текущем периоде
% объектов накопительно	 Вещественный	% объектов по каждой когорте, для которых событие произошло к концу текущего периода

Структура таблицы "Доля событий"

Метка	Тип	Описание
Период	 Целый	Порядковый номер периода
% объектов накопительно	 Вещественный	% объектов по всем когортам, для которых событие произошло к текущему периоду

Переменные в порте "Точка стабилизации"

№	Метка	Тип	Описание
1	Точка стабилизации	12 Целый	Номер периода, в котором был достигнут заданный процент объектов, с которыми событие произошло

Алгоритмы

1. Определение периода для каждой даты и выделение когорт.
2. Расчет количества периодов между датой первого и второго событий.
3. Подсчет количества объектов, для которых событие произошло в определенный номер периода.
4. Подсчет накопительного процента объектов, для которых событие произошло к концу текущего периода.

Кластерные силуэты

[Назад к списку компонентов](#)

Назначение

Позволяет сформировать набор данных для графического построения кластерных силуэтов и проведения общей оценки качества кластеризации, выполненной внешним алгоритмом. В компоненте проводится:

- расчет индекса оценки силуэта для заданного числа точек (но не более 1000);
- расчет индекса силуэта для каждого кластера;
- расчет индекса силуэта кластеризации в целом;
- формирование метки качества кластеризации.

Внешний алгоритм кластеризации должен быть основан на каком-либо центроидном методе (т.е. оперировать понятием центра кластера), например, алгоритм k-средних.

Входные порты

Название	Тип
Таблица объектов	Таблица
Переменные	Переменные

Структура таблицы "Таблица объектов"

Метка	Тип	Описание
Номер кластера	12 Целый	Номер кластера, к которому относится объект
Идентификатор	ab Строковый	Идентификатор объекта
Поле	ab Строковый	Наименование поля, в котором содержится значение характеристики объекта кластера
Значение	9.0 Вещественный	Значение поля

Переменные в порте "Переменные"

№	Метка	Тип	Значение
1	Стандартизация данных	0/1 Логический	true

Стандартизация данных — флаг, при котором значения будут приведены к единому виду. По умолчанию **true**.

Выходные порты

Название	Тип
Кластерные силуэты	Таблица
Индексы силуэтов	Таблица
Качество кластеризации	Переменные

Структура таблицы "Кластерные силуэты"

Метка	Тип	Описание
Номер строки	12 Целый	Порядковый номер строки
Номер кластера	12 Целый	Номер кластера, к которому относится объект
Идентификатор	ab Строковый	Идентификатор объекта
Индекс оценки силуэта	90 Вещественный	Значение индекса оценки силуэта для объекта

Таблица отсортирована по возрастанию номера кластера и убыванию индекса силуэта.

Структура таблицы "Индексы силуэтов"

Метка	Тип	Описание
Номер кластера	12 Целый	Уникальный номер кластера
Индекс силуэта	90 Вещественный	Индекс силуэта кластера

Переменные в порте "Качество кластеризации"

№	Метка	Тип	Описание
1	Индекс силуэта	90 Вещественный	Значение индекса силуэта кластеризации
2	Качество кластеризации	ab Строковый	Интерпретация значения индекса силуэта, см. Алгоритмы

Алгоритмы

Общее описание ([ссылка 1](#)) и подробный алгоритм с реализацией на языке Питон ([ссылка 2](#)).

1. Индекс оценки силуэта **SilhouetteScore** рассчитывается по формуле:

$$SilhouetteScore = \frac{BI - AI}{\max(AI, BI)}, \text{ где:}$$

- AI — Расстояние до центра кластера;
 - BI — Расстояние до центра соседнего кластера.
2. Среднее значение индекса оценки силуэта **SilhouetteCoefficient** варьируется от -1 до 1.
 3. Метка, отвечающая за качество кластеризации, формируется по правилу:
 - Качество кластеризации **Низкое**, если **SilhouetteCoefficient** < 0,2;
 - Качество кластеризации **Среднее**, если **SilhouetteCoefficient** в интервале [0,2; 0,5)
 - Качество кластеризации **Высокое**, если **SilhouetteCoefficient** > 0,5.

Дополнительная литература

1. Паклин, Н.Б., Орешков, В.И. Кластерные силуэты (2016)

Сравнение кластеризаций

[Назад к списку компонентов](#)

Назначение

Предназначен для расчета относительных метрик оценки качества двух кластеризаций (мер близости):

- индекса Рэнда;
- индекса Фолкса-Маллоуса;
- индекса Жаккара;
- разностного разделения.


Входные порты

Название	Тип
Входной набор данных	Таблица
Доля выборки	Переменные

Структура таблицы "Входной набор данных"

Метка	Тип	Описание
Идентификатор	ab Строковый	Идентификатор объекта
Номер кластера 1	12 Целый	Номер кластера объекта по первой кластеризации
Номер кластера 2	12 Целый	Номер кластера объекта по второй кластеризации

Переменные в порте "Доля выборки"

№	Метка	Тип	Значение
1	Доля выборки, %	 Вещественный	100

Для выборки, % — позволяет задать долю объектов, которые будут участвовать в формировании счетных пар. По умолчанию участвуют все объекты (100%).

Выходные порты

Название	Тип
Индексы	Переменные

Переменные в порте "Индексы"

№	Метка	Тип	Описание
1	Индекс Рэнда	 Вещественный	Значение лежит в диапазоне от 0 до 1, 1 говорит об идентичности кластеризаций
2	Индекс Фолкса-Маллоуса	 Вещественный	Обычно используется при иерархической кластеризации. 1 указывает на идентичность кластеризаций
3	Индекс Жаккара	 Вещественный	Рассчитывается при исключении числа наблюдений, попавших в обеих кластеризациях в разные кластеры. Чем выше значение индекса, тем ближе кластеризации
4	Разностное разделение	 Целый	Количество пар объектов, которые принадлежат различным кластерам в кластеризациях

Дополнительная литература

1. [Clustering performance evaluation](#) — портал scikit-learn.org
2. [Fowlkes–Mallows index](#) — Wikipedia
3. [Коэффициент Жаккара](#) — Wikipedia

Сравнение метрик

[Назад к списку компонентов](#)



Назначение

Компонент для сравнения текущих показателей с эталонной шкалой AUC, Gini, KS, Индекс силуэта, а так же с пользовательской шкалой (на выбор) и с построением визуализации.



Входные порты

Название	Тип
Метрика 1	Переменные
Метрика 2	Переменные
Метрика 3	Переменные
Тип метрик	Переменные

Переменные в порте "Метрика 1"

№	Метка	Тип	Значение
1	Метрика 1.Метка	 Строковый	Обучающая выборка
2	Метрика 1.Значение	 Вещественный	0,6

Переменные в порте "Метрика 2"

№	Метка	Тип	Значение
1	Метрика 2.Метка	 Строковый	Тестовая выборка
2	Метрика 2.Значение	 Вещественный	0,7

Переменные в порте "Метрика 3"

№	Метка	Тип	Значение
1	Метрика 3.Метка	ab Строковый	null
2	Метрика 3.Значение	9.0 Вещественный	null

Переменные в порте "Тип метрик"

№	Метка	Тип	Значение
1	Тип метрики	ab Строковый	Custom
2	Неудовлетворительное	9.0 Вещественный	0,30
3	Среднее	9.0 Вещественный	0,50
4	Хорошее	9.0 Вещественный	0,70
5	Очень хорошее	9.0 Вещественный	0,90
6	Отличное	9.0 Вещественный	0,99

Тип метрики — позволяет выбрать тип шкалы для сравнения метрик. Возможные значения:

- **AUC** — индекс AUC;
- **Gini** — индекс Gini;
- **KS** — статистика KS;
- **SI** — индекс силуэта;
- **Custom** — пользовательская шкала (по умолчанию).

Выходные порты

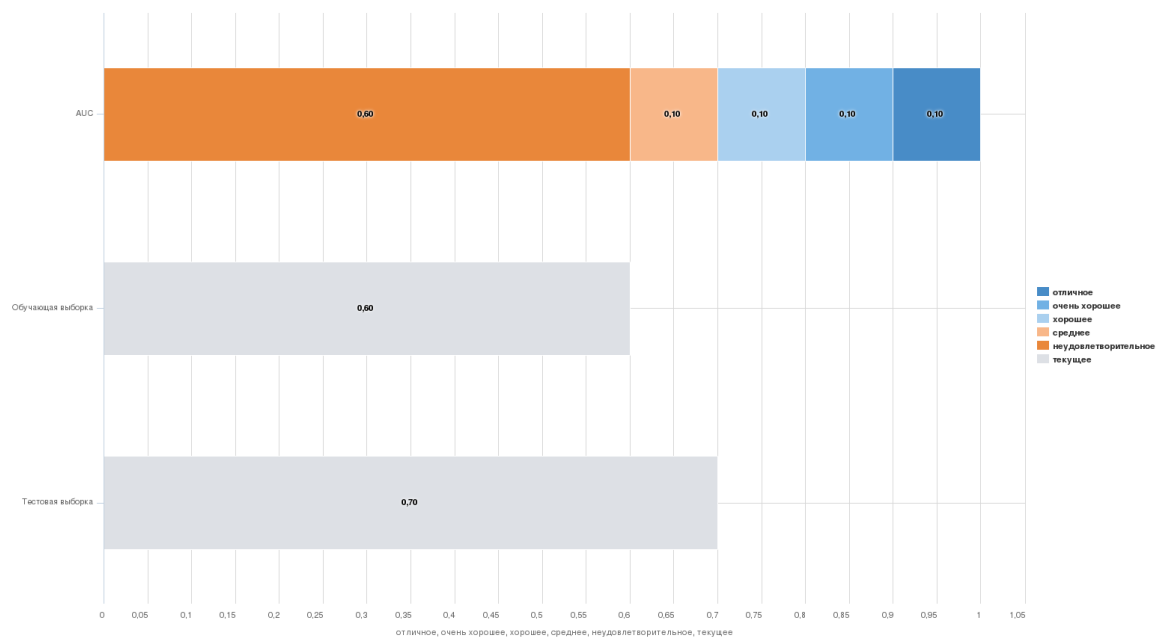
Название	Тип
Сравнение метрик	Таблица

Структура таблицы "Сравнение метрик"

Метка	Тип	Описание
Тип метрики	ab Строковый	Название эталонной шкалы и меток текущих метрик
текущее	90 Вещественный	Значение текущих метрик
неудовлетворительное	90 Вещественный	Неудовлетворительное значение эталонной шкалы
среднее	90 Вещественный	Среднее значение эталонной шкалы
хорошее	90 Вещественный	Хорошее значение эталонной шкалы
очень хорошее	90 Вещественный	Очень хорошее значение эталонной шкалы
отличное	90 Вещественный	Отличное значение эталонной шкалы

Визуализаторы

Диаграмма "Сравнение метрик"



Статистическая значимость

[Назад к списку компонентов](#)

Назначение

Компонент для проверки статистической значимости с помощью одностороннего Т-теста, двустороннего Т-теста, теста Манна-Уитни (U-тест) на выбор.

Входные порты

Название	Тип
Процент по группам	Таблица
Переменные	Переменные

Структура таблицы "Процент по группам"

Метка	Тип	Описание
Контрольная группа, %	9.0 Вещественный	Процент события по контрольной группе (группе, на которую не производилось воздействие)
Тестовая группа, %	9.0 Вещественный	Процент события по тестовой группе (группе, на которую производилось воздействие)

Переменные в порте "Переменные"

№	Метка	Тип	Значение
1	Уровень значимости	9.0 Вещественный	0,05
2	Тип теста	12 Целый	0

1. **Уровень значимости** — это величина, используемая для оценки истинности некоторого результата или гипотезы. Популярными уровнями значимости являются:

- **0,05%** (по умолчанию);
- **0,01%**;
- **0,001%**.

2. **Тип теста** — может принимать значения:

- **0** (по умолчанию) — односторонний Т-тест;
- **1** — двусторонний Т-тест;
- **2** — теста Манна-Уитни (U-тест).

Выходные порты

Название	Тип
Переменные	Переменные

Переменные в порте "Переменные"

№	Метка	Тип	Описание
1	Статистически значимая	ab Строковый	Принимает значения Да или Нет (см. Дополнительная литература)
2	Контрольная группа: кол-во наблюдений	12 Целый	Кол-во наблюдений (записей) по контрольной группе в исходных данных
3	Тестовая группа: кол-во наблюдений	12 Целый	Кол-во наблюдений по тестовой группе в исходных данных
4	Уровень значимости	9.0 Вещественный	Выбранный уровень значимости
5	Значение критерия	9.0 Вещественный	Значение Т-критерия или U-критерия, рассчитанное для групп
6	Тип теста	ab Строковый	Выбранный тест
7	Влияние воздействия, %	9.0 Вещественный	(см. Алгоритмы)
8	Контрольная группа: среднее	9.0 Вещественный	Среднее значение по контрольной группе (Только для Т-теста)
9	Тестовая группа: среднее	9.0 Вещественный	Среднее значение по тестовой группе (Только для Т-теста)
10	Контрольная группа: дисперсия	9.0 Вещественный	Дисперсия по контрольной группе (Только для Т-теста)
11	Тестовая группа: дисперсия	9.0 Вещественный	Дисперсия по тестовой группе (Только для Т-теста)

Алгоритмы

1. **Влияние воздействия, %** — если в результате проверки поле *Статистически значимая* приняло значение **Да**, то расчет влияния воздействия для Т-теста будет происходить по формуле:

$$\frac{(TestAvg - ControlAvg)}{ControlAvg} * 100$$

, где:

- **TestAvg** — среднее значение по тестовой выборке;
- **ControlAvg** — среднее значение по контрольной выборке.

Дополнительная литература

1. t-критерий Стьюдента - материал из [Wikipedia \(рус.\)](#)
2. U-критерий Манна — Уитни - материал из [Wikipedia \(рус.\)](#)

Метрики классификации

[Назад к списку компонентов](#)

Назначение

Позволяет рассчитать ряд метрик для оценки качества классификации.

Входные порты

Название	Тип
Набор данных	Таблица

Структура таблицы "Набор данных"

Метка	Тип	Описание
Событие: истинное	12 Целый	Истинное значение события: 1 — событие произошло; 0 — событие не произошло
Событие: предсказанное	12 Целый	Предсказанное значение события: 1 — модель предсказала, что событие произошло; 0 — модель предсказала, что событие не произошло

Выходные порты

Название	Тип
Метрики	Переменные

Переменные в порте "Метрики"

№	Метка	Тип	Описание
1	Точность	9.0 Вещественный	см. Алгоритмы
2	Полнота	9.0 Вещественный	см. Алгоритмы
3	F1-мера	9.0 Вещественный	см. Алгоритмы
4	Ассурасу	9.0 Вещественный	см. Алгоритмы
5	TN	12 Целый	Количество истинноотрицательных случаев
6	FP	12 Целый	Количество ложноположительных случаев
7	FN	12 Целый	Количество ложноотрицательных случаев
8	TP	12 Целый	Количество истинноположительных случаев

Алгоритмы

Точность = $TP / (TP + FP)$

Полнота = $TP / (TP + FN)$

F1-мера = $2 (Точность Полнота) / (Точность + Полнота)$

Ассурасу = $(TN + TP) / (TN + FP + FN + TP)$

Дополнительная литература

1. **F-score**

Тривиальная модель

[Назад к списку компонентов](#)




Назначение

Построение тривиальной (base) модели, оценивающей вероятность наступления события на одном входном факторе (непрерывного вида, целого или вещественного типа). Используется логистическая регрессия с предварительной процедурой оптимального квантования. Модели автоматически переобучаются.

Входные порты

Название	Тип
Обучающая выборка	Таблица
Тестовая выборка	Таблица

Структура таблиц "Обучающая выборка", "Тестовая выборка"

Метка	Тип	Описание
Идентификатор	 Строковый	Идентификатор объекта
Тривиальный показатель	 Вещественный	Поле с тривиальным показателем
Событие	 Логический	Поле с событием

1. Поле с тривиальным показателем должно быть непрерывного вида, вещественного или целого типа.
2. Тривиальное поле для тестовой и обучающей выборок должно совпадать, но этот контроль лежит на пользователе компонента.
3. Событие содержит два уникальных значения логического типа. `True`

соответствует наступлению события.

Выходные порты

Название	Тип
Вероятности событий	Таблица
Метрики качества	Переменные

Структура таблицы "Вероятности событий"

Метка	Тип	Описание
Идентификатор	ab Строковый	Идентификатор объекта
Вероятность события	90 Вещественный	Поле с тривиальным показателем

Объекты отсортированы в порядке убывания вероятности события.

Переменные в порте "Метрики качества"

№ | Метка | Тип | Описание | |:-|:-----|:-----|:---
-----| | 1 | Индекс AUC | 90 Вещественный |
Расчитанное значение индекса AUC для модели | | 2 | Качество | ab Строковый |
Словесное описание качества модели (см. **Алгоритмы**) | | 3 | Стандартная ошибка
| 90 Вещественный | Стандартное отклонение по набору | | 4 | Доверительный
интервал нижний | 90 Вещественный | Нижняя граница доверительного
интервала | | 5 | Доверительный интервал верхний | 90 Вещественный | Верхняя
граница доверительного интервала | | 6 | Z-оценка | 90 Вещественный |
Расчитанное значение Z-оценки |

Переменные выходного порта соответствуют компоненту **AUC** и рассчитываются для тестовой выборки. Диаграммы ROC-кривых для тестовой и обучающей выборок можно увидеть, зайдя внутрь компонента.

IV-отбор

[Назад к списку компонентов](#)

Назначение

Компонент предназначен для постобработки результатов компонента **Конечные классы**:

- отбор столбцов, которые содержат метку конечного класса (по умолчанию имя такого столбца имеет постфикс `_ClassMark`);
- отбор значимых столбцов (список значимых столбцов подается отдельно и может быть взят с выхода конечных классов).

Входные порты

Название	Тип
Входной набор данных	Таблица
Значимости столбцов	Таблица
Переменные	Переменные

Структура таблицы "Входной набор данных"

На вход нужно подавать выходной набор данных из первого выходного порта компонента **Конечные классы**.

Структура таблицы "Значимости столбцов"

Метка	Тип	Описание
Имя столбца	ab Строковый	
Метка столбца	ab Строковый	
Значимость столбца	ab Строковый	

Данный набор данных может быть взят из третьего выходного порта компонента **Конечные классы**, предварительно отфильтровав нужные по значимости столбцы.

Переменные в порте "Переменные"

Метка	Тип	Значение
Исключаемые столбцы	ab Строковый	Id,Event
Постфикс метки	ab Строковый	_ClassMark

Переменную **Постфикс метки** менять нежелательно, если используются результаты работы компонента **Конечные классы**. В переменную **исключаемые столбцы** задаются через разделитель "," имена полей, которые необходимо добавить "как есть" на выходе компонента (обычно это поле с идентификатором и событием).

Выходные порты

Название	Тип
Выходной набор данных	Таблица

Структура таблицы "Выходной набор данных"

Структура таблицы соответствует входному набору после преобразования.

Генератор календаря

[Назад к списку компонентов](#)

Назначение




Компонент позволяет сгенерировать временной ряд между заданными датами. Промежуток между датами задается переменной **Тип периода**. Возможные значения:

- **d** — день;
- **w** — неделя (в соответствии с международным стандартом **ISO 8601**);
- **m** — месяц;
- **q** — квартал;
- **y** — год.

Входные порты

Название	Тип
Даты	Переменные

Переменные в порте "Даты"

№	Метка	Тип	Значение
1	Период начальный	 Дата/Время	01.01.2023, 0:00
2	Период конечный	 Дата/Время	31.12.2023, 0:00
3	Тип периода	 Строковый	d

Выходные порты

Название	Тип
Набор данных	Таблица

Структура таблицы "Набор данных"

Метка	Тип	Описание
Календарь	 Дата/Время	Сформированный временной ряд

Генератор пустых полей

[Назад к списку компонентов](#)

Назначение

Как правило, требуется, когда на входе какого-либо компонента есть обязательное поле, которого нет в анализируемом наборе данных.

Предназначен для создания четырех пустых полей разных типов:

- Строковый;
- Вещественный;
- Целый;
- Дата/Время.

Созданные поля можно присоединить к набору данных узлом **Соединение**.

Входные порты

У компонента нет входных портов.

Выходные порты

Название	Тип
Набор данных	Таблица

Структура таблицы "Набор данных"

Метка	Тип	Описание
Объект	ab Строковый	Пустое поле строкового типа
Дата	31 Дата/Время	Пустое поле типа Дата/Время
Вещественное значение	90 Вещественный	Пустое поле вещественного типа
Целое значение	12 Целый	Пустое поле целого типа

Генератор списка

[Назад к списку компонентов](#)

Назначение

Компонент позволяет сгенерировать нумерованный список. Длина списка задается переменными **Начало списка** и **Конец списка**.

Входные порты

Название	Тип
Переменные	Переменные

Переменные в порте "Переменные"

№	Метка	Тип	Значение
1	Начало списка	12 Целый	1
2	Конец списка	12 Целый	100

Если значение переменной **Начало списка** > **Конец списка**, то выводится пустой набор данных.

Если значение переменной **Начало списка** = **Конец списка**, то выводится одна строка.

Выходные порты

Название	Тип
Набор данных	Таблица

Структура таблицы "Набор данных"

Метка	Тип	Описание
Список	12 Целый	Сформированный список

Категория к числу

[Назад к списку компонентов](#)

Назначение

Компонент осуществляет кодирование каждой категории в целое число. Расстояние между соседними целыми числами зависит от переменной **Инкремент**.

Входные порты

Название	Тип
Категории	Таблица
Переменные	Переменные

Структура таблицы "Номинальные данные"

Метка	Тип	Описание
Категория	ab Строковый	Наименование или идентификатор категории

Переменные в порте "Переменные"

№	Метка	Тип	Значение
1	Начальное значение	12 Целый	0
2	Инкремент	12 Целый	1
3	Считать пустые	0/1 Логический	true

1. **Начальное значение** — значение, которое будет присвоено первой категории.

2. **Инкремент** — категории присваивается номер по принципу: *Начальное значение + Инкремент * Позиция категории*.
3. **Считать пустые** — задает способ обработки пустых значений. При значении **true** (по умолчанию) пустой категории будет присвоен номер, **false** — пустые категории останутся пустыми.

Выходные порты

Название	Тип
Набор данных	Таблица
Таблица замен	Таблица

Структура таблицы "Набор данных"

Метка	Тип	Описание
Категория (замена)	12 Целый	Числовые значения категорий

Структура таблицы "Таблица замен"

Метка	Тип	Описание
Категория (Category)	ab Строковый	Список уникальных значений категорий
Категория (замена)	12 Целый	Список уникальных номеров, присвоенных категориям

Разность дат

[Назад к списку компонентов](#)

Назначение

Компонент позволяет рассчитать разность между двумя датами. Может принимать на вход как набор данных из двух колонок, так и две переменные.

Входные порты



Название	Тип
Даты	Таблица
Даты	Переменные

Структура таблицы "Даты"

Метка	Тип	Описание
Дата 1	 Дата/Время	Временной ряд 1
Дата 2	 Дата/Время	Временной ряд 2

Данный входной порт является необязательным.













Переменные в порте "Даты"

№	Метка	Тип	Значение
1	Дата 1	 Дата/Время	01.01.2023 12:30:48
2	Дата 2	 Дата/Время	03.01.2023 10:25:12

Выходные порты

Название	Тип
Набор данных	Таблица
Переменные	Переменные

Структура таблицы "Набор данных"

Метка	Тип	Описание
Разность	 Вещественный	Математическая разность между датами
Дней (Всего)	 Целый	Разница в полных днях
Часов (Всего)	 Целый	Разница в полных часах
Минут (Всего)	 Целый	Разница в полных минутах
Секунд (Всего)	 Целый	Разница в секундах
Часов (Остаток)	 Целый	Количество часов с учетом полных дней
Минут (Остаток)	 Целый	Количество минут с учетом часов
Секунд (Остаток)	 Целый	Количество секунд с учетом минут
Разница с учетом дней	 Строковый	Прошедшее время между датами с учетом полных дней
Разница в часах	 Строковый	Прошедшее время между датами в часах
Дата 1	 Дата/Время	Исходный временной ряд 1
Дата 2	 Дата/Время	Исходный временной ряд 2

Переменные в порте "Переменные"

№	Метка	Тип	Значение
1	Разность	 Вещественный	1,91
2	Дней (Всего)	 Целый	1
3	Часов (Всего)	 Целый	45
4	Минут (Всего)	 Целый	2754
5	Секунд (Всего)	 Целый	165264
6	Часов (Остаток)	 Целый	21
7	Минут (Остаток)	 Целый	54
8	Секунд (Остаток)	 Целый	24
9	Разница с учетом дней	 Строковый	21:54:24
10	Разница в часах	 Строковый	45:54:24
11	Дата 1	 Дата/Время	01.01.2023 12:30:48
12	Дата 2	 Дата/Время	03.01.2023 10:25:12

Сдвиг дат

[Назад к списку компонентов](#)


Назначение

Компонент для сдвига даты на одну или несколько единиц измерения времени.

Входные порты

Название	Тип
Даты	Таблица
Переменные	Переменные

Структура таблицы "Даты"

Метка	Тип	Описание
Список дат	 Дата/Время	Временной ряд

Переменные в порте "Переменные"

№	Метка	Тип	Значение
1	Направление сдвига	ab Строковый	+
2	Годы	12 Целый	0
3	Месяцы	12 Целый	0
4	Дни	12 Целый	0
5	Часы	12 Целый	0
6	Минуты	12 Целый	0
7	Секунды	12 Целый	0

Направление сдвига — переменная, принимающая два значения:

- **+** (плюс) — прибавляет указанные единицы измерения времени к исходной дате (по умолчанию);
- **-** (минус) — вычитает указанные единицы измерения времени к исходной дате;

Выходные порты

Название	Тип
Выходной набор данных	Таблица

Структура таблицы "Выходной набор данных"

Метка	Тип	Описание
Список дат	31 Дата/Время	Преобразованный временной ряд

Перцентиль N%

[Назад к списку компонентов](#)


Назначение

Компонент рассчитывает значение N-го перцентиля в ряде данных. Перцентиль выражается в процентах и указывает, какое значение находится ниже указанного процента данных. Например, если мы говорим о 75-перцентиле, то это означает, что 75% значений находятся ниже этого значения, а оставшиеся 25% значений выше.



Входные порты

Название	Тип
Ряд данных	Таблица
Переменные	Переменные

Структура таблицы "Ряд данных"

Метка	Тип	Описание
Значения	 Вещественный	Исходный ряд данных

Переменные в порте "Переменные"

№	Метка	Тип	Значение
1	Значение N	 Целый	95
2	Сравнение	 Строковый	>

В переменной **Значение N** задается требуемое значение перцентиля в процентах. Переменная **Сравнение** может принимать значения **>** или **<** . Она влияет на логику формирования выходного столбца **Флаг**.

Выходные порты

Название	Тип
Ряд данных	Таблица
Перцентиль	Переменные
Перцентиль	Таблица

Структура таблицы "Ряд данных"

Метка	Тип	Описание
Значения	90 Вещественный	Исходный ряд данных
Флаг	0/1 Логический	См. ниже

Значение в столбце **Флаг** принимает **True** , если значение ряда больше или меньше (в зависимости от переменной **Сравнение**) перцентиля.

Переменные в порте "Перцентиль"

№	Метка	Тип	Значение
1	Перцентиль	90 Вещественный	Рассчитанный перцентиль

Структура таблицы "Перцентиль"

Метка	Тип	Описание	
Перцентиль	90 Вещественный	Рассчитанный перцентиль	

Случайная категория JS

[Назад к списку компонентов](#)

Назначение

Присваивает каждому объекту из списка n-случайных категорий с учетом вероятности выбора категории (или без учета вероятности). Для одного объекта категории не повторяются. Компонент использует код на языке JS (компонент **JavaScript**).

Входные порты

Название	Тип
Объекты	Таблица
Категории	Таблица
Значения	Таблица
Переменные	Переменные

Структура таблицы "Объекты"

Метка	Тип	Описание
Объект	ab Строковый	Идентификатор или наименование объекта (например, клиента)

Структура таблицы "Категории"

Метка	Тип	Описание
Категория	ab Строковый	Идентификатор или наименование категории (например, товара)

Структура таблицы "Значения"

Метка	Тип	Описание
Значение	90 Вещественный	Значение, относящее к категории (например, количество покупок)

Данный порт является необязательным.

Переменные в порте "Переменные"

№	Метка	Тип	Значение
1	Количество категорий	12 Целый	3

Выходные порты

Название	Тип
Набор данных	Таблица

Структура таблицы "Набор данных"

Метка	Тип	Описание
Объект	ab Строковый	Идентификатор или наименование объекта
Категория	ab Строковый	Категория, присвоенная объекту

Алгоритмы

1. Расчет вероятности выбора категории: Показатель по категории / Общая сумма показателей по всем категориям.
2. Формирование границ по категориям. Верхняя граница — накопительное значение вероятности.
3. Генерация случайного числа по каждому объекту и проверка, в границы

какой категории оно попало.

4. Проверка, не была ли эта категория уже присвоена объекту ранее. Если была, повторяем пункты 3,4.

Фильтр столбцов JS

[Назад к списку компонентов](#)

Назначение

Предназначен для отбора столбцов входного набора данных по списку. Работает в двух режимах:

- оставляет в выходном наборе столбцы, имена которых есть в списке;
- исключает из выходного набора столбцы, имена которых есть в списке.

Компонент использует код на языке JS (компонент **JavaScript**).

Входные порты

Название	Тип
Набор данных	Таблица
Список столбцов	Таблица
Переменные	Переменные

Структура таблицы "Набор данных"

Структура таблицы не определена, в порте включена автосинхронизация.

Структура таблицы "Список столбцов"

Метка	Тип	Описание
Имена столбцов	ab Строковый	Список имен столбцов входного набора, которые необходимо оставить/исключить из выходного

Данный порт является необязательным.

Переменные в порте "Переменные"

№	Метка	Тип	Значение
1	Исключать поля по списку	0/1 Логический	true
2	Список полей	ab Строковый	null

Исключать поля по списку:

- **true** — из выходного набора будут исключены столбцы, список которых подан на вход **Список столбцов** или указан в переменной **Список полей**, режим по умолчанию;
- **false** — в выходной набор попадут только столбцы, список которых подан на вход **Список столбцов** или указан в переменной **Список полей**.

Список полей — имена столбцов, которые необходимо оставить/исключить из выходного набора, перечисленные через запятую. Данный список является приоритетным по отношению к таблице **Список столбцов** в случае, когда он заполнен.

Выходные порты

Название	Тип
Набор данных	Таблица

Структура таблицы "Набор данных"

Структура таблицы будет соответствовать списку столбцов на входном порте **Список столбцов**.

IF-объединение JS

[Назад к списку компонентов](#)

Назначение

Компонент принимает на вход два набора данных, один из которых обязательно является пустым. На выходе остается только тот набор данных, который не является пустым. Используется как один из вариантов объединения ветвей после выполнения узла **Условие**.

Входные порты

Название	Тип
Входной источник данных 1	Таблица
Входной источник данных 2	Таблица

Структура таблицы "Входной источник данных 1"

Структура таблицы не определена, в порте включена автосинхронизация.

Структура таблицы "Входной источник данных 2"

Структура таблицы не определена, в порте включена автосинхронизация.

Выходные порты

Название	Тип
Выходной набор данных	Таблица

Структура таблицы "Набор данных"

Структура таблицы соответствует структуре набора данных с порта **Входной источник данных 1** или **Входной источник данных 2**.

Список полей JS

[Назад к списку компонентов](#)

Назначение

Компонент позволяет получить информацию по входным полям: имя, метку, тип и вид поля.

Входные порты

Название	Тип
Входной источник данных	Таблица

Структура таблицы "Входной источник данных"

Структура таблицы не определена, в порте включена автосинхронизация.

Выходные порты

Название	Тип
Выходной источник данных	Таблица

Структура таблицы "Выходной источник данных"

Метка	Тип	Описание
Поле.Имя	ab Строковый	Выводит имя поля
Поле.Метка	ab Строковый	Выводит метку поля
Поле.Тип	ab Строковый	Выводит тип поля: Integer, Boolean, DateTime, Float, String, Continuous
Поле.Вид	ab Строковый	Выводит вид поля: Undefined, Continuous, Discrete

JSON-парсер JS

[Назад к списку компонентов](#)

Назначение

Простой компонент для разбора JSON строк и извлечения данных из них.

Входные порты

Название	Тип
Входной источник данных	Таблица
Входные переменные	Переменные

Структура таблицы "Входной источник данных"

Метка	Тип	Описание
Строка JSON	ab Строковый	Строка формата JSON

Переменные в порте "Входные переменные"

№	Метка	Тип	Значение
1	Строка JSON	ab Строковый	null

Выходные порты

Название	Тип
Выходной набор данных	Таблица

Структура таблицы "Выходной набор данных"

Метка	Тип	Описание
Индекс ответа	 Строковый	Идентификатор строки входного набора
parent	 Строковый	parent элемента JSON
key	 Строковый	идентификатор элемента JSON
type	 Строковый	тип элемента JSON
value	 Переменный	значение элемента JSON

Объединение столбцов JS

[Назад к списку компонентов](#)

Назначение

Компонент для объединения столбцов в одну строку через произвольный разделитель.

Входные порты

Название	Тип
Набор данных	Таблица
Входные переменные	Переменные

Структура таблицы "Набор данных"

Структура таблицы не определена, в порте включена автосинхронизация.

Переменные в порте "Входные переменные"

№	Метка	Тип	Значение
1	Разделитель	ab Строковый	,

Выходные порты

Название	Тип
Выходной набор данных	Таблица

Структура таблицы "Выходной набор данных"

Метка	Тип	Описание
Выходная строка	ab Строковый	Строка после объединения