

Библиотека «Loginom Scorecard Quality Kit»

Версия 1.0.0

Библиотека **Loginom Scorecard Quality Kit** позволяет построить KPI панель, включающую все необходимые метрики для оценки качества скоринговой карты. Визуализаторы строятся в формате отчетов в Loginom.

Компоненты находятся в пакете **loginom_scorecard_quality_kit.lgp**.

Комплект поставки

Библиотека **Loginom Scoring Engine Kit** состоит из следующих каталогов:

Файл/Каталог	Назначение
loginom_scorecard_quality_kit.lgp	Пакет с компонентами библиотеки
Кейсы использования.lgp	Пакет с примерами использования компонентов библиотеки
data	Содержит демонстрационный набор данных в формате .lgd

Установка и настройка

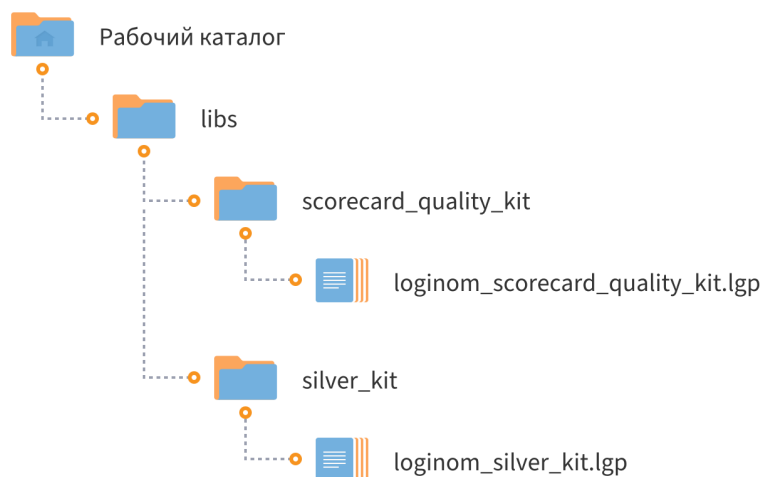
Требования

Для работы библиотеки **Loginom Scorecard Quality Kit** необходимо:

- Клиент доступа к Loginom Studio или любая локальная редакция Loginom. Версия не ниже 7.0.2
- Библиотека Loginom Silver Kit ([Скачать на GitHub](#)). Версия не ниже 3.0.0.

Порядок установки

1. Перейдите в файловое хранилище клиента Loginom Studio или назначьте каталог на локальном диске в случае Loginom CE.
2. Создайте каталог **libs**.
3. Распакуйте архив **scorecard_quality_kit.zip** и поместите все файлы из архива в каталог **libs** так, как они лежали в архиве.
4. Убедитесь, что библиотека Loginom Silver Kit находится в каталоге **libs** в папке **silver_kit**.



Требования к данным






Для построения KPI панели требуется следующий набор данных:

- **Рейтинг** — подаётся на компонент **Scorecard KPI 1** и **Scorecard KPI 2**.

Рейтинг

Набор данных, где для каждой скоринговой карты содержится информация о набранном балле и наличии отклика по каждому запросу. Включает тестовую и обучающую выборки.

Структура таблицы

Метка	Тип	Описание
Идентификатор	 Строковый	Идентификатор запроса
Балл	 Вещественный	Балл, назначенный каждому идентификатору
Скоринговая карта	 Строковый	Идентификатор скоринговой карты
Тестовая выборка	 Логический	true — тестовая, false — обучающая
Отклик	 Логический	true — событие, false — не-событие

Компоненты

- Scorecard KPI 1
- Scorecard KPI 2

Компоненты поставляются в зашифрованном виде и могут быть использованы только через компонент **Выполнение узла**.

Производные компоненты

- Scorecard KPI 1. Отчеты
- Scorecard KPI 2. Отчеты

Scorecard KPI 1

Назначение

Компонент формирует наборы данных, которые позволяют построить ряд отчетов для оценки качества скоринговой карты: распределение баллов, шансов, CAP-, KS-, ROC-кривые и прочее.

Входные порты

Название	Тип
Рейтинг	Таблица
Переменные	Переменные

Структура таблицы "Рейтинг"

Структура таблицы описана в разделе [Требования к данным](#).

Переменные в порте "Переменные"

№	Метка	Тип	Значение
1	Скоринговая карта	ab Строковый	1
2	Тестовая выборка	0/1 Логический	true
3	Количество интервалов	12 Целый	15
4	Количество квантилей (Распределение)	12 Целый	50
5	Количество квантилей (Шансы)	12 Целый	20

1. **Скоринговая карта** — идентификатор карты, по которой будут рассчитаны KPI.

2. **Тестовая выборка** — флаг, указывающий по какому типу выборки будут проводиться расчеты:
 - **true** — тестовая выборка (по умолчанию);
 - **false** — обучающая выборка.
3. **Количество интервалов** — позволяет задать количество интервалов для **распределения скоринговых оценок**.
4. **Количество квантилей (Распределение)** — позволяет задать количество квантилей для **распределения счетов с событием и не-событием**.
5. **Количество квантилей (Шансы)** — позволяет задать количество квантилей для **шансов событий/не-событий**.

Выходные порты

Название	Тип
Распределение скоринговых оценок	Таблица
Распределение счетов с событием и не-событием	Таблица
Шансы событий/не-событий (рисковые сегменты)	Таблица
CAP-кривая	Таблица
KS-кривая, ROC-кривая, Кумулятивное распределение счетов	Таблица
Переменные	Переменные

Структура таблицы "Распределение скоринговых оценок"

Метка	Тип	Описание
Номер интервала	12 Целый	
Интервал балла	ab Строковый	Метка интервала, которая содержит верхнюю и нижнюю границы
Идентификатор (Количество)	12 Строковый	Количество запросов, попавших в интервал

Структура таблицы "Распределение счетов с событием и не-событием"

Метка	Тип	Описание
Балл	9.0 Вещественный	
Событие	9.0 Вещественный	Количество событий для соответствующего балла
Не-событие	9.0 Вещественный	Количество не-событий для соответствующего балла

Структура таблицы "Шансы событий/не-событий (рисковые сегменты)"

Метка	Тип	Описание
Балл	9.0 Вещественный	
Шансы События	9.0 Вещественный	Нормированный шанс события для соответствующего балла
Шансы Не-события	9.0 Вещественный	Нормированный шанс не-события для соответствующего балла

Структура таблицы "CAP-кривая"

Метка	Тип	Описание
Балл	9.0 Вещественный	Максимальный балл из попавших в интервал
Объем выборки	12 Целый	Номер интервала от 1 до 100
Случайное угадывание	12 Целый	Диагональная линия, соответствующая «бесполезному» классификатору, который выполняет предсказания случайным образом
Событие кумулятив., %	9.0 Вещественный	CAP-кривая для оценки модели
Идеальная линия	9.0 Вещественный	Идеальная CAP-кривая с максимальным числом положительных результатов

Структура таблицы "KS-кривая, ROC-кривая, Кумулятивное распределение счетов"

Метка	Тип	Описание
Балл	9.0 Вещественный	Пороговое значение
KS	9.0 Вещественный	$\left(\frac{TP}{(TP + FN) \cdot 100} - \frac{FP}{(TN + FP) \cdot 100} \right) \cdot 100$
Специфичность, %	9.0 Вещественный	$\frac{TN}{(TN + FP) \cdot 100}$
Чувствительность, %	9.0 Вещественный	$\frac{TP}{(TP + FN) \cdot 100}$
% ложноположительных	9.0 Вещественный	$\frac{FP}{(TN + FP) \cdot 100}$
% ложноотрицательных	9.0 Вещественный	$\frac{FN}{(TP + FN) \cdot 100}$
Кумулятивная доля событий, %	9.0 Вещественный	Процент отношения кумулятивной суммы события к общему количеству событий
Кумулятивная доля не событий, %	9.0 Вещественный	Процент отношения кумулятивной суммы не-события к общему количеству не-событий
Диагональ	9.0 Вещественный	Диагональная линия, соответствующая «бесполезному» классификатору, который выполняет предсказания случайным образом

- **TP** — количество истинноположительных случаев;
- **FP** — количество ложноположительных случаев;
- **TN** — количество истинноотрицательных случаев;
- **FN** — количество ложноотрицательных случаев;

Переменные в порте "Переменные"

№	Метка	Тип	Описание
1	Качество	ab Строковый	Словесное описание качества модели
2	Гипотеза о равенстве двух функций	ab Строковый	(см. Алгоритмы)
3	Дивергенция	9.0 Вещественный	(см. Алгоритмы)
4	KS, %	9.0 Вещественный	Значение статистики Колмогорова-Смирнова
5	Индекс AUC	9.0 Вещественный	Расчитанное значение индекса AUC для модели
6	Стандартная ошибка	9.0 Вещественный	Стандартное отклонение по набору
7	AUC доверительный интервал нижний	9.0 Вещественный	Нижняя граница доверительного интервала
8	AUC доверительный интервал верхний	9.0 Вещественный	Верхняя граница доверительного интервала
9	Z-оценка	9.0 Вещественный	Расчитанное значение Z-оценки

Алгоритмы

1. Гипотеза о равенстве двух функций.

Если **KS** меньше **порогового значения**, то гипотезу о равенстве двух функций распределений отвергают. Иначе гипотеза подтверждается.

$z(\alpha)\sqrt{\frac{m+n}{mn}}$ — пороговое значение, где:

- $z(\alpha) = 0,05$ — выбранный уровень значимости;

- m — количество Событий;
- n — количество Не-событий.
- **Дивергенция.**

Чем больше значение коэффициента дивергенции, тем лучше качество модели с точки зрения ее классификационной способности.

$$D = \frac{2(\bar{x}_1 - \bar{x}_2)^2}{(s_1^2 + s_2^2)} \text{ — коэффициент дивергенции, где:}$$

- \bar{x}_1 — среднее значение скорингового балла для Не-событий;
- \bar{x}_2 — среднее значение скорингового балла для Событий;
- s_1^2 — дисперсия скорингового балла для Не-событий;
- s_2^2 — дисперсия скорингового балла для Событий.

Дополнительная литература

1. Сорокин, А. К вопросу о валидации модели логистической регрессии в кредитном скоринге — интернет-журнал «Науковедение», выпуск 2, март – апрель 2014.

Scorecard KPI 1. Отчеты

Назначение

Компонент предназначен для формирования отчётов по результатам работы одноименного компонента **Scorecard KPI 1**:

- Распределение скоринговых оценок;
- Распределение счетов с событием и не-событием;
- Шансы событий/не-событий (рисковые сегменты);
- CAP-кривая;
- KS-кривая;
- ROC-кривая;
- Кумулятивное распределение счетов.

Входные порты

Название	Тип
Распределение скоринговых оценок	Таблица
Распределение счетов с событием и не-событием	Таблица
Шансы событий/не-событий (рисковые сегменты)	Таблица
CAP-кривая	Таблица
KS-кривая, ROC-кривая, Кумулятивное распределение счетов	Таблица

Структура таблицы "Распределение скоринговых оценок"

Структура таблицы описана в разделе **Scorecard KPI 1**.

Структура таблицы "Распределение счетов с событием и не-событием"

Структура таблицы описана в разделе [Scorecard KPI 1](#).

Структура таблицы "Шансы событий/не-событий (рисковые сегменты)"

Структура таблицы описана в разделе [Scorecard KPI 1](#).

Структура таблицы "CAP-кривая"

Структура таблицы описана в разделе [Scorecard KPI 1](#).

Структура таблицы "KS-кривая, ROC-кривая, Кумулятивное распределение счетов"

Структура таблицы описана в разделе [Scorecard KPI 1](#).

Scorecard KPI 2

Назначение

Компонент предназначен для сравнения метрик AUC, KS, Gini обучающей и тестовой выборки с обычными моделями.

Входные порты

Название	Тип
Рейтинг	Таблица
Переменные	Переменные

Структура таблицы "Рейтинг"

Структура таблицы описана в разделе [Требования к данным](#).

Переменные в порте "Переменные"

№	Метка	Тип	Значение
1	Скоринговая карта	 Строковый	1

Выходные порты

Название	Тип
Статистика AUC, KS %, Gini %	Таблица

Структура таблицы "Статистика AUC, KS %, Gini %"

Метка	Тип	Описание
Тип метрики	ab Строковый	AUC, индекс Gini, статистика KS
Тип модели	ab Строковый	Обучающая, тестовая, обычная
текущее	90 Вещественный	Значение метрики для обучающей и тестовой модели
неудовлетворительное	90 Вещественный	Неудовлетворительное значение метрики для обычной модели
среднее	90 Вещественный	Удовлетворительное значение метрики для обычной модели
хорошее	90 Вещественный	Хорошее значение метрики для обычной модели
очень хорошее	90 Вещественный	Очень хорошее значение метрики для обычной модели
отличное	90 Вещественный	Отличное значение метрики для обычной модели

Значения для обычной модели

Значения метрик для обычной модели выводятся в виде накопительной шкалы, с которой будут сравниваться значения метрики для обучающей и тестовой модели.

Тип метрики	Тип модели	неудовлетворительное	среднее	хорошее
AUC	*Обычная	0,60	0,10	0,10
Индекс Gini	*Обычная	35,00	20,00	15,00
Статистика KS	*Обычная	20,00	20,00	10,00

Scorecard KPI 2. Отчеты

Назначение

Компонент предназначен для формирования отчётов по результатам работы одноименного компонента **Scorecard KPI 2**:

- AUC;
- Статистика KS;
- Индекс Gini.

Входные порты

Название	Тип
Статистика AUC, KS %, Gini %	Таблица

Структура таблицы "Статистика AUC, KS %, Gini %"

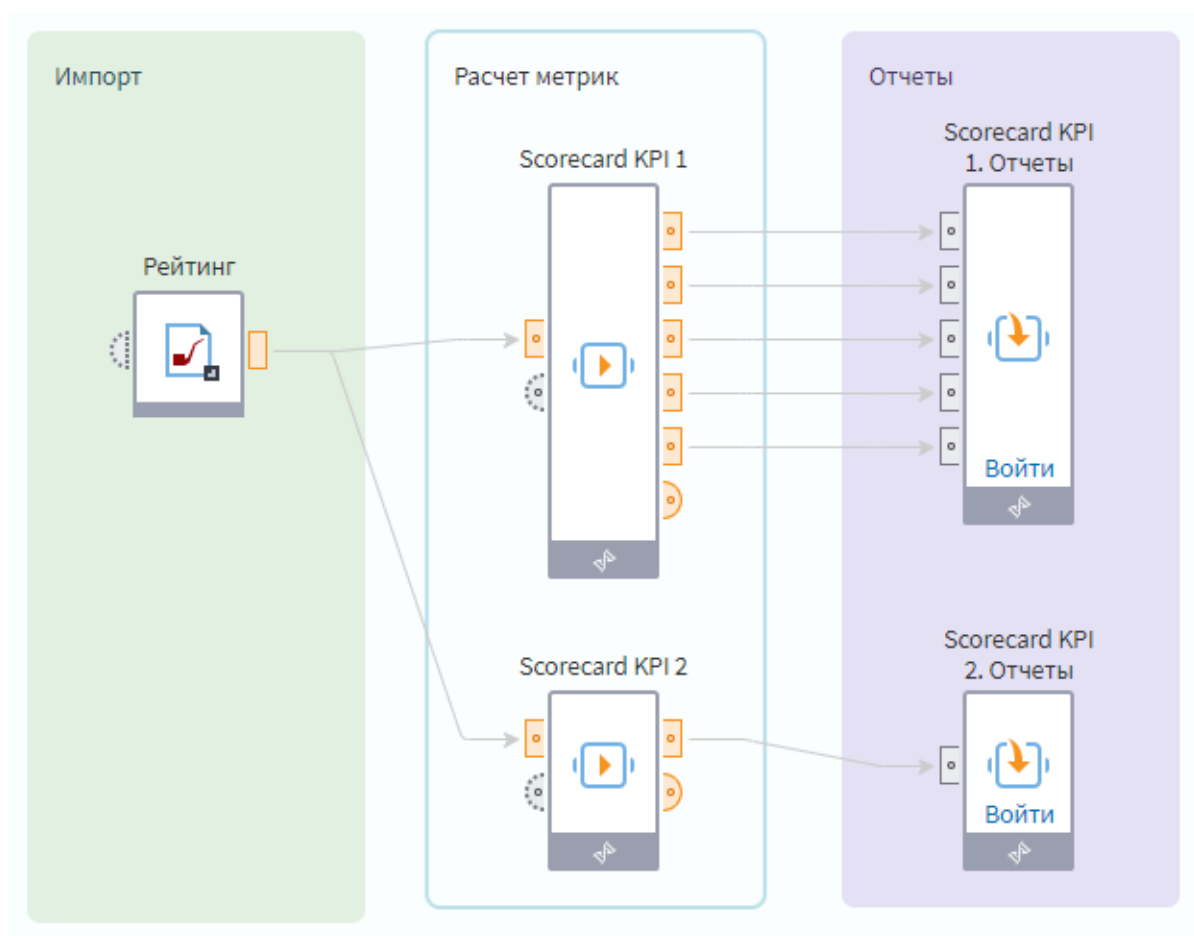
Структура таблицы описана в разделе **Scorecard KPI 2**.

Кейсы использования

Кейс 1. Отчеты

Отчеты формируются в следующей последовательности:

1. Scorecard KPI 1 — Scorecard KPI 1. Отчеты;
2. Scorecard KPI 2 — Scorecard KPI 2. Отчеты.



В данном кейсе отчеты распределены по следующим группам:

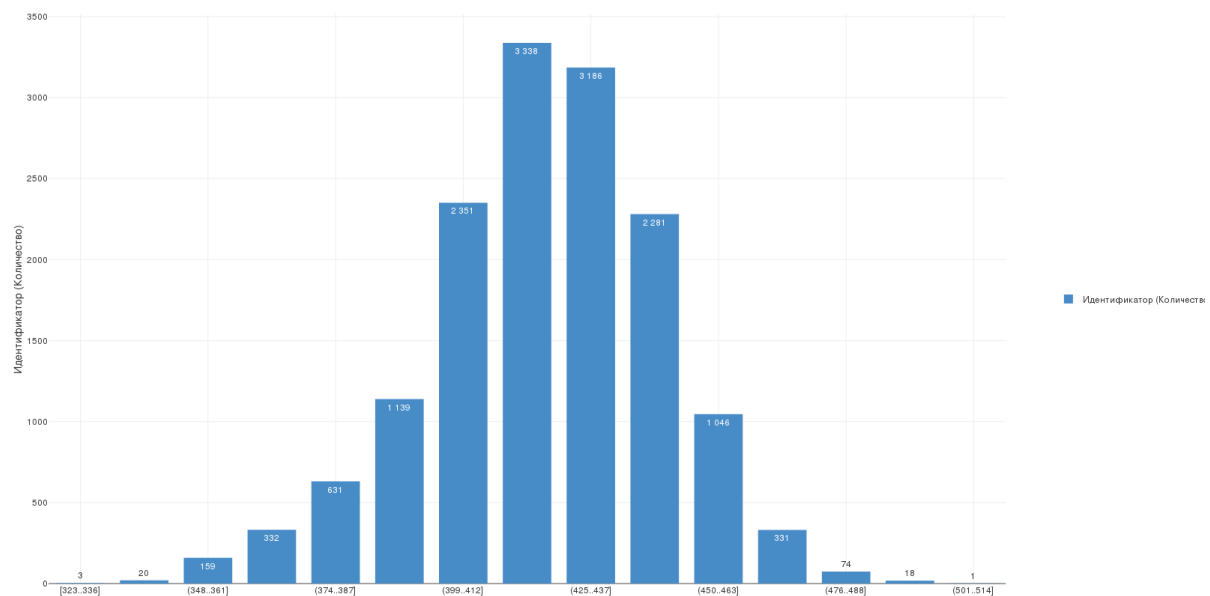
1. Главные метрики:
 - Распределение скоринговых оценок;
 - CAP-кривая;
 - Индекс Gini;
 - Статистика KS.
2. Вспомогательные метрики:

- Распределение счетов с событием и не-событием;
- Кумулятивное распределение счетов;
- KS-кривая;
- Шансы событий/не-событий (рисковые сегменты).

3. ROC, AUC:

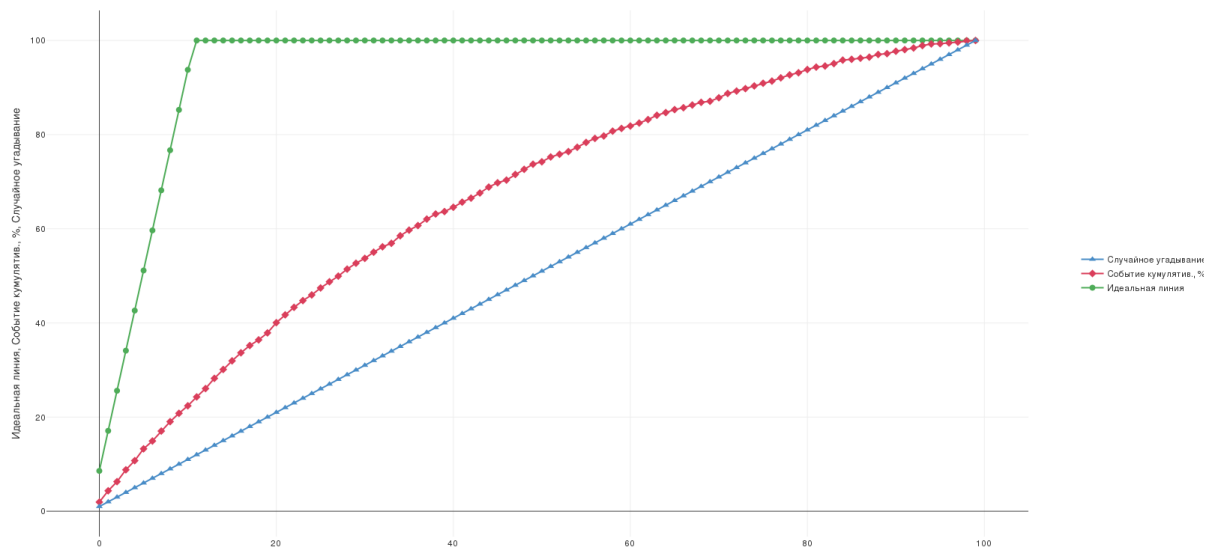
- ROC-кривая;
- AUC.

Распределение скоринговых оценок



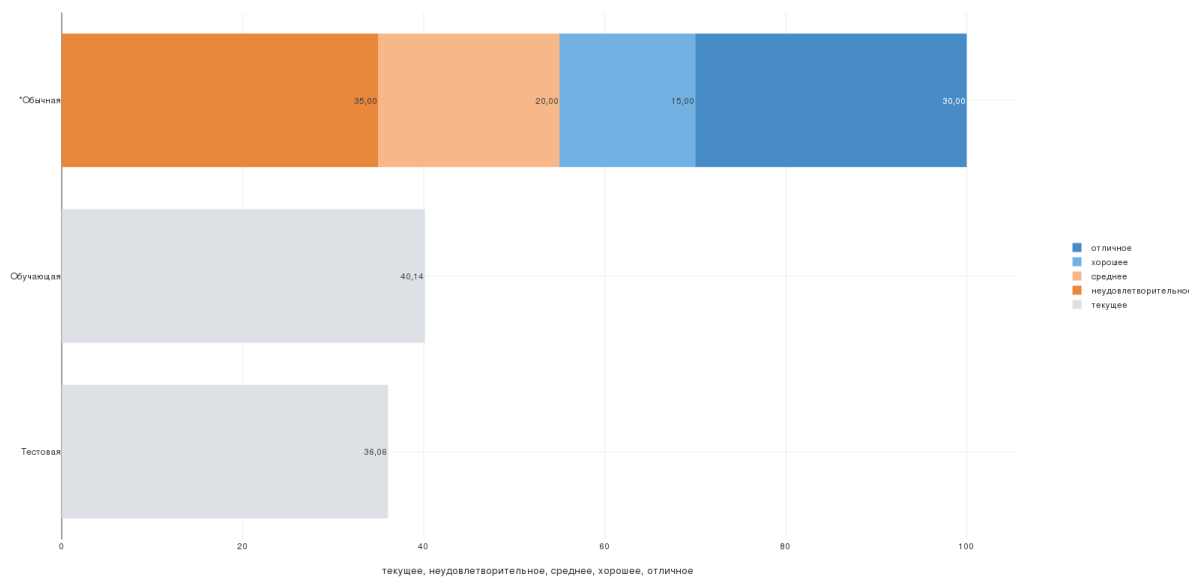
- **Поле X:** Номер интервала
- **Поле Y:** Идентификатор (Количество)
- **Поле меток:** Интервал балла

САР-кривая



- **Поле X:** Объем выборки
- **Поле Y1:** Идеальная линия
- **Поле Y2:** Событие кумулятив., %
- **Поле Y3:** Случайное угадывание

Индекс Gini

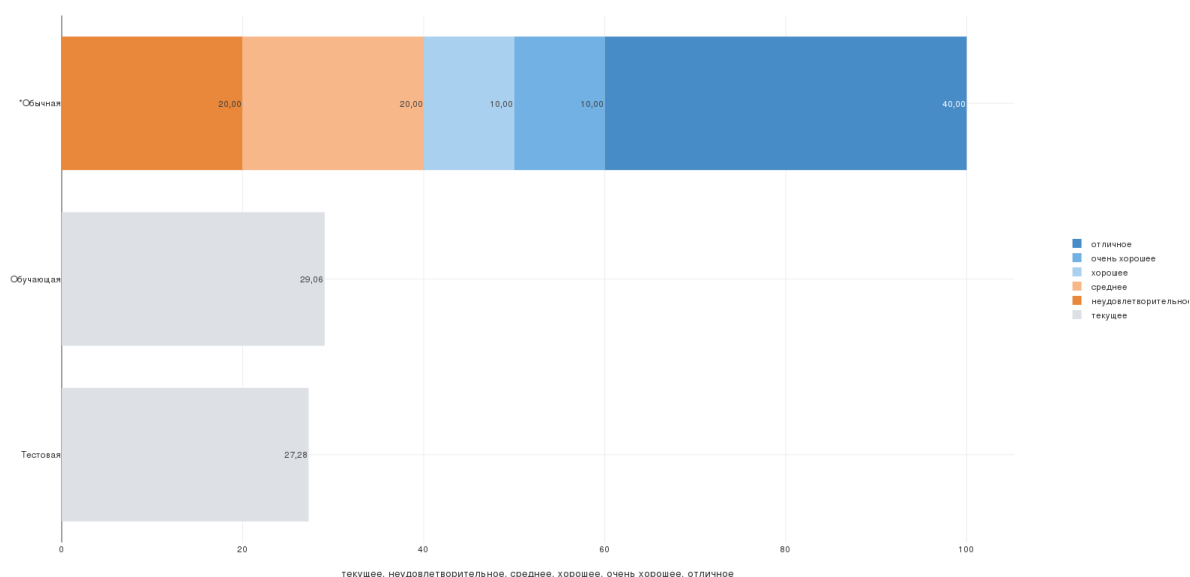


- **Поле Y1:** отличное
- **Поле Y2:** хорошее
- **Поле Y3:** удовлетворительное
- **Поле Y4:** неудовлетворительное

- Поле Y5: текущее
- Поле меток: Тип модели

Примечание: В разделе *Общие настройки* необходимо указать Тип накопления: Обычное.

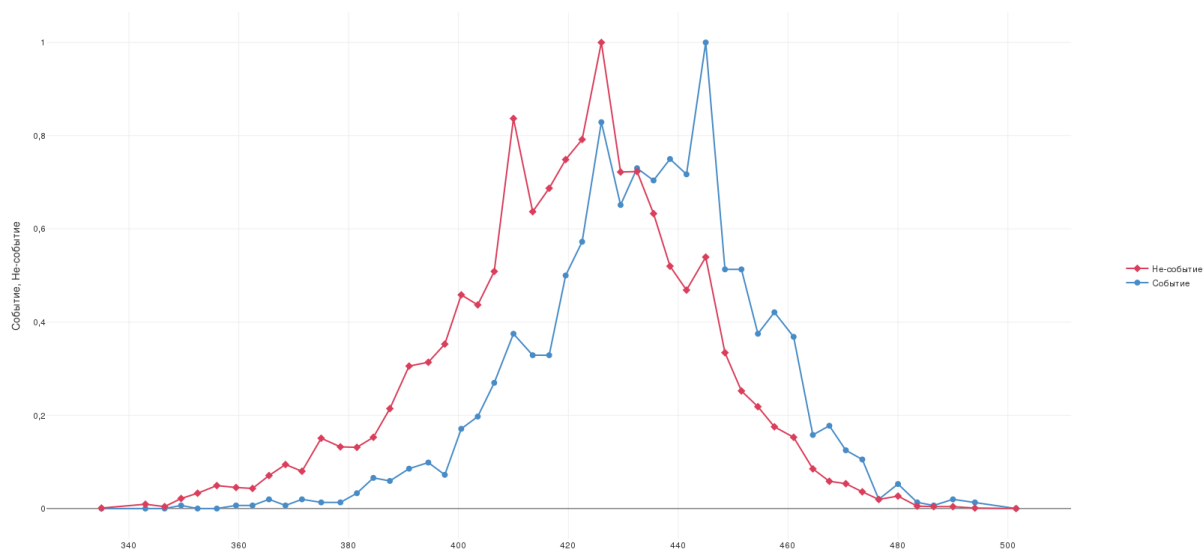
Статистика KS



- Поле Y1: отличное
- Поле Y3: очень хорошее
- Поле Y4: хорошее
- Поле Y5: удовлетворительное
- Поле Y6: неудовлетворительное
- Поле Y7: текущее
- Поле меток: Тип модели

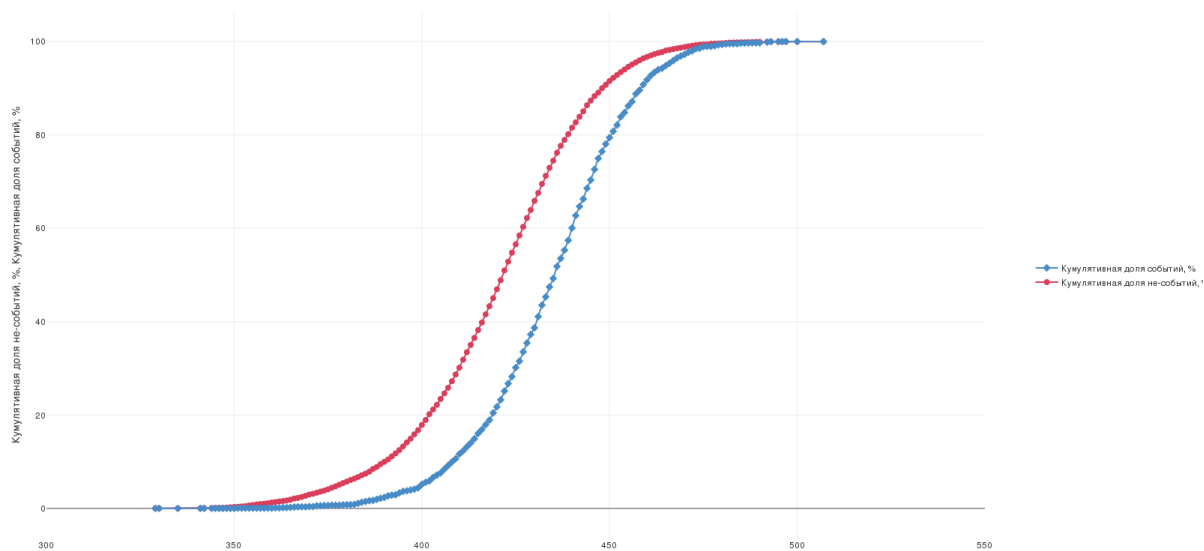
Примечание: В разделе *Общие настройки* необходимо указать Тип накопления: Обычное.

Распределение счетов с событием и не-событием



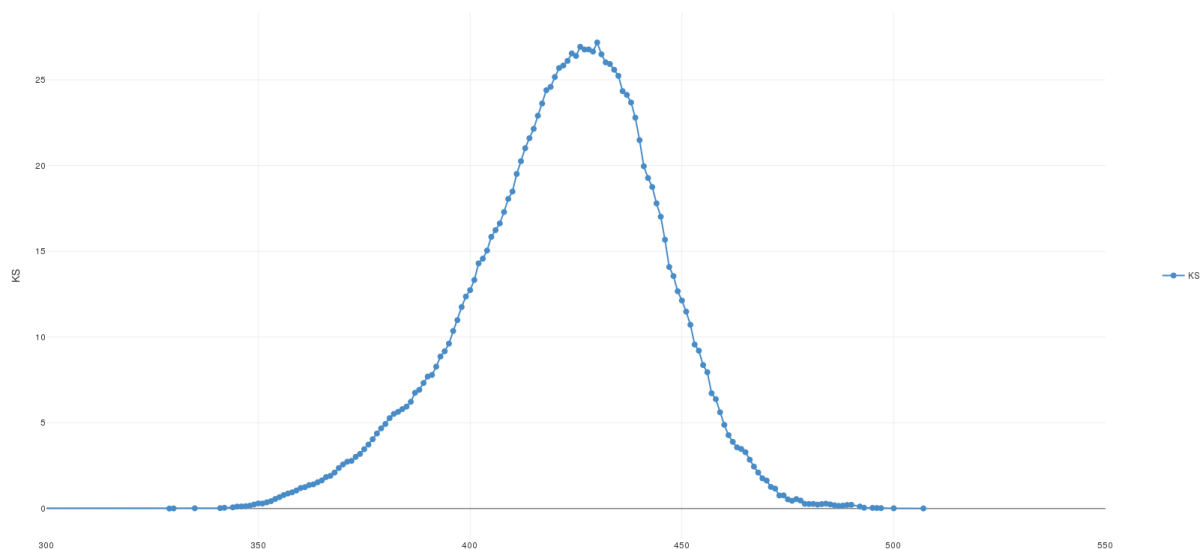
- Поле X: Балл
- Поле Y1: Событие
- Поле Y2: Не-событие

Кумулятивное распределение счетов



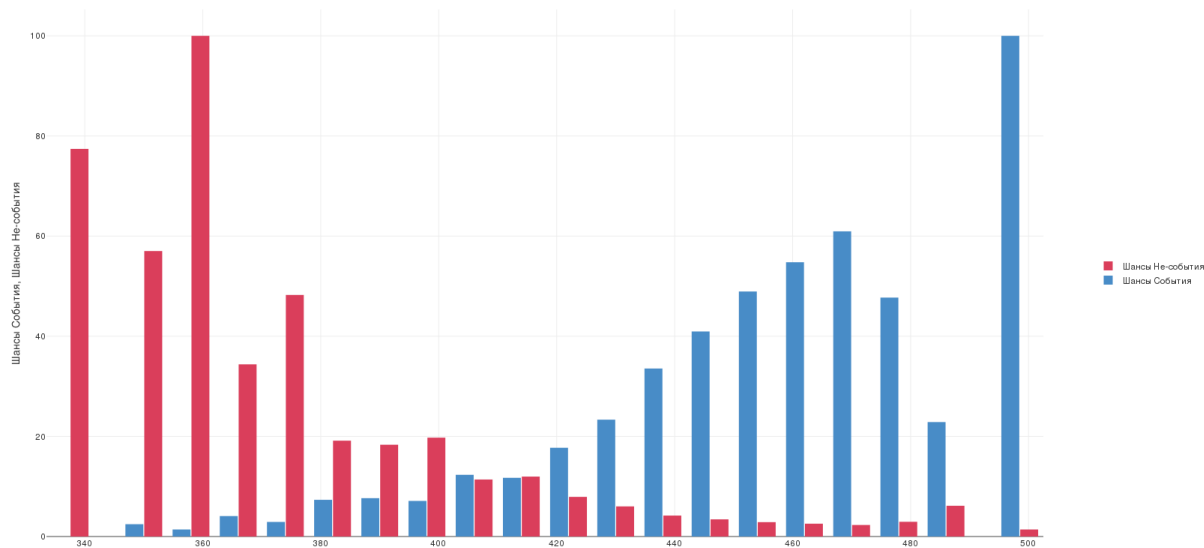
- Поле X: Балл
- Поле Y1: Кумулятивная доля событий, %
- Поле Y2: Кумулятивная доля не-событий, %

KS-кривая



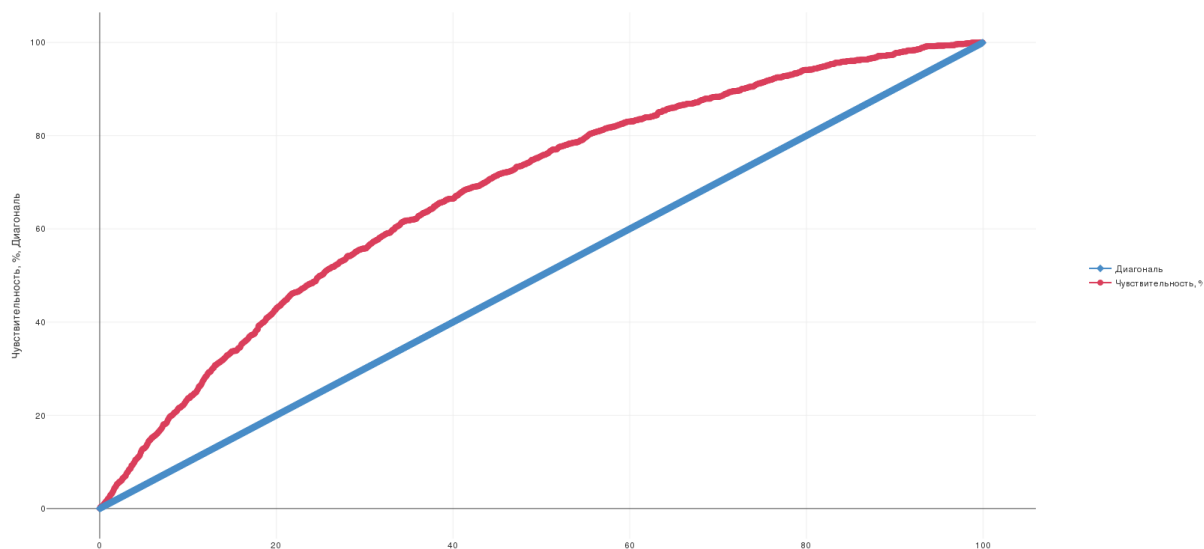
- Поле X: Балл
- Поле Y: KS

Шансы событий/не-событий (рисковые сегменты)



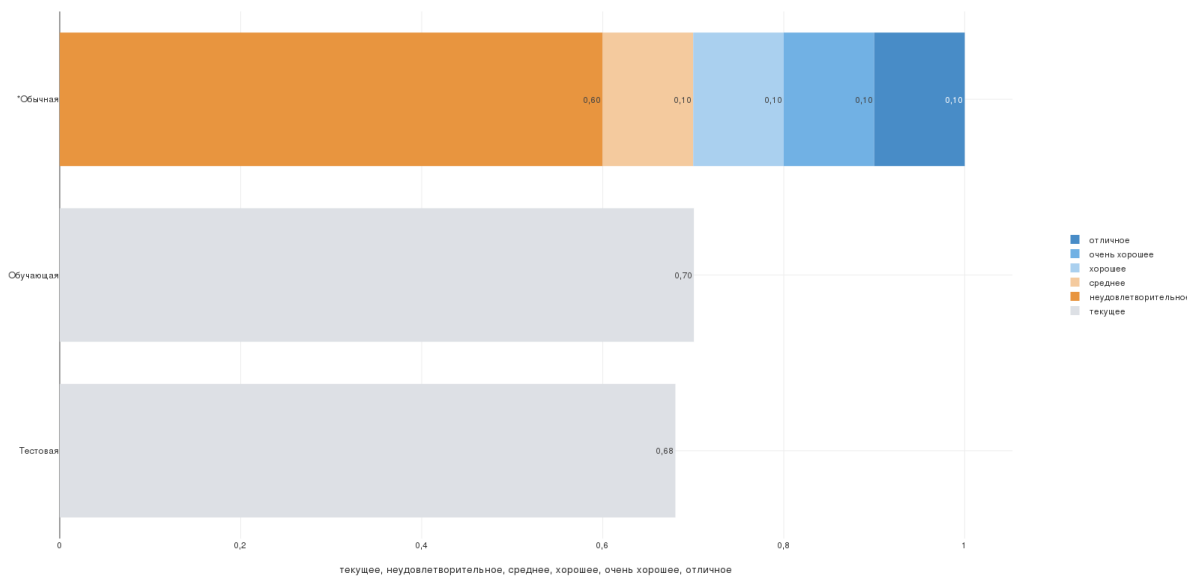
- Поле X: Балл
- Поле Y1: Шансы Не-События
- Поле Y2: Шансы События

ROC-кривая



- Поле X: % ложноположительных
- Поле Y1: Чувствительность, %
- Поле Y2: Диагональ

AUC



- Поле Y1: отличное
- Поле Y3: очень хорошее
- Поле Y4: хорошее
- Поле Y5: удовлетворительное
- Поле Y6: неудовлетворительное

- **Поле Y7:** текущее
- **Поле меток:** Тип модели

Примечание: В разделе *Общие настройки* необходимо указать *Тип накопления: Обычное*.

Loginom Skills

Методология и подробные кейсы использования в виде пошаговых симуляций в Loginom рассматриваются в бизнес-курсе [Оценка качества скоринговой карты](#).

История изменений документа

№	Дата	Описание
1	28.03.2023	Первая версия документации