

# Алгоритмы машинного обучения: как применять и рассчитать эффект

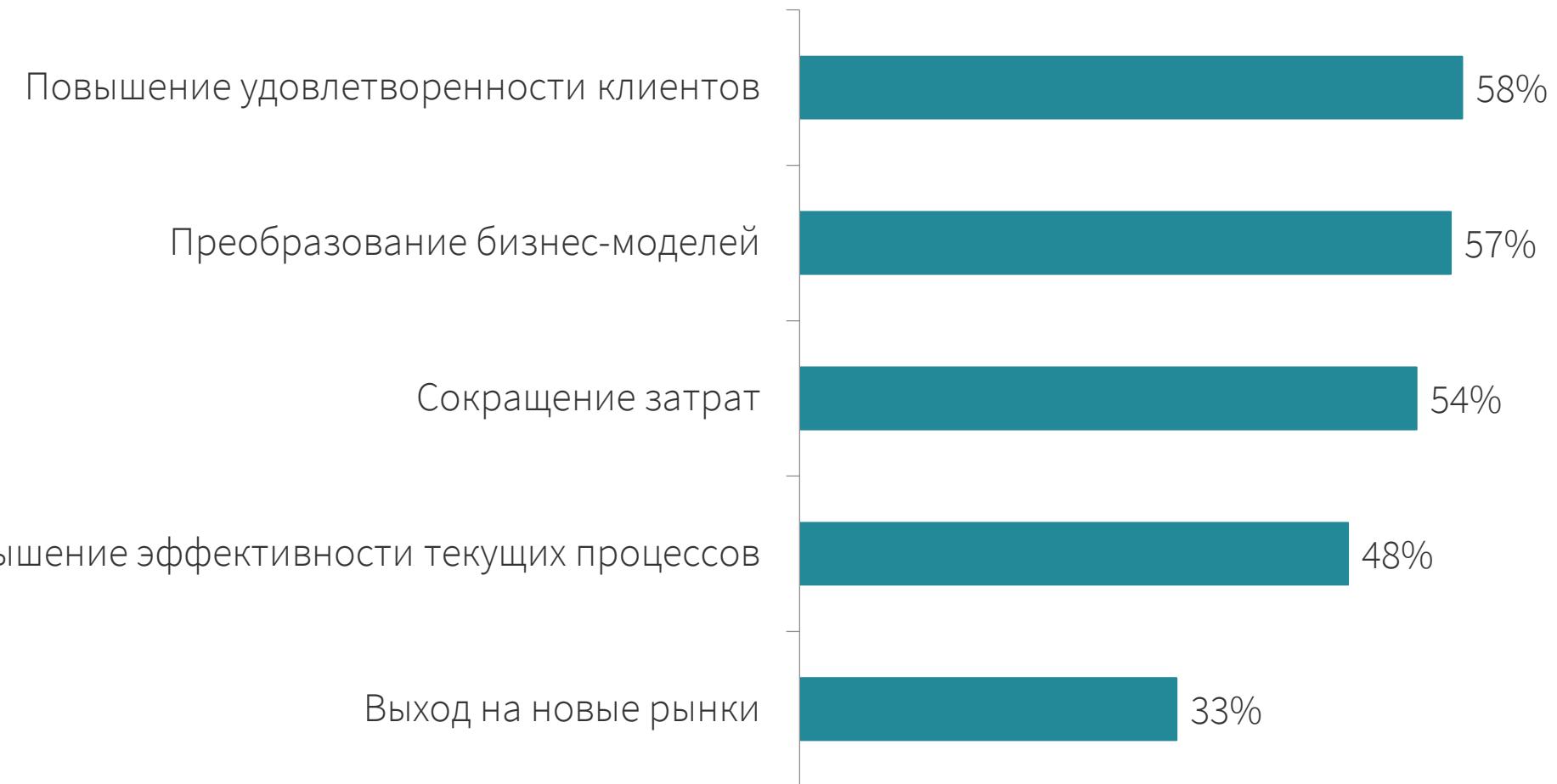
Арутамов Алексей

Loginom Company

Цифровая трансформация — это процесс интеграции цифровых технологий во все аспекты деятельности организаций:

- Бизнес-процессы
- Продукты и услуги
- Культура производства
- Коммуникации

# Основные цели цифровой трансформации



Исследование Forbes Insights и Hitachi Data Systems , опрос 573 топ-менеджеров компаний из разных стран мира



Машинное обучение –  
ключевой компонент  
цифровой трансформации  
для гибкого и быстрого  
принятия оптимальных  
решений с минимальным  
участием людей.

# Препятствия к цифровой трансформации

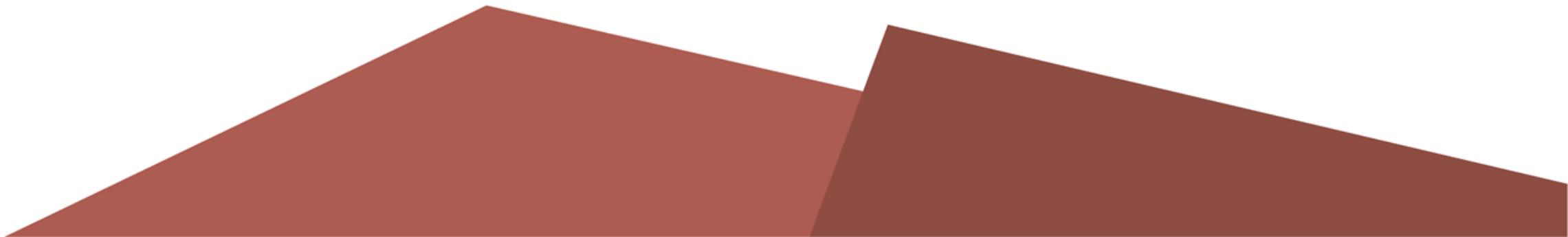


Исследование Forbes Insights и Hitachi Data Systems , опрос 573 топ-менеджеров компаний из разных стран мира



Возможность обосновать  
эффективность в деньгах  
значительно повышает  
готовность реально  
применять машинное  
обучение на практике

Пример: оценка вероятности  
оттока клиентов –  
доказательство окупаемости



## Вариант 1: Удачные кейсы

Логика: Компания «Х» благодаря машинному обучению снизила отток клиентов на 20% –  
таков эффект систем противодействия оттоку.

Проблема: Игнорирование систематической ошибки выжившего.

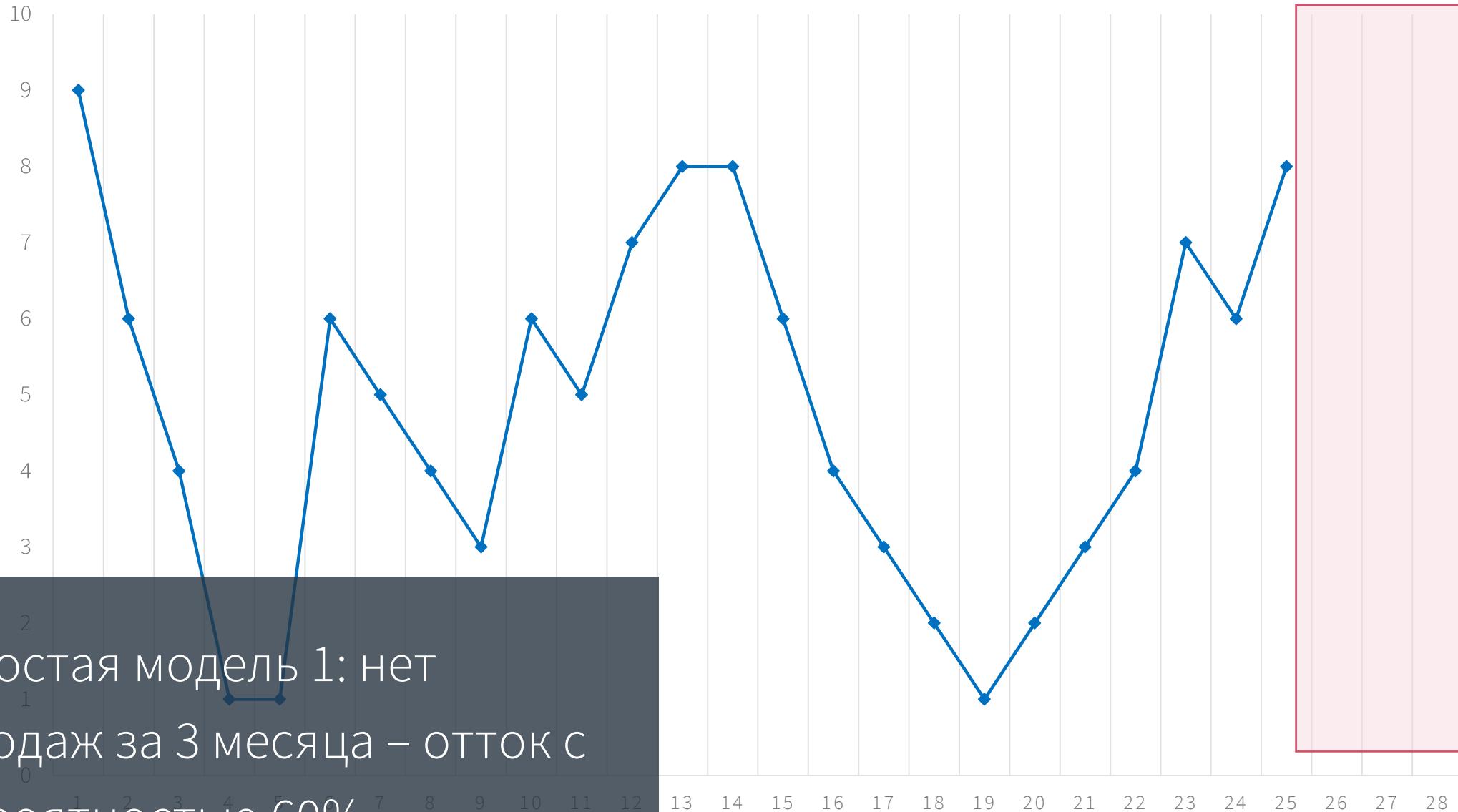


Хорошо известны удачные кейсы,  
хотя большинство попыток не  
приносят желаемой отдачи

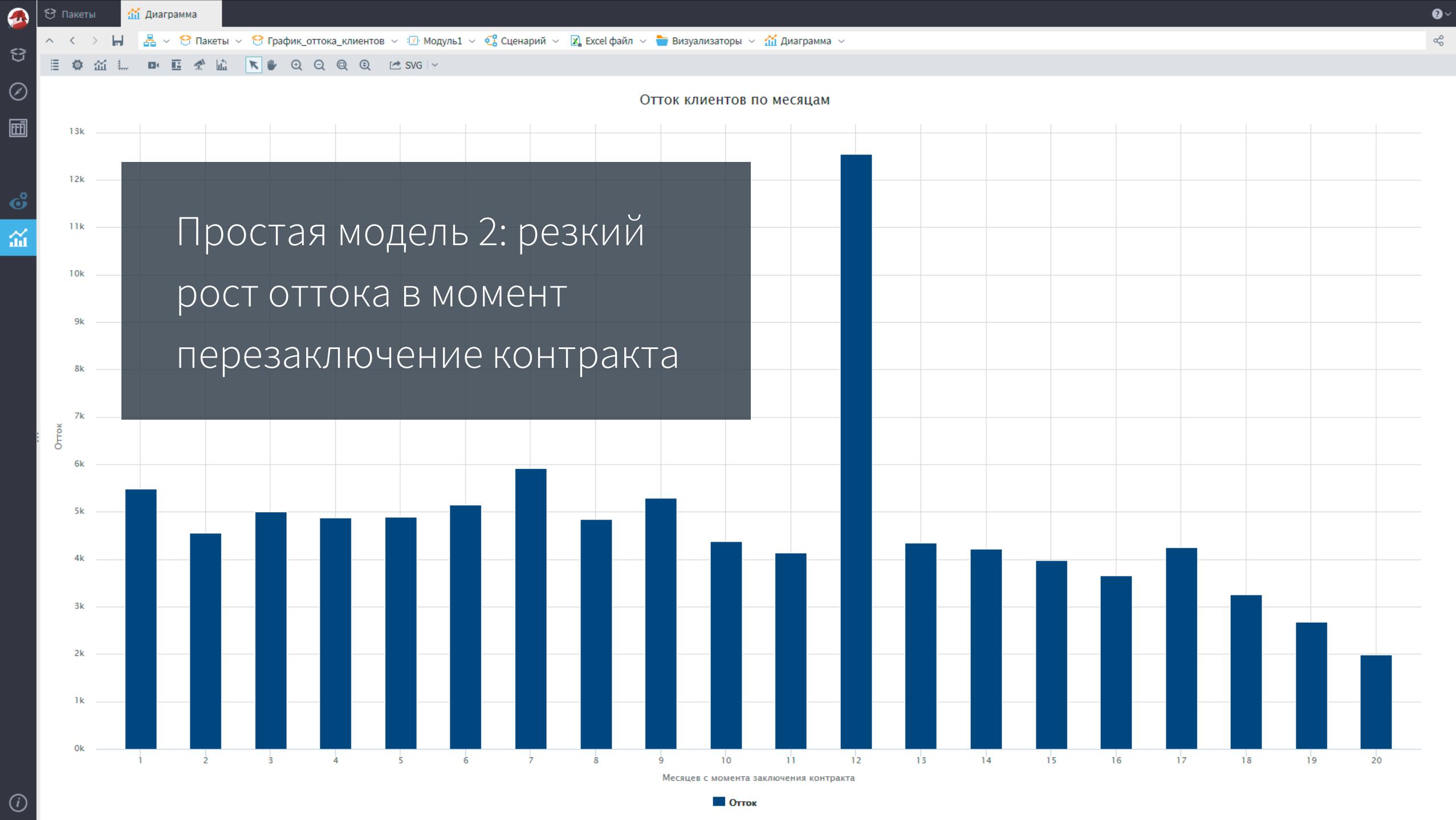
## Вариант 2: Сравнение с нулем

Логика: Машинное обучение обнаружило 80% склонных к оттоку, все доходы от удержанных – эффект от применения модели.

Проблема: Игнорирование простых методов предсказания оттока клиентов.



Простая модель 1: нет  
продаж за 3 месяца – отток с  
вероятностью 60%



## Вариант 3: А/В тестирование

Логика: Создаем 2 группы одну оставляем без изменений, а для второй применяем ML модель, разница между ними – эффект от модели машинного обучения.

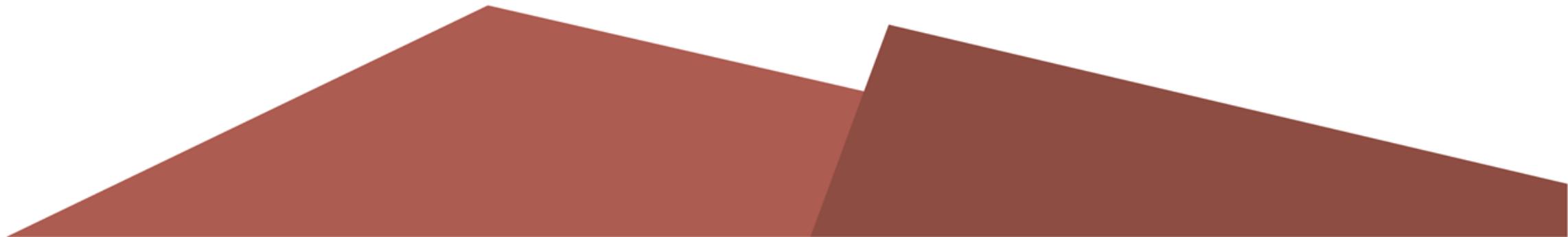
Проблема: нужен работающий процесс со встроенными инструментами А/В тестирования.

## Вариант 4: Сравнение с тривиальной моделью

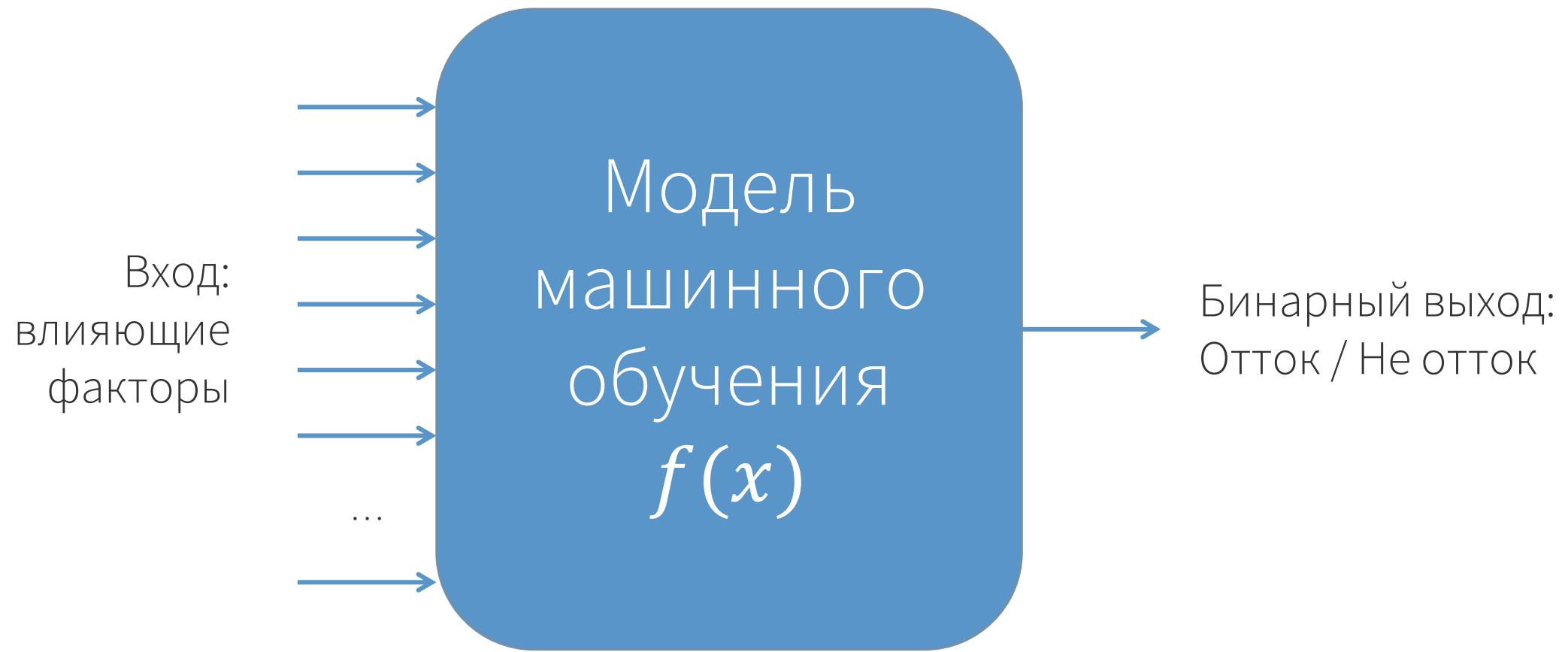
Логика: На одном наборе данных строим тривиальную модель и модель с машинным обучением, разница между ними – эффект от модели машинного обучения.

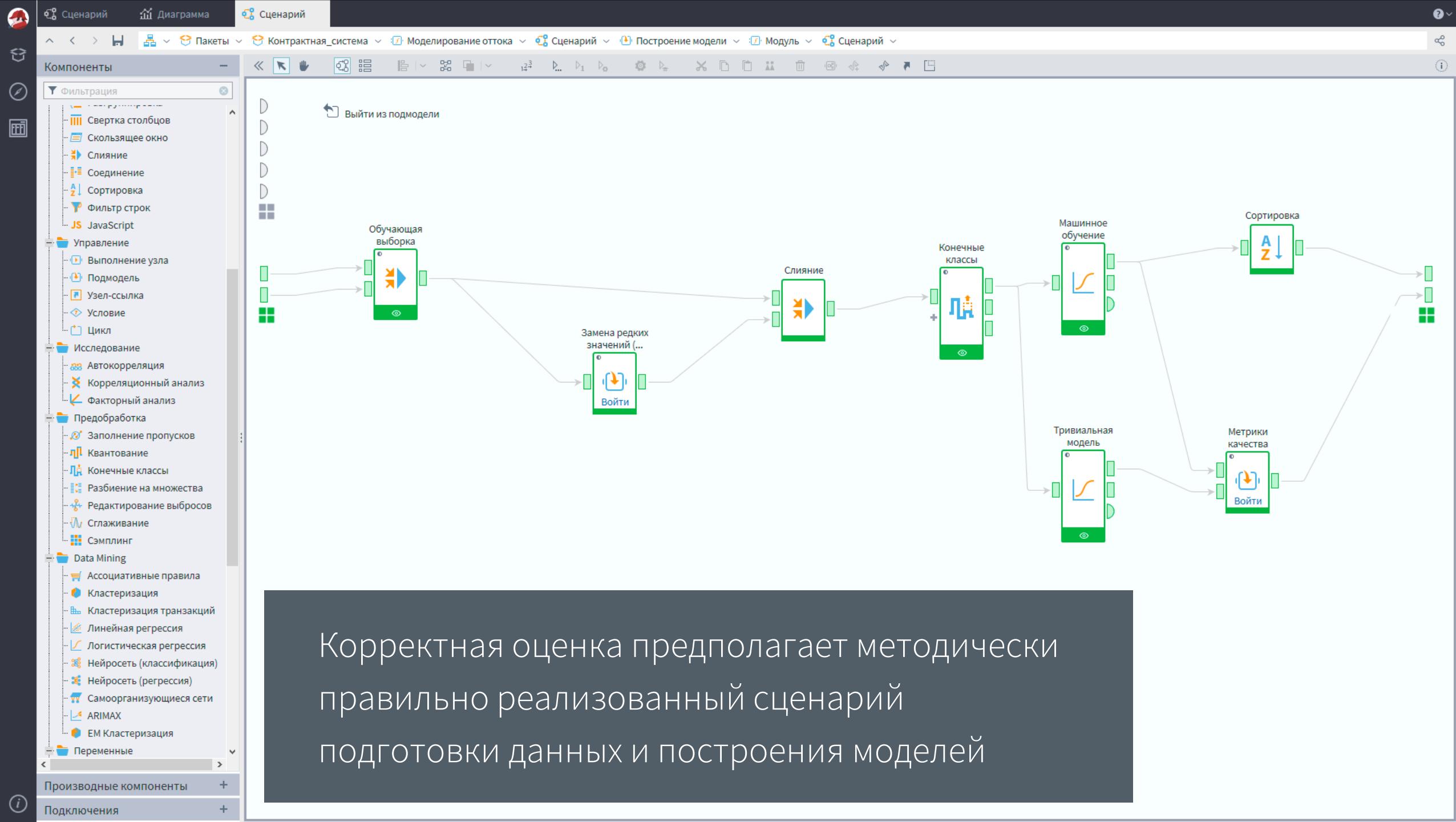
Проблема: эффект ML модели зависит от «тривиальности» простой модели.

# Кейс: Сравнение с тривиальной моделью



# Задача бинарной классификации





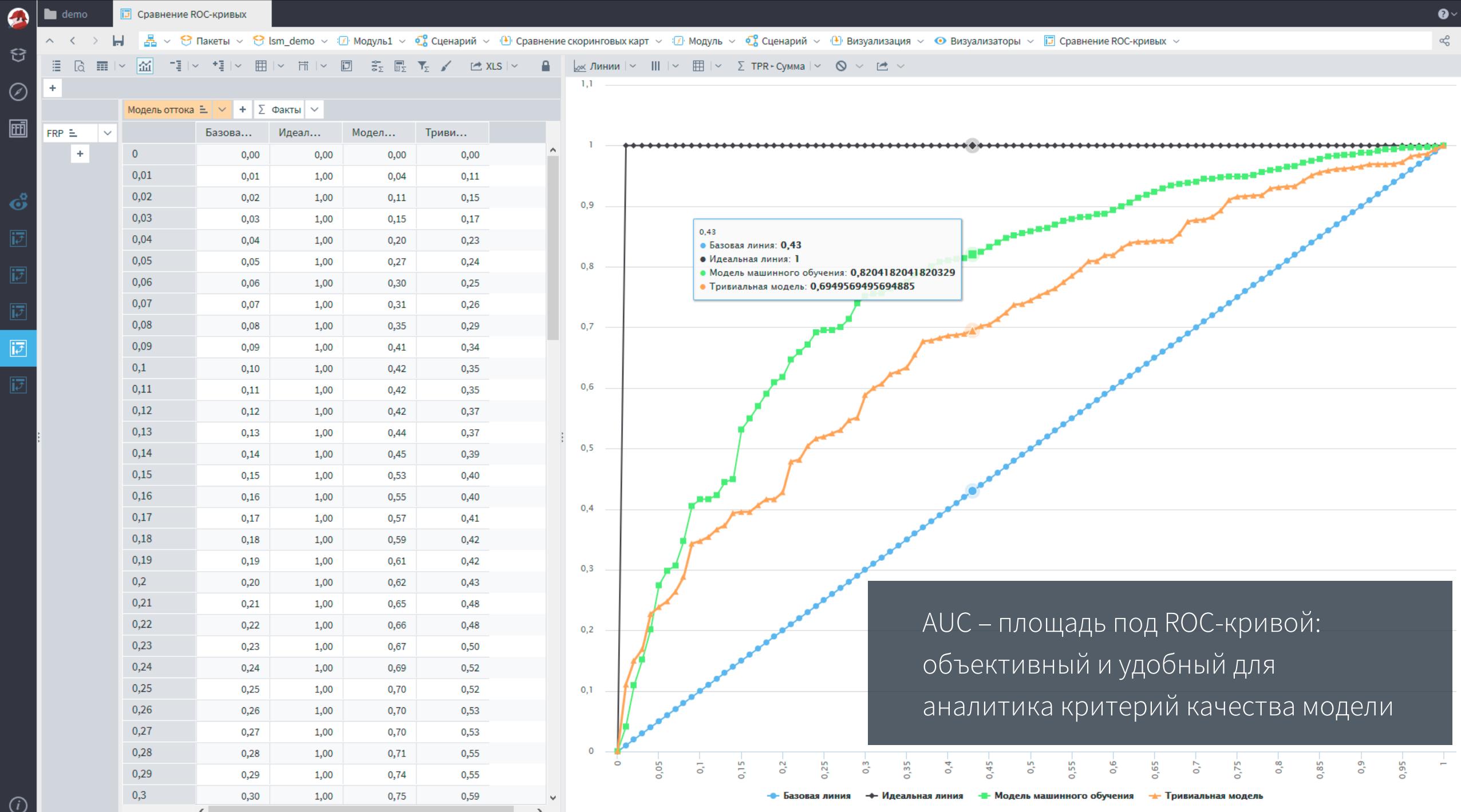
## Тривиальная модель

1. Простое бизнес-правило
2. Использование одного значимого фактора
3. Минимум внешней информации

## Сложная модель

1. Применение сложных алгоритмов
2. Использование множества факторов
3. Обогащение данных перед построением модели

VS



# Площадь под ROC-кривой

Достоинство: стандартный и общепринятый в машинном обучении критерий сравнения нескольких моделей.

Недостаток: нет простого способа оценки разницы моделей в деньгах.

# Ошибка первого и второго рода

## Тривиальная модель

Модель	Факт		
	Отток	Не отток	Итого
Отток	6 512 (63.9%)	856 (1.1%)	7 368
Не отток	3 667 (36.1%)	78 903 (98.9%)	82 570
Итого	10 179	79 759	89 938

## Сложная модель

Модель	Факт		
	Отток	Не отток	Итого
Отток	7 025 (69.0%)	343 (0.4%)	7 368
Не отток	3 154 (31.0%)	79 416 (99.6%)	82 570
Итого	10 179	79 759	89 938

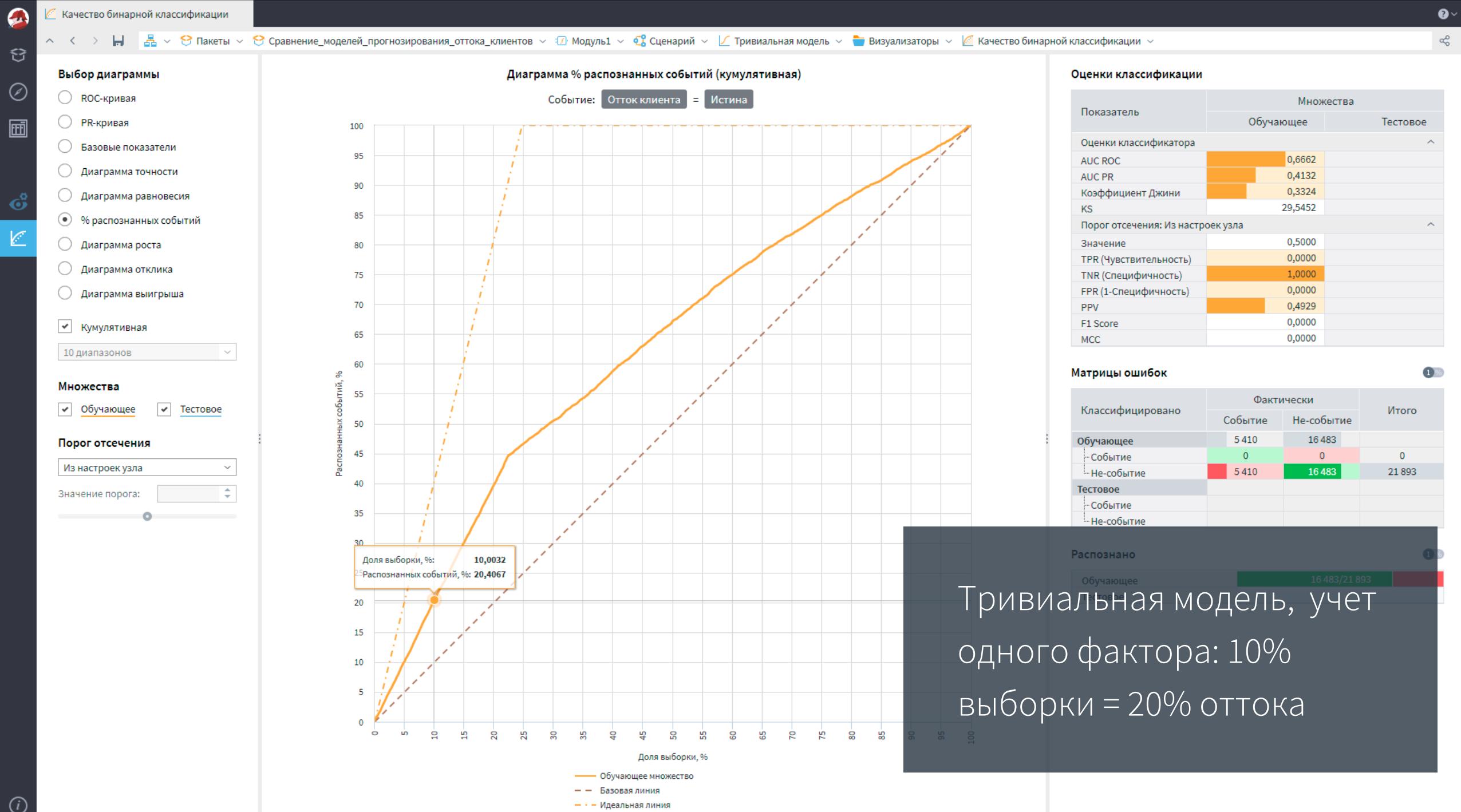
# Ошибка первого и второго рода

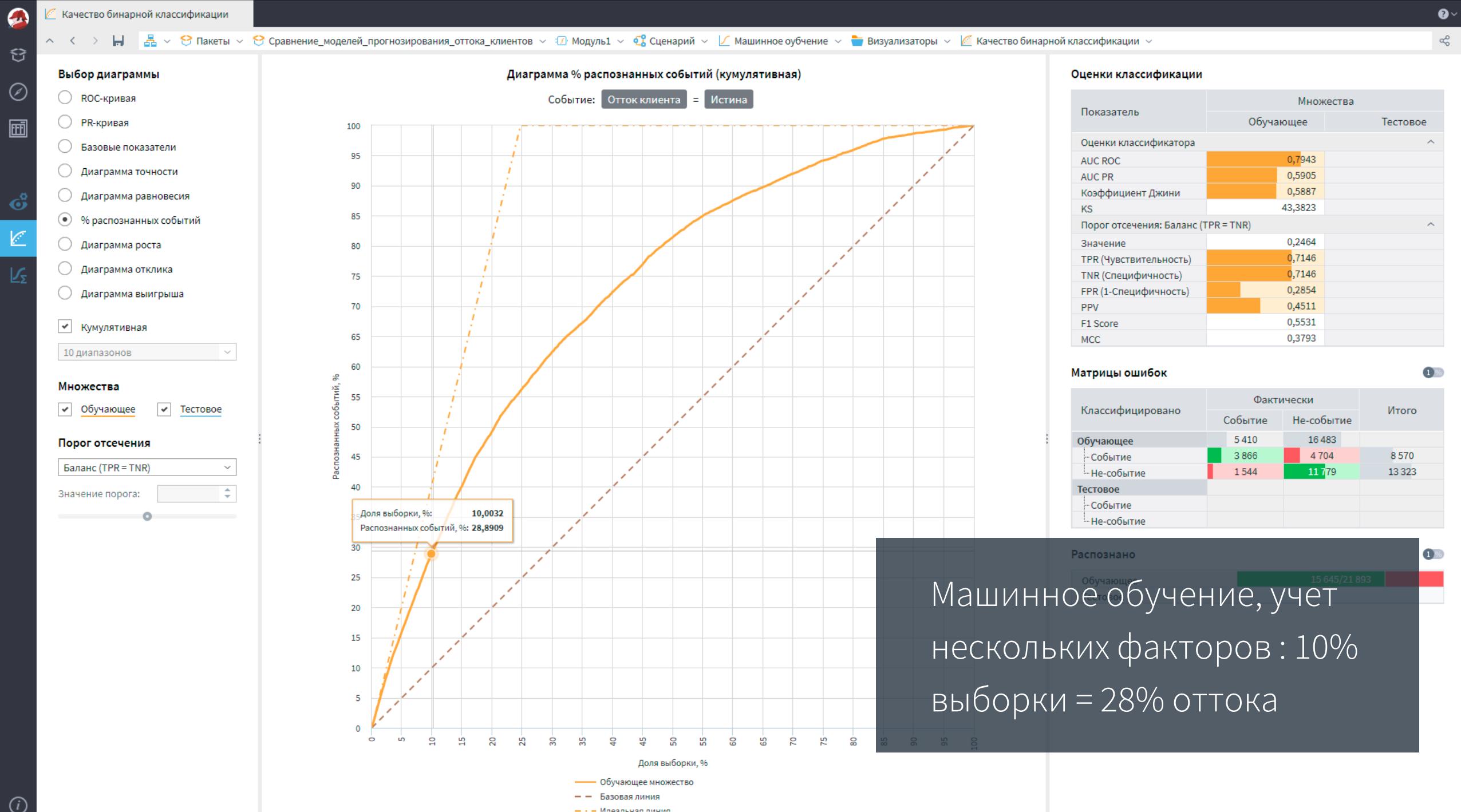
Достоинство: стандартный критерий сравнения нескольких моделей, позволяющий учесть цену ошибки.

Недостаток: сложно цену ошибки напрямую свести к финансовым потерям.

## Процент распознанных событий

После построения модели для каждого клиента рассчитывается вероятность оттока, записи сортируются по убыванию вероятности и строится кумулятивный график угаданных моделью случаев оттока.





Показатель	Значение
Активных клиентов	21 893
Средний отток	960
Процент оттока	4.38%
Средний доход на год (руб.)	50 000
Потери от оттока	48 000 000
Стоимость удержания одного клиента	1 000
На какой процент клиентов будем воздействовать	10.00%
Процент удержания после воздействий	50.00%
<b>Тривиальная модель</b>	
Распознали склонных к оттоку	20.40%
На сколько клиентов надо воздействовать	2 190
Стоимость мероприятий по удержанию	2 190 000
Сколько удержим (в клиентах)	98
Сколько удержим (в рублях)	4 896 000
Выгода от удержания	2 706 000
<b>Машинное обучение</b>	
Распознали склонных к оттоку	28.90%
На сколько клиентов надо воздействовать	2 190
Стоимость мероприятий по удержанию	2 190 000
Сколько удержим (в клиентах)	139
Сколько удержим (в рублях)	6 936 000
Выгода от удержания	4 746 000
Выгода от ML модели	2 040 000

## Процент распознанных событий

Достоинство: можно свести к понятной бизнесу методике оценке эффекта до начала применения модели.

Недостаток: требуются определенные допущения для расчета эффекта.

Описанные методики позволяют, хоть и с некоторыми допущениями, доказать окупаемость инвестиций в машинное обучение повышают вероятность их реального применения для решения практических задач.

loginom.ru