Antifraud

team 2

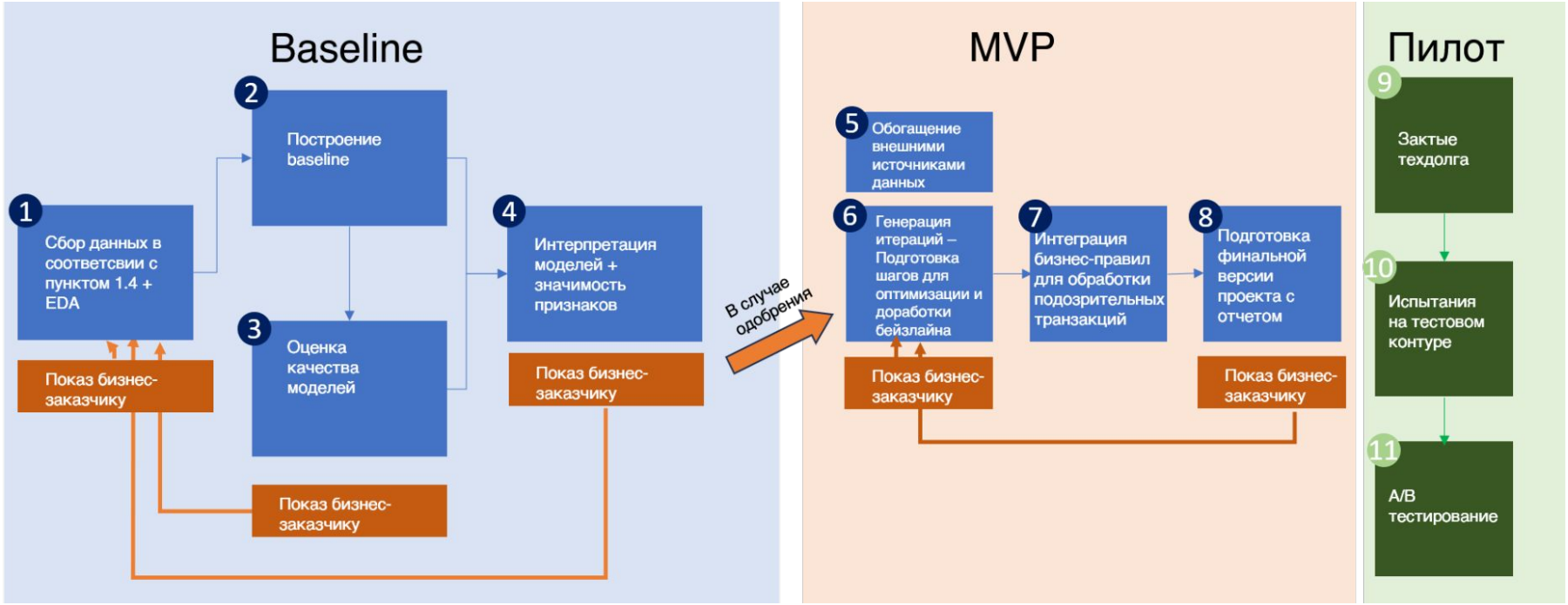
Александр Ковязин DS, СПБГУ, Sber
Александра Оберемок, DS, Samokat.Tech
Антон Петров DS
Ханду Ринчинова НГУ

Методология

Постановка задачи

Бинарная классификация + второй шаг (выходное значение меньше определенного порога отправляем на валидацию аналитикам)

Блок-схема решения



Этапы решения задач

Сбор данных

Внутренние данные

- Переводы физических лиц
- Данные клиента
- Данные устройства, с которого переводят
- Данные по “черным спискам” из имеющейся rule-based системы
- Данные по движению средств на счетах
- История транзакции клиента

Внешние данные

- Кредитная история
- Информация по клиенту от налоговой службы (сколько платит налогов)
- Наличие у человека ИП/ООО/Юрлиц/Банкростства физического лица
- Справка о доходах/род деятельности
- Данные службы судебных приставов + данные от коллекторов (если есть)

Этапы решения задач

Требования к данным + EDA

Требования к данным

- Полнота данных, процент пропусков < 5%
- Реплицированные данные совпадают с первоисточником
- Исторические данные за 2 года
- Маппинг клиентов из разных источников

EDA

- корреляцию, отбрасываем коррелированные данные
- анализ категориальных признаков - one-hot, label-enc, KNN
- пропущенные столбцы-строки - изучаем характер пропуска, если случайные (< 5% удаляем, если > заполняем среднее/медиана), если пропуски, зависящие от значения признака - то флаг
- проверка наличия дубликатов
- выбросы - обрабатывает по той же логике, что и пропущенные
- смотрим на распределения внутри признаков (если по графику видим не нормальное распределение, то преобразуем, например, логарифмируем)

Этапы решения задач

Бейзлайн

Бэйзлайн предсказание вероятности фродовой транзакции, ориентир для сравнения качества моделей.

Простое прогнозирование эвристиками из rule-based модели либо дамми классификатор, бинарная классификация

Таргет - фрод или нет от аналитиков или по обращениям/жалобам

Риски

- сильный дисбаланс,
- могут прийти данные сильно отличающиеся от тренировочной выборки

Метрики - F-мера (Beta>1) macro,

Мы хотим снижать сумму штрафов, что коррелирует с технической метрикой f-metric с уклоном на recall берем коэффициент > 1 и из-за дисбаланса классов берем в расчет macro.

Pipeline обучения MVP

Разбиение выборок - train, OOS (shuffle) - без пересечения по клиентам, стратифицированно
Гранулярность - отдельная транзакция
Частота - ASAP, переобучение раз в сутки ночью, когда меньше всего нагрузки
Таргет - фрод или нет, по обращениям/жалобам/наличие штрафа/аналитики
Метрики - F-мера (Beta>1) macro выше бейзлайна
Риски

- неправильное распределение таргета из-за отложенной разметки
- дороговизна внешних данных,
- разреженные данные,
- некорректные данные,
- вероятность поймать даталики,
- меняются виды мошеничества

Feature eng

- разбиваем на элементы дату и время, адрес (если геоданные),
- склонность к определенным транзакциям,
- собираем статистики по транзакциям за предыдущие периоды,
- сегменты пользователей, которые делают переводы др др,
- частота взаимодействия между отправителем и получателем

Алгоритм -

- кросс-валидация group stratified kfold,
- бустинг (class weighted) + логрег для калибровки,
- подбор гиперпараметров hyperopt,
- важность фичей (permutation importance, SHAP),
- отбор фичей по вкладу в метрику

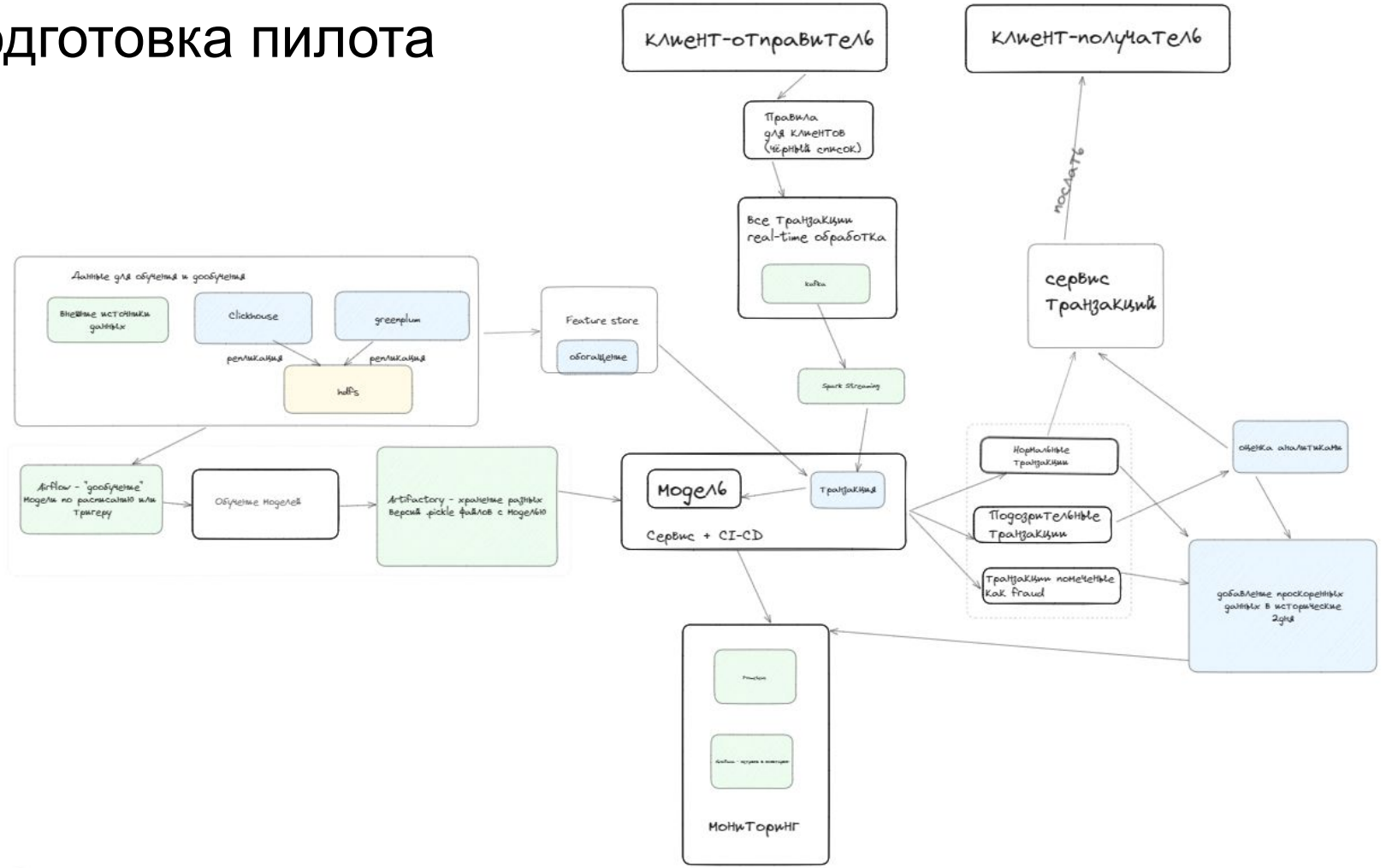
Оценка качества модели

- на отличных выборках, применяем ту же предобработку и проверяем метрики,
- проверка на чувствительность к шуму,
- оценка качества на разных кластерах (по регионам, полу, возрасту)

Демо - интерпретация (согласование с бизнес-логикой)

Вопросы

Подготовка пилота



Ссылка https://excalidraw.com/#room=dba2f5245eb88593a88e,LJEUq833bL-qBCMUNtavkg

Подготовка пилота - АВ тестирование

Критерий успеха - Сокращение суммы штрафов на Y%

H0 - эффекта от МЛ системы нет

H1 - эффект есть, уменьшится сумма штрафов

Мощность = 80%, значимость = 0.05, MDE условно 5% (Y)

1. Размер выборки
2. Продолжительность (размер_выборки * 2/100_000 транзакций) дней
3. Разбили клиентов на 2 равные группы рандомно.
4. Эксперимент - на контрольной ничего не меняем, на тестовой выборке ML-решение.
5. Тест Манна-Уитни
6. Стат значимость по p-value

Вычислительные мощности:

- сервер на 200 gb, Ram 30Gb, 8 cpu для baseline расчетов
- в пилоте для 100_000 транзакций в день, соответственно, соблюдать пропорцию на большее количество

Вопросы