

АНАЛИЗ ПОВЕДЕНИЯ ПОЛЬЗОВАТЕЛЕЙ

ЧТО ТАКОЕ ПОЛЬЗОВАТЕЛЬСКИЕ ДАННЫЕ?



- › Персональные данные
- › Данные о взаимодействии с продуктом!

ЗАЧЕМ АНАЛИЗИРОВАТЬ ПОЛЬЗОВАТЕЛЕЙ

- › Понять (и простить) пользователей
- › Оценить ключевые характеристики аудитории
- › Работать с аудиторией

- › Описание целевой аудитории
- › Привлечение пользователей
- › Работа с аудиторией:
 - ▶ вовлечение
 - ▶ влияние на ключевые показатели
- › Прогнозирование оттока
- › Удержание пользователей



- › Выделение ядра аудитории
- › Сегментация пользователей

ПРИВЛЕЧЕНИЕ ПОЛЬЗОВАТЕЛЕЙ



- › Что предлагаем? способы привлечения
- › Как связываемся? каналы взаимодействия
- › Какой результат? поведение пользователей в разрезе каналов привлечения
- › Давайте оптимизировать!

РАБОТА С АУДИТОРИЕЙ



- › Анализ пользовательского опыта
- › Основные сценарии
- › Мотивация достижения целей





ЧЕГО НЕ БУДЕТ

- Привлечение пользователей
- Продвижение
- Работа с аудиторией
- Оптимизация сайта, SEO-оптимизация
- Реклама и маркетинг

- Анализ пользовательских данных:
 - ▶ Описание и сегментация аудитории
 - ▶ Аудиторные метрики
 - ▶ Прогнозирование оттока

АУДИТОРНЫЕ МЕТРИКИ: ПРИВЛЕЧЕНИЕ

ПРИВЛЕЧЕНИЕ

- › Количество пользователей
- › Конверсии
- › Стоимость привлечения пользователей
- › Эффективность инвестиций

КОЛИЧЕСТВО ПОЛЬЗОВАТЕЛЕЙ



- › Total users
- › New users

- › Конверсия первого дня — % новых пользователей, вернувшихся хотя бы 1 раз после дня регистрации
- › Конверсия первой недели
- › Конверсия первого месяца

- Конверсии можно считать между чем угодно!
 - ▶ Конверсия игроков N -го уровня в игроков $N+k$ уровня
 - ▶ Конверсия зарегистрированных пользователей в покупателей

СТОИМОСТЬ ПРИВЛЕЧЕНИЯ ПОЛЬЗОВАТЕЛЯ



- › CPA — cost per acquisition
- › $CPA = (\text{total advertisement spend}) / (\text{number of registered users})$
- › CPI — cost per install
- › $CPI = (\text{total advertisement spend}) / (\text{number of installs})$

- » Кого считать привлечённым пользователем?
 - ▶ посещение страницы?
 - ▶ регистрация?
 - ▶ совершение целевого действия?

- › ROI – return on investment
- › $ROI = (total\ revenue - total\ cost) / (total\ cost) * 100$
- › Investment A: total revenue – 1100\$, total cost – 1000\$
- › Investment B: total revenue – 2500\$, total cost – 2000\$

ЭФФЕКТИВНОСТЬ ИНВЕСТИЦИЙ

- $ROI(A) = (1100 - 1000)/1000 * 100 = 10\%$
- $ROI(B) = (2500 - 2000)/2000 * 100 = 25\%$
- Важно учитывать длительность периода оценки эффективности инвестиции

- › Total users
- › New users
- › Conversion
- › CPA, CPI
- › ROI

АУДИТОРНЫЕ МЕТРИКИ: АКТИВНОСТЬ

АКТИВНОСТЬ



- › Количество активных пользователей
- › Сессионные метрики
- › Временные метрики

КОЛИЧЕСТВО АКТИВНЫХ ПОЛЬЗОВАТЕЛЕЙ



- › DAU (daily active users)
- › WAU (weekly active users)
- › MAU (monthly active users)

КОЛИЧЕСТВО АКТИВНЫХ ПОЛЬЗОВАТЕЛЕЙ

- Насколько стабильна аудитория?
 - ▶ Sticky factor = DAU/MAU
 - ▶ Loyal users

КОЛИЧЕСТВО АКТИВНЫХ ПОЛЬЗОВАТЕЛЕЙ

- › *L – loyal users
- › DAU → LDAU
- › WAU → LWAU
- › MAU → LMAU
- › Каких пользователей считать лояльными?

СЕССИОННЫЕ ПОКАЗАТЕЛИ

- › Сессия — последовательность действий пользователя на сервисе в рамках одного визита
- › Средняя длина сессии = $\frac{\text{суммарная длина всех сессий}}{\text{количество сессий}}$
- › Хорошо оценивать в разрезе сегментов

- › **Lifetime** – количество дней между первым и последним визитами, усредненное по пользователям
- › Как определить, что визит последний?
 - ▶ Явные сигналы
 - ▶ Неявные сигналы

- › DAU, WAU, MAU
- › LDAU, LWAU, LMAU
- › Average session length
- › Lifetime

АУДИТОРНЫЕ МЕТРИКИ: МОНЕТИЗАЦИЯ

- › Доли и конверсии
- › Платежи
- › Доход

- › Paying share
- › Paying conversion

- › Общее количество платежей (total transactions)
- › Среднее количество платежей на пользователя (transactions by user)

ДОХОД

- › Gross — доход
- › Revenue — прибыль, "доход" — "затраты"

СРЕДНИЙ ДОХОД



- › **ARPU** — average revenue per user
- › **ARPPU** — average revenue per paying user
- › **Average Check**

LIFETIME VALUE

- **LTV** — Lifetime Value — сколько в среднем денег принёс пользователь за всё время использования сервиса
- Зависит от:
 - ▶ lifetime
 - ▶ **ARPU**

- › Paying share, paying conversion
- › Total transactions, transactions by user
- › Gross, revenue
- › ARPU, ARPPU, Average Check
- › Lifetime Value

АУДИТОРНЫЕ МЕТРИКИ: УДЕРЖАНИЕ

УДЕРЖАНИЕ И ОТТОК



- » Возвращаемость
- » Отток

- › Return rate
- › $RR = (\text{current number of customers from the original set}) / (\text{number of customers at the original set}) * 100$

ВОЗВРАЩАЕМОСТЬ



- › 1-day retention
- › 7-day retention
- › 28-day retention

ВОЗВРАЩАЕМОСТЬ



- › 1-day rolling retention
- › 7-day rolling retention
- › 28-day rolling retention

› Churn rate

› $CR = (\text{number of churned customers}) / (\text{total number of customers}) * 100$

- › Return rate
- › X-day retention
- › Rolling retention
- › Churn rate

ПРОГНОЗИРОВАНИЕ ОТТОКА ПОЛЬЗОВАТЕЛЕЙ

ОПИСАНИЕ КЕЙСА

- › Больше пользователей приносят больше прибыли
- › Больше пользователей = больше новых + меньше отток существующих
- › Часто удержать одного пользователя стоит дешевле, чем привлечь одного пользователя

ОПИСАНИЕ КЕЙСА



- › Удерживать всех пользователей дорого
- › Нужно адресное удержание
- › Удержание пользователей происходит не мгновенно

- › Определение оттока
- › Тип модели
- › Горизонт прогнозирования
- › Методика оценки качества модели
- › Дизайн эксперимента
- › Требования к модели

- › Отток — разрыв договора подключения к сервису
- › Модель — бинарная классификации
- › Горизонт прогнозирования — 2 недели
- › Методика оценки — метрика AUC
- › Дизайн эксперимента — А/Б тестирование на 10% сегменте случайных пользователей
- › Требования к модели: вероятностная модель

ОПРЕДЕЛЕНИЕ ОТТОКА



- › Разрыв договора
- › Отсутствие платных транзакций более 10/90 дней
- › Отсутствие на сервисе более 14/28 дней

КАКАЯ ДОЛЯ ПОЛЬЗОВАТЕЛЕЙ ПОПАДАЕТ ПОД ЭТО ОПРЕДЕЛЕНИЕ?

- › Разрыв договора — 0,1%
- › Отсутствие платных транзакций > 10 дней — 22%
- › Отсутствие платных транзакций > 90 дней — 5%
- › Отсутствие в сети более 14 дней — 16%
- › Отсутствие в сети более 28 дней — 9%

КАКАЯ ДОЛЯ ПОЛЬЗОВАТЕЛЕЙ ПОСЛЕ ВОЗВРАЩАЕТСЯ?



- › Разрыв договора — 0%
- › Отсутствие платных транзакций > 10 дней — 80%
- › Отсутствие платных транзакций > 90 дней — 12%
- › Отсутствие в сети более 14 дней — 90%
- › Отсутствие в сети более 28 дней — 78%

- › Как быстро мы можем связаться с пользователем?
- › Какие методики удержания мы используем?
- › Сколько времени занимает процесс удержания?

- › Какие данные доступны?
- › За какой исторический период?
- › Как объединять данные ?
- › Есть ли в данных сигнал?
- › Как данные следует обработать?
- › Как рассчитать признаки на основе данных?

- › Ключевые характеристики пользователей?
- › Различаются ли пользователи из групп отток/не отток?
- › Можем ли мы решать задачу для всех сегментов?
- › Одинаково ли важно решать задачу для всех сегментов?

- › Кластеризация пользователей
- › Анализ получившихся сегментов
- › Оценка аудиторных показателей
- › «Ручной» анализ случайных пользователей из разных сегментов

- › Убедиться в обоснованности задачи с точки зрения бизнеса
- › Формализовать постановку задачи
- › Провалидировать полученную постановку

ПРОГНОЗИРОВАНИЕ ОТТОКА ПОЛЬЗОВАТЕЛЕЙ

- » Данные разного типа
 - ▶ числовые
 - ▶ номинальные
 - ▶ порядковые
- » Временные ряды

НЕСБАЛАНСИРОВАННАЯ ВЫБОРКА

- › Доля целевого класса может быть намного меньше доли нецелевого класса (0.1% vs 99.9%)
- › Несбалансированность выборки может негативно сказаться на качестве модели
- › Важно заметить это в процессе построения модели!

- Задать веса для объектов таким образом, чтобы:
 - ▶ Скомпенсировать количество объектов меньшего класса их важностью
 - ▶ Задать стоимость ошибки классификации разного рода

- Сгенерировать больше объектов меньшего класса:
 - ▶ Дублирование объектов
 - ▶ Генерация новых объектов путем изменения некоторых признаков существующих объектов
 - ▶ Генерация новых объектов на основе нескольких существующих объектов

- Исключить из обучения объекты преобладающего класса:
 - ▶ Удаление из выборки случайных объектов преобладающего класса
 - ▶ Удаление из выборки групп схожих объектов из преобладающего класса

- › Обучение на данных, доступных НЕ только за исторический период
- › Контроль обучения на данных из будущего
- › Контроль переобучения

КРОСС-ВАЛИДАЦИЯ



- › По объектам
- › По времени

ПОДБОР ПАРАМЕТРОВ

- › Используем кросс-валидацию
- › Сразу фиксируем hold-out dataset
- › Их может быть несколько
- › Используем для финальной проверки решения

- › Одна целевая метрика
- › Оффлайн метрика совпадает или коррелирует с целевой метрикой
- › Хорошо оценивать модель «скользящим окном» по времени

- На каких группах объектов модель ошибается?
- Является ли инвестиция в дальнейшее улучшение модели экономически оправданной?

- › Какие факторы внесли наибольший вклад в модель?
- › Гипотезы относительно причин оттока пользователей?
- › Какие объекты классифицируются наиболее/наименее уверенно?
- › Какие еще данные могли бы быть полезны?

- › Как меняется качество модели во времени?
- › Как быстро она «протухает»?
- › Сколько времени занимает переобучение модели?
- › Сколько времени требуется на переключение с одной модели на другую?

- › Изменились ли данные?
- › Изменилось ли качество модели?
- › Хорошо оценивать модель с разных сторон с помощью набора метрик

- › Оценивать качество решения задачи на всех этапах
- › Заранее продумать список потенциальных «узких» мест