



ИГОРЬ АШМАНОВ

# БОЛЬШИЕ ПОЛЬЗОВАТЕЛЬСКИЕ ДАННЫЕ И НОВЫЙ МАРКЕТИНГ

ПОДГОТОВЛЕНО  
«АШМАНОВ И ПАРТНЕРЫ»  
МАРТ 17



# БОЛЬШИЕ ПОЛЬЗОВАТЕЛЬСКИЕ ДАННЫЕ – НОВАЯ РЕАЛЬНОСТЬ

- ❖ Накоплены большие объёмы БПД
- ❖ Возникли новейшие технологии обработки
- ❖ Возник рынок, в условиях правового вакуума
- ❖ Происходит массовая утечка БПД за рубеж
- ❖ В США 29 марта 2017 разрешили продажу БПД
- ❖ Есть большие риски и большие возможности



# ЧТО ТАКОЕ «БОЛЬШИЕ ДАННЫЕ»

- Модное, «шумовое» слово
- Не научное, а скорее медийное понятие
- Не новое понятие (ему примерно 30 лет, раньше называлось Data Mining и т.п.)
- Не алгоритмы обработки, а сами данные
- Большие массивы данных со специфическими свойствами



# БОЛЬШИЕ ДАННЫЕ – ЧТО ЭТО?



# «БОЛЬШИЕ ДАННЫЕ» - ОПРЕДЕЛЕНИЕ

- Большие (нельзя обработать вручную)
- **Целое больше суммы частей** (сверху видны закономерности и связи)
- Данные из разных каналов (соцсети, поисковики, камеры, телефоны, счётчики)
- Данные разного формата (текст, связи, картинки, видео, голос)
- С ретроспективой (накопление истории, признаков и маркеров)
- **Данные о людях**

# ЧТО ТАКОЕ БОЛЬШИЕ ПОЛЬЗОВАТЕЛЬСКИЕ ДАННЫЕ

Неформальное определение:

**Большие пользовательские данные** – это данные, собираемые в сети различными сервисами, устройствами и приложениями о поведении пользователей, понимаемых как пользовательские профили.

- ❖ Данные о профилях пользователей
- ❖ Бывают первичные (сырые) и производные (вычисляемые)
- ❖ Не технология, а данные

# КАТЕГОРИИ БПД С ТОЧКИ ЗРЕНИЯ ПОЛЬЗОВАТЕЛЯ

- **Явно отдаваемые.** Адреса почты, посты/комментарии, профили, фото, лайки и т.п. «Регулируются» пользовательскими соглашениями.
- **Ограниченного распространения.** Переписка, закрытые обсуждения, география, биометрия.
- **Косвенные данные.** Поведенческие, география, биометрия, фото, видео, привычки, взгляды, граф связей.
- **Воруемые данные.** Фото, видео, СМС, голос, клавиатурный ввод, пароли, переписка (смартфоны, смарт-ТВ, keyloggers, трояны)



# КТО СОБИРАЕТ БПД

- **Навигаторы:** поисковики и браузеры (запросы, интересы, клики, посещения третьих сайтов, география)
- **Счётчики** разных типов (до 50 на странице, клики, география, интересы, запросы)
- **Рекламные сети** (интересы, клики, география)
- **Коммуникационные сервисы:** социальные сети, публичные почты, мессенджеры (соцдем, адреса, фото, интересы, граф связей)
- **Платёжные системы:** финансовая и экономическая информация
- **Платформы:** ОС, провайдеры, платёжные системы (запросы, видео, аудио, тексты, сообщения, аккаунты, пароли)



# НОВЫЕ СУЩНОСТИ

- **Смартфоны и приложения** (тексты, СМС, снимки, геоположение, контакты, ...)
- **Смарт-ТВ** отправляют поток данных из квартиры (видео, жесты, трёхмерные сцены, звук, потребительское поведение).
- **Носимые устройства, фитнес-браслеты** – изначально «в облаке» (биометрия, медицинское состояние, перемещения)
- **Анализ ДНК** – данные о национальности, родственниках, болезнях
- **Автомобили** (видео, звук, география, поведенческие данные)
- **«Интернет вещей»** - данные об окружении, устройствах, производстве, ремонте



# КАКИЕ ДАННЫЕ СОБИРАЮТСЯ

- **Поведенческие:** поисковые запросы, клики, посещения сайтов, география, перемещения, точки пребывания
- **Социальные:** посты, комменты, соцдем, фото, видео, политические взгляды, работа, привычки, хобби
- **Биологические:** фото, видео, ДНК, отпечатки пальцев, пульс, давление, физическая активность, сигнатура голоса и т.п.
- **Финансово-экономические:** счета, активы, покупки, транзакции, уровень доходов, активы
- **Личные, интимные:** семья, связи, круг общения, привычки, половая ориентация, развлечения, психология, взгляды



# МЕТОДЫ СВЕДЕНИЯ ДАННЫХ И АНАЛИЗА

- **Анализ по всей линейке сервисов:** поиск, новости, рекламная система, почта, социальная сеть, карта, мессенджер, платёжная система, браузер, фотохостинг, счётчик, ...
- **Кросс-канальный анализ:** сравнение векторов с разных источников, правило 4 точек для разных сфер, сведение времён выхода в сеть и постов в соцсетях и т.п.
- **Кросс-платформенный анализ** - сведение воедино разных устройств пользователя (отпечатки браузера, ультразвук и т. п.)



# БПД И ОБЩЕСТВО



# РИСКИ БПД

- **Слежка**, утечка данных граждан РФ к частным компаниям и за рубеж. Большинство платформ - американские
- **Нарушение приватности**, продажа пользовательских данных, навязчивая реклама, раскрытие интимных данных
- **Влияние и пропаганда**, персональное и массовое информационное давление, создание «информационных пузырей»



# ЭТИЧЕСКИЕ И ПРАВОВЫЕ АСПЕКТЫ

- Проблема информированности пользователей
- Проблема безопасности
- Коммерческий интерес собирателей БПД
- Проблема «наблюдающих сервисов»
- Отсутствие этического кодекса и саморегуляции отрасли
- Отсутствие регулирования оборота, правовой вакуум
- Отсутствие отраслевого регулятора и оператора БПД



# КОМУ ДОЛЖНЫ ПРИНАДЛЕЖАТЬ БПД?

- **Пользователю?** Он не знает, кто собирает, что собирают, не может контролировать или запретить сбор и использование
- **Площадкам?** Они по факту владеют БПД, но их пользовательские соглашения незаконны; право использования и оборота надо регулировать.
- **Вывод: это собственность нации,** нужна защита государством.
- **Метафора «цифровых недр».** Владелец площадки не владеет недрами под ней. Для разработки и оборота «цифровой нефти» нужно регулирование.



# ВОЗМОЖНОСТИ БПД

- **Алгоритмы анализа:** промышленная революция, нельзя упустить, чтобы потом не импортировать
- **Знания о нации:** медицина, образование, настроения, выборы, занятость, развлечения, отдых и т.п.
- **Знания об инфраструктуре и производстве:** транспорт, дороги, ЖКХ, электросети, производства.
- **Знания об экономике:** доходы, структура расходов и сбережений, страхование, льготы, планирование развития, прочее
- **Большой рынок:** профиль пользователя стоит денег, при упорядоченном обороте БПД – большой рынок





# БПД: НОВЫЙ МАРКЕТИНГ



# МАРКЕТИНГ И СОЦИОЛОГИЯ В МАСШТАБЕ 1:1

- Знать **каждого** пользователя каждой категории (расширенный профиль пользователя, вычисляемый косвенно)
- **Адресное** сообщение, общение «лицом к лицу»
  - **«Стадо Навального»** – 50-70 тысяч человек, известны поимённо, легко мониторить или послать сообщение каждому
  - **«Группы смерти», экстремистские группы** - видны по лексике, пользователи и модераторы известны поимённо
  - **Карта читателей СМИ:** все, кто советует и обсуждает ссылки на статьи – известны поимённо, размечены категориями и названиями СМИ
  - **Социальная CRM** Сбербанка, МТС и других: личный контакт в аккаунте через 15-20 минут после жалобы в сети



# ПРИМЕР: КАМПАНИЯ ТРАМПА

- Модная тема: новый медийный трюк от Трампа
  - Фабула – беллетристика, формат – вброс в Das Magazin и «отмывка» в СМИ (причём российских)
  - Одна полезная мысль: сегментирование аудитории и адресная доставка; остальное – ненаучная фантастика.
- Уроки кампании Трампа
  - Детальное сегментирование аудитории, адресные сообщения
  - Реальная причина победы – не просто полезная медийная технология, а сообщение, соответствующее запросу на справедливость в США
  - Анагажированные СМИ и пропаганда врагов по площадям – не сработали
  - Сообщение – **необходимое** условие, адресная медийная доставка – **достаточное**



# ЧТО МЫ ЗНАЕМ О ПОЛЬЗОВАТЕЛЕ

## Сообщают сами в профилях:

- ФИО, пол – 99%
- город, страна – 70%
- возраст, день рождения – 30%
- семейное положение – 10%
- образование, место работы, должность – 2%
- интересы - меньше 1-2%

## Можно вычислить по текстам:

- **Соцдем:** семейность, география, возраст, уровень дохода, образование
- **Семантика:** взгляды, хобби, интересы, политическая позиция, религиозность, половая ориентация, прочее

## По социальному графу:

- окружение, схожее поведение, паттерны медиапотребления, лидеры мнений, любимые СМИ и группы



# НОВЫЙ МАРКЕТИНГ: КАК ЭТО ДЕЛАЕТСЯ

# СЕГМЕНТАЦИЯ АУДИТОРИИ

- **Гипотеза:** начальные категории от маркетологов и социологов, параметры (соцдем, доход, география, прочее)
- **Построение модели** категории и сбор начальных списков аудитории:
  - **Ареал обитания:** группы, форумы, блоги, тематика текстов и сайтов
  - **Содержание** общения: посты, комменты, ссылки на СМИ
  - **Социальный граф:** с кем связаны, кого репостят и комментят
- **Расширение списков:**
  - **Социальное замыкание** (френды, комменты, лайки)
  - **Семантические близнецы** (товарные или политические предпочтения)



# ТЕСТИРОВАНИЕ И КОРРЕКЦИЯ МОДЕЛИ

- **Чистка** от ботов, спамеров, биороботов, виртуалов, кибербойцов
- **Тестовые сообщения:**
  - Опросы в соцсетях
  - Таргетированная реклама
  - SMM, посты, комменты, ссылки на СМИ
- **Мониторинг реакции** в комментариях, A/B-тестирование
- **Коррекция** модели и списков категорий



# ВЫЯВЛЕНИЕ АВТОРИТЕТОВ ДЛЯ КАТЕГОРИИ

- Вычисление **лидеров мнений** по социальному графу, лайкам, комментариям
- Вычисление **авторитетных СМИ** – по размещению в соцсетях ссылок на статьи и новости
- Вычисление популярных **мест посещения** (групп, пабликов) – по постам и комментариям, ссылкам





# КАК ДОСТАВИТЬ СООБЩЕНИЕ?



# ДОСТАВКА: МЕТОДЫ

- **Таргетированная реклама в соцсетях:** рекламные сообщения на основе параметров пользователя и списков
  - Таргетированная реклама «по площадям» на основе списков параметров
  - Таргетированная реклама 1:1 на основе списков аккаунтов
- **Программатика:** размещение рекламного сообщения в рекламных сетях на основе параметров категории
- **Нативная реклама** в соцсетях по спискам аккаунтов и авторитетов: SMM, посты, комменты, ссылки на СМИ
- **Нативная реклама в СМИ:** сообщения в СМИ, являющихся авторитетными для данной категории



# ПЛЮСЫ И МИНУСЫ РЕКЛАМНЫХ МЕТОДОВ

## Минусы:

- **Динамическая аудитория:** завтра эти куки и браузеры – другие
- **Динамическое сообщение:** сообщение исчезает, его не остаётся в сети после показа, нет последствия
- **Баннерная слепота:** пользователь не видит сообщения или оно его раздражает, доверие низкое
- **Нельзя проверить доставку:** конверсия в действие 1-2%
- **Дорого:** стоимость доставки 2-5 рублей на пользователя

## Плюсы:

- Быстрое, массовое и автоматическое размещение.



# СОЦИАЛЬНАЯ ДОСТАВКА: + / -

## Плюсы:

- Категории и списки аккаунтов – **постоянные**, это люди
- Сообщение **остаётся в сети**, есть последствие – комментирование, расшаривание, лайки
- **Есть обратная связь**, сообщение видно на 100%, можно померить реакцию (выхлоп, тональность)
- **Дешевле**: можно доставлять не всем, а в авторитетные СМИ, группы, в аккаунты лидеров мнений

## Минусы:

- Нельзя доставлять автоматически
- Дольше, требует ручной работы



# ДОСТАВКА: ОБРАТНАЯ СВЯЗЬ

- Мониторинг реакции в постах и комментариях
- Уточнение модели
- Уточнение списков аккаунтов
- Уточнение списков авторитетов
- Уточнение сообщения
- Повторная доставка



# СПАСИБО!

# ВОПРОСЫ?

ИГОРЬ АШМАНОВ

Управляющий партнер  
«Ашманов и партнеры»

+7 495 269 0630

+7 985 222 5451

[igor@ashmanov.com](mailto:igor@ashmanov.com)