



МАРКЕТИНГОВЫЙ КОНСАЛТИНГ
И АУТСОРСИНГ

BIG DATA в маркетинге.

Просто о непонятном:

задачи, возможности, реальность

Антон Липатов

Руководитель Datasense, МГ «Текарт»

17/04/2015 Минск, Неделя Байнета. Истоки

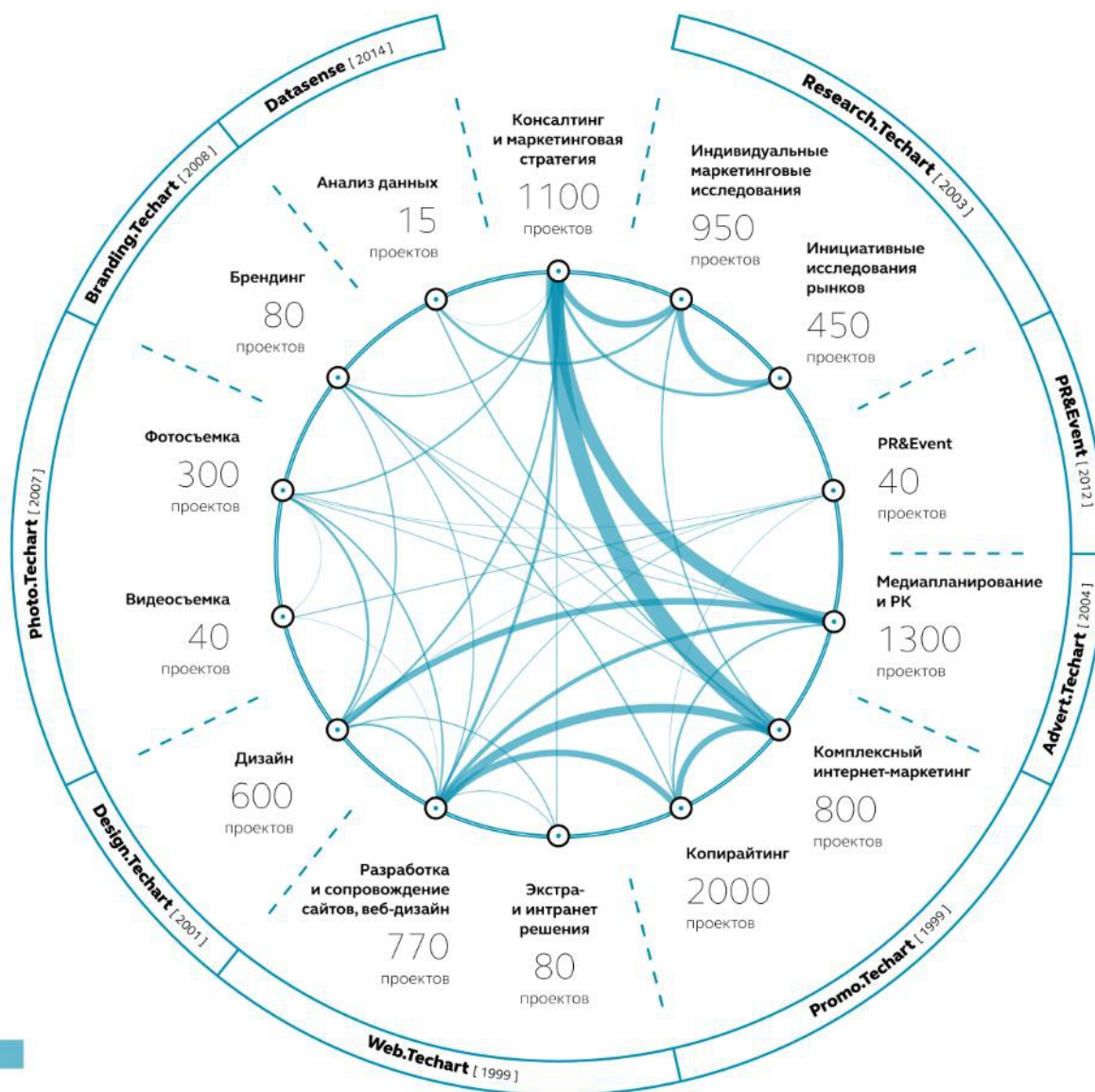
МАРКЕТИНГОВАЯ ГРУППА «ТЕКАРТ»

2/30

1999-2015

*Департаменты, опыт,
и их взаимосвязи*

20 100 200 400 500



DATASENSE

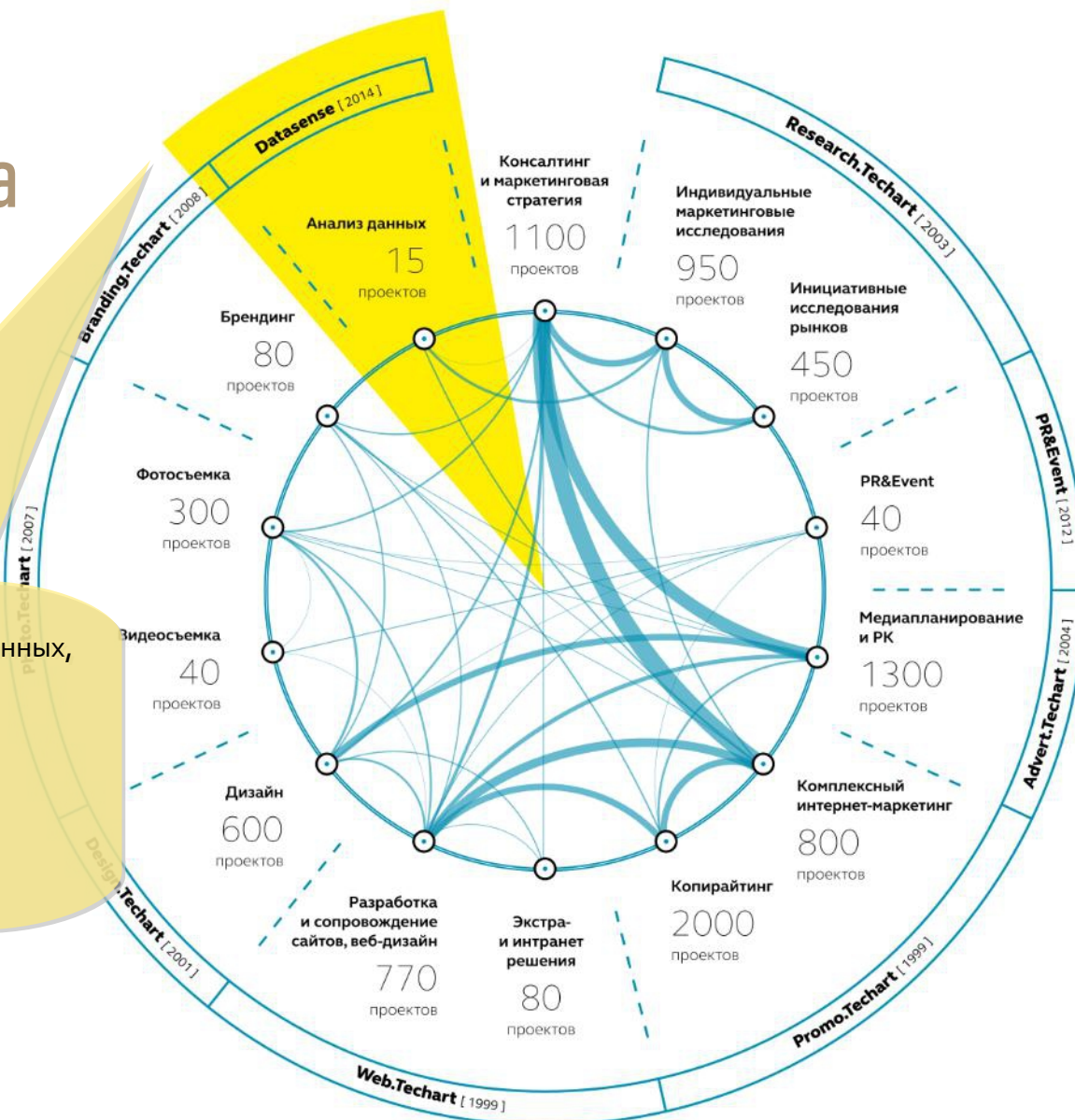
3/30

Дизайн и аналитика данных

Направление, связанное с аналитикой и дизайном данных, Datasense, было открыто в 2014 году.

Задача — собрать, проанализировать и упаковать маркетинговую информацию таким образом, чтобы она стала понятной клиенту.

Дата-майнинг — одно из направлений работы.



Большие данные — это красивая маркетинговая упаковка, которая объединяет, во многом, уже известные методы работы с данными.

Содержание этой упаковки интуитивно непонятно, поэтому вызывает такой интерес, в том числе среди непрофессионалов. В свое время так в России было с нанотехнологиями...

БОЛЬШИЕ ДАННЫЕ

«BIG DATA IS LIKE TEENAGE SEX:
EVERYONE TALKS ABOUT IT, NOBODY REALLY KNOWS
HOW TO DO IT, EVERYONE THINKS EVERYONE ELSE
IS DOING IT, SO EVERYONE CLAIMS THEY ARE DOING IT»

«Большие данные как подростковый секс: все подростки говорят о нем, никто на самом деле не знает, как им заниматься; все думают, что остальные это делают, и потому заявляют, что занимаются тоже»

DAN ARIELY

Автор предыдущей цитаты,
Дэн Ариели. Совершенно никакого
отношения не имеет к большим данным...



DAN ARIELY

Но тоже о них говорит!

... EVERYONE TALKS ABOUT IT ...



A man with short dark hair and blue eyes, wearing a dark suit jacket over a light-colored shirt, is shown from the chest up. He has a thoughtful expression, with his right hand resting against his chin. The background is a dark, textured blue. A yellow rounded rectangular callout box is positioned in the upper left area of the image.

Но о больших данных говорят и профессионалы.

«СМЫСЛ В БОЛЬШИХ ДАННЫХ ЕСТЬ ТОЧНО.
ВОПРОС, КАК ИМЕННО БУДУТ ЗАРАБАТЫВАТЬСЯ ДЕНЬГИ.
ЭТО ОДНА ИЗ ИДЕЙ, ГДЕ МОЖЕТ БЫТЬ
НАШ СЛЕДУЮЩИЙ МИЛЛИАРД»

ОБЪЕМ РЫНКА МИЛЛИАРДЫ ДОЛЛАРОВ?

9/30

	2013	2014	...	2017	2018
IDC		17	...		42
Wikibon	19		...	50	
ABI Research	31		...		114
SNS Research		30	...		56

Оценки и прогнозы рынка BIG DATA, сделанные в 2014-2015 гг., млрд долл.

Оценки и прогнозы рынка огромны. Непонятно, что за ними скрывается, к чему относятся миллиарды долларов?

4V БОЛЬШИХ ДАННЫХ

10/30



ЧТО ВКЛЮЧАЮТ В СЕБЯ МИЛЛИАРДЫ?

ЭТО «ИНВЕСТИЦИИ» В БОЛЬШИЕ ДАННЫЕ

А объем рынка составляют все инвестиции в подобные процессы. В том числе затраты на железо. Такая методология сильно раздувает оценку.

38% ЖЕЛЕЗО

40% УСЛУГИ

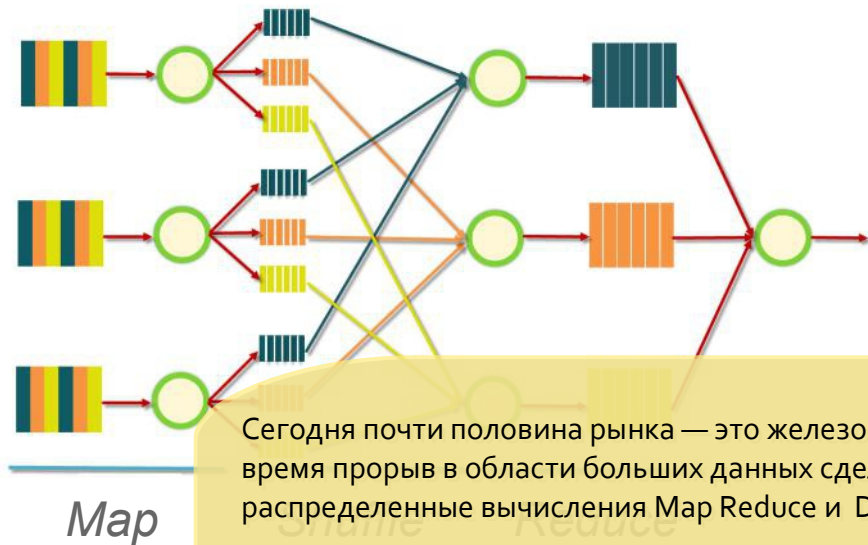
22% СОФТ

! ЗАРПЛАТА

НА САМОМ ДЕЛЕ

12/30

MAP REDUCE

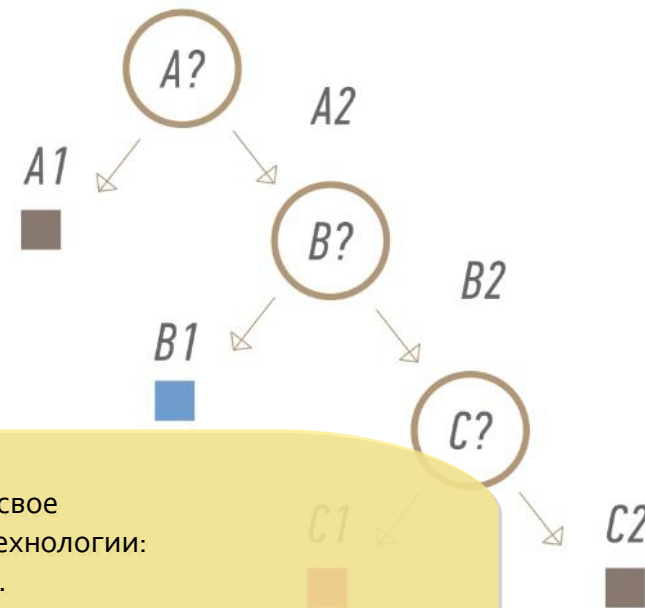


Сегодня почти половина рынка — это железо. Однако в свое время прорыв в области больших данных сделали две технологии: распределенные вычисления Map Reduce и Data Mining.

Data Mining, релевантные софт и услуги, сегодня можно считать фронтендом темы (наиболее часто под большими данными понимают именно поиск неочевидных связей). Он объединил различные методы, позволяющие находить скрытые нетривиальные знания в больших массивах данных.

Бэкенд — необходимая инфраструктура для анализа данных.

ДАТА-МАЙНИНГ



ЭКСПЕРИМЕНТАЛЬНЫЙ ПОДХОД

13/30

Data Mining предполагает статистический, экспериментальный подход. Обычно задача формулируется: есть исторический массив данных, нужно на основании этих данных дать прогноз. Такой подход не новый, но возможность обработки именно больших исторических массивов позволила сделать прорыв в ряде отраслей, например, в машинном переводе.

АНАЛИТИЧЕСКИЙ ПЕРЕВОД

- *морфология*
- *семантика*
- *синтаксис*

Но для Data Mining нужен стабильный поток данных. Если этот поток есть, то получаются отличные результаты.

Мы стоим на пороге новых великих потоков: информация от датчиков, интернет вещей.

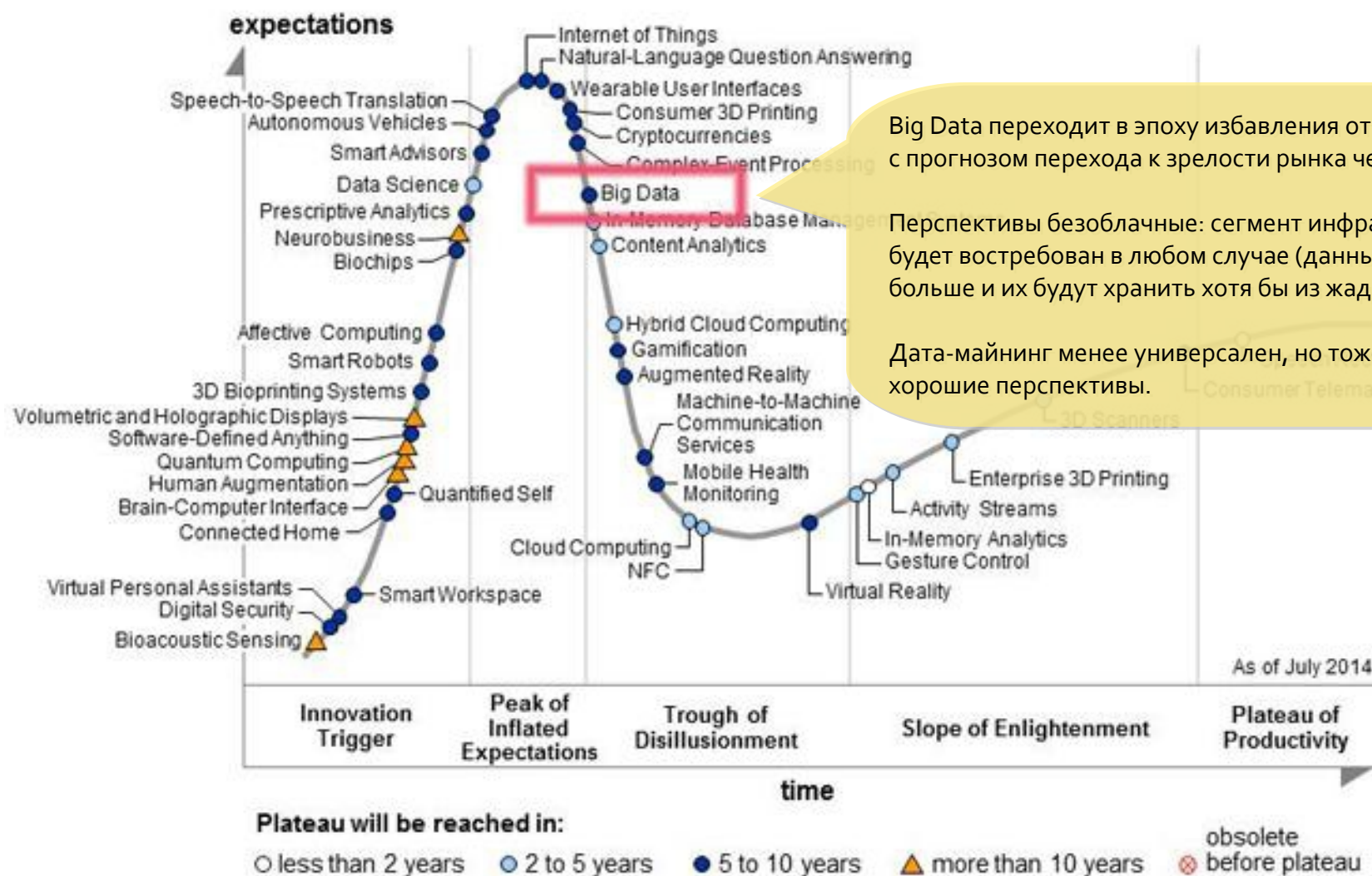
СТАТИСТИЧЕСКИЙ ПЕРЕВОД

Google.Translate
Яндекс.Перевод и др.

А еще: обработка результатов научных исследований, загруженность дорог, выявление фактов мошенничества, медицинских исследованиях и проч.

ЗРЕЛОСТЬ ТЕХНОЛОГИЙ 2014

14/30



Big Data переходит в эпоху избавления от иллюзий с прогнозом перехода к зрелости рынка через 5-10 лет.

Перспективы безоблачные: сегмент инфраструктуры будет востребован в любом случае (данных будет все больше и их будут хранить хотя бы из жадности).

Дата-майнинг менее универсален, но тоже имеет хорошие перспективы.

БОЛЬШИЕ ДАННЫЕ В МАРКЕТИНГЕ

ОЖИДАНИЯ?

В теории, большие данные позволяют отследить каждого пользователя и перейти в коммуникации и анализе на «уровень пользователя» (user-level)

БОЛЬШИЕ ДАННЫЕ —
БОЛЬШИЕ ВОЗМОЖНОСТИ?

БОЛЬШИЕ ДАННЫЕ В МАРКЕТИНГЕ

ПЕРЕЙТИ НА УРОВЕНЬ КОНКРЕТНОГО КЛИЕНТА
И КОНКРЕТНОЙ ТРАНЗАКЦИИ

ГОВОРИТЬ С КЛИЕНТОМ ТЕТ-А-ТЕТ, ПОНИМАТЬ
ЕГО ПРОБЛЕМЫ И ЖЕЛАНИЯ

ПОВЫШАТЬ ЭФФЕКТИВНОСТЬ МАРКЕТИНГА
И СНИЖАТЬ ЗАТРАТЫ

БОЛЬШИЕ ДАННЫЕ В МАРКЕТИНГЕ

РЕАЛЬНОСТЬ

БОЛЬШИЕ ДАННЫЕ —
БОЛЬШИЕ ОГРАНИЧЕНИЯ

Данные в маркетинге имеют ряд особенностей, которые накладывают особые ограничения, в первую очередь, на дата-майнинг.

User-level — не панацея.

ДА! USER-LEVEL ДАННЫМ

18/30

ПЕРСОНАЛИЗАЦИЯ ВЕБ-САЙТА

АВТОМАТИЗАЦИЯ EMAIL-РАССЫЛОК

USER-LEVEL РЕКЛАМА (В ТОМ ЧИСЛЕ RTB)

НЕТ?! USER-LEVEL ДАННЫМ

19/30

SEO

User-level, т.е. знание особенностей каждого пользователя, не имеет смысла в традиционных методах интернет-маркетинга: SEO, классическом контексте. Здесь в лучшем случае возможно мыслить и работать с группами и сегментами целевой аудитории.

Еще меньше персонализация применима к оффлайновой рекламе.

КОНТЕКСТНАЯ РЕКЛАМА

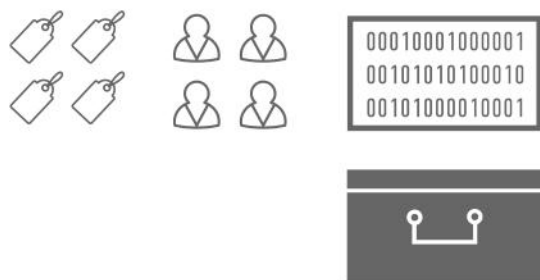
ОФФЛАЙНОВАЯ РЕКЛАМА

01 ПРОБЛЕМА ЗНАНИЙ. ПОЧЕМУ?

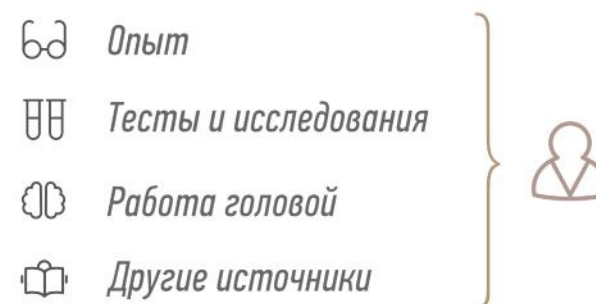
20/30

КАКУЮ СКИДКУ УСТАНОВИТЬ?

Дата-майнинг



Традиционный анализ



??? ← ПОЧЕМУ? → Потому что...

02 ПРОБЛЕМА ДАННЫХ

21/30

ПРОЦЕСС ДАТА-МАЙНИНГА (УПРОЩЕННО):

1. Подготовка данных
2. Вычисления
3. Интерпретация результатов

Ограничение 2. Проблема качества данных.

Подготовка данных, как правило, обязательный этап перед вычислениями. Он повышает затраты на анализ.

02 ПРОБЛЕМА ДАННЫХ. ШУМЫ

22/30

НАПРИМЕР, ДАННЫЕ В ИНТЕРНЕТ:

- *роботы, парсеры*
- *множественные одинаковые события*
- *несовершенство систем веб-аналитики*

Ограничение 2. Проблема качества данных.

Данные подвержены различным шумам.

Если анализ ведется на пользовательском уровне влияние шумов на качество анализа усиливается.

02 ПРОБЛЕМА ДАННЫХ

ПОЛНОТА

23/30

ВКЛЮЧЕНИЕ В АНАЛИЗ МНОГИХ
ФАКТОРОВ **ТРЕБУЕТ ДАННЫХ** ПО
КАЖДОМУ ИЗ ФАКТОРОВ

Ограничение 2. Проблема качества данных.

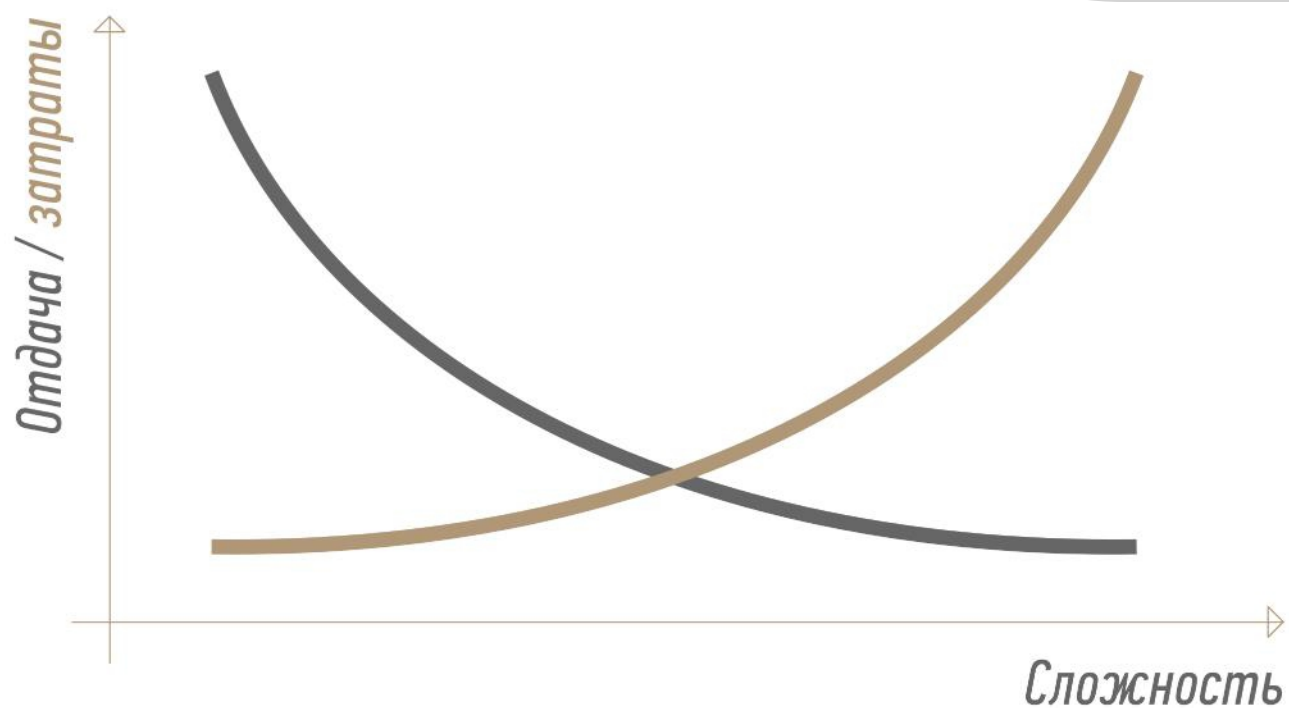
Преимущество data-mining в том, что в анализе могут быть задействованы много факторов. Однако по каждому фактору должны быть качественные данные.

В современном маркетинге данные обрывочны и часто не позволяют полностью реализовать мощь анализа.

03 ПРОБЛЕМА ЭФФЕКТА

Ограничение 3. Соотношение затраты/эффект

С ростом сложности анализа отдача от анализа стремительно снижается при росте затрат.





2009

NETFLIX

DATE: 09-21-09

PAY TO THE ORDER OF: BellKor's Pragmatic Chaos

AMOUNT: ONE MILLION \$1,000,000 ⁰⁰/₁₀₀

FOR: The Netflix Prize

Reed Hastings

03 ПРОБЛЕМА ЭФФЕКТА

26/30

Ограничение 3. Соотношение затраты/эффект

NetFlix Prize — колоссальные трудозатраты.

NETFLIX PRIZE (2006–2009 гг.)

20 000 команд зарегистрировались в первый год

2 000 часов потратила команда-лидер в первый год

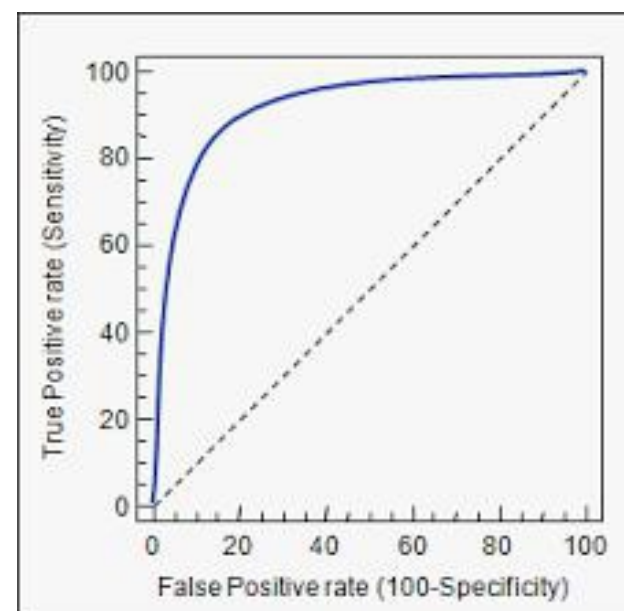
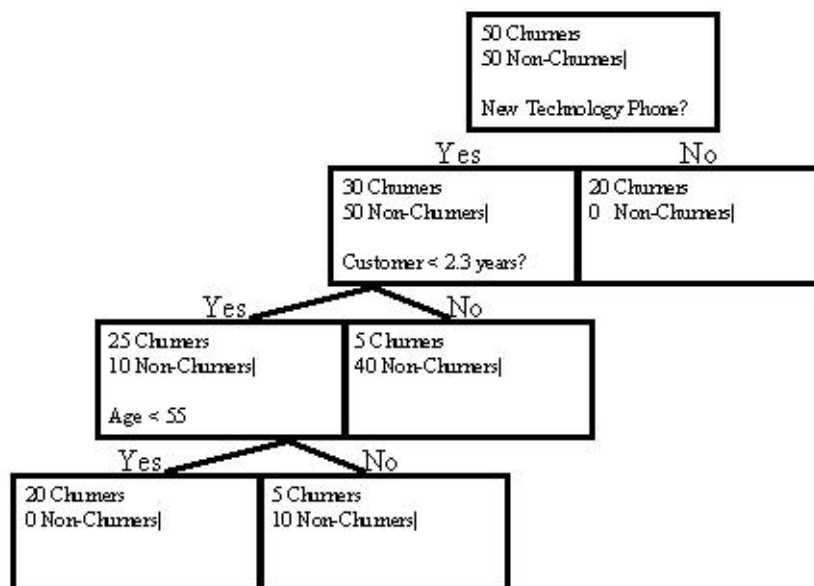
3 года продлился конкурс

Суммарные трудозатраты?

04

ПРОБЛЕМА ИНТЕРПРЕТАЦИИ
И ВИЗУАЛИЗАЦИИ

27/30



ОГРАНИЧЕНИЯ «БОЛЬШИХ ДАННЫХ»

28/30

1. ПРОБЛЕМА ЗНАНИЙ

2. ПРОБЛЕМА КАЧЕСТВА ДАННЫХ

3. ВОПРОС ЭФФЕКТИВНОСТИ

4. ВИЗУАЛИЗАЦИЯ И ИНТЕРПРЕТАЦИЯ

Маркетинг — не отрасль-локомотив для практического приложения статистических методов и больших данных.

В маркетинге ограничены стабильные качественные потоки данных. Есть более дешевые методы повышения эффективности.

Современный маркетинг нельзя отождествлять только с маркетингом на уровне пользователей: для этого недостаточно качественные данные мы собираем.



Эмоция.

Сегодня на подъеме data-driven подход, ориентированный на постоянное повышение эффективности маркетинга за счет использования данных (CRM, персонализация, big data...)

Это здорово, но не стоит забывать, что за маркетингом, кроме данных, алгоритмов и задач повышения эффективности, стоят еще идеи и психология, понимание клиентов.

Возможно, именно по этой причине, статистические методы в маркетинге не имеют такого эффекта, как в «холодных» областях, вроде финансов, страхования и машинного перевода.

АЛГОРИТМ ВМЕСТО ИДЕИ?

Вопросы?

Антон Липатов
lipatov@techart.ru

+7 (495) 790-75-91, доб. 125
www.facebook.com/anton.y.lipatov

