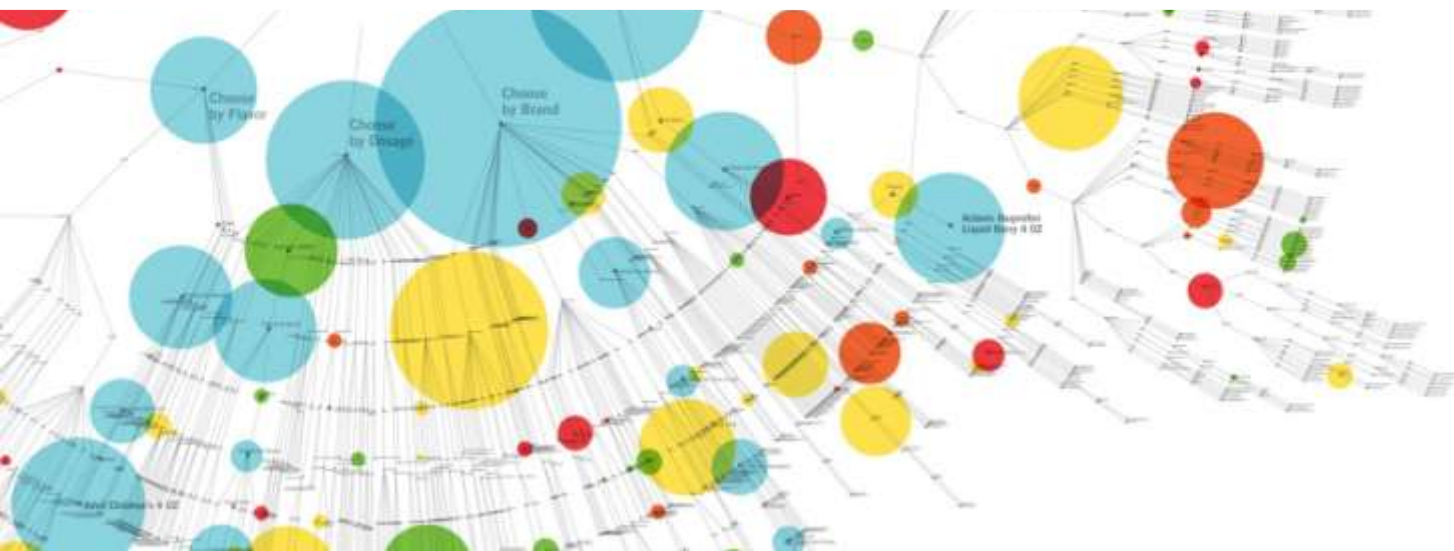


Данные как ценнейший актив организации:
как заставить их работать?



Вахмянин Иван
CEO Visiology
ivan@visiology.su



В ближайшие пять лет все компании на рынке разделятся на победителей и побежденных в зависимости от качества их аналитики.

Джинни Рометти, CEO IBM



Data is the new oil. It's valuable, but if unrefined it cannot really be used. It has to be changed into gas, plastic, chemicals, etc to create a valuable entity that drives profitable activity; so must data be broken down, analyzed for it to have value.

Клайв Хамби, британский математик и основатель компании Dunnhumby



О компании Polymedia

Polymedia — лидер в области решений для визуализации информации.

Направления деятельности:

- Системная интеграция, арендные решения, дистрибуция.
- Разработка и реализация комплексных решений для диспетчерских залов, ситуационных центров, центров управления, конференц-залов, пресс-центров, входных групп, систем распределенных информационных дисплеев, учебных аудиторий и классов.
- Интеграция и внедрение программно-аппаратных решений в области анализа и визуализации данных (BI, Big Data) на базе информационно-аналитической платформы Visiology
- Центр Разработки - разработка специализированного оборудования программного обеспечения;
- Центр Технической Поддержки Пользователей и Сервисный Центр.



Интеграционные проекты Polymedia



Ситуационные центры



Диспетчерские



Конференц-залы



Переговорные комнаты



Мультимедийные классы



Залы заседаний





Visiology – российская ИТ компания. Мы разрабатываем программное обеспечение для анализа и визуализации данных, а также осуществляем внедрение своих разработок в коммерческих компаниях и государственных организациях.



Polywall

Управление видеостенами



Flipbox Software Suite

ПО для интерактивных
дисплеев



**Аналитическая платформа
Visiology**

Part of  group

Поиск...

Подробная аналитика

- Стратегическое управление
- Основная деятельность
- Обеспечивающая деятельность
- Финансы
- Пользовательская аналитика

Справочная информация

\$ 65.89 -1.72%

€ 73.46 -2.10%

¥ 10.04 -1.82%

21°

759мм

52%

2м.с.

Индекс ММВБ 10:40:01 1898.75

10:00

14:10

18:20

Индекс РТС

10:40:00

907.97

0.96%

Широкий рынок

10:39:46

1343.19

0.01%

Голубые фишки

10:40:00

12463.09

0.03%

Индекс FOM

10:39:46

35.14

-2.7%

Избранные показатели



Иванов Иван

ПРОСРОЧЕННЫЕ ПОРУЧЕНИЯ

Задание	Отв.лицо	Дата
Провести анализ рынка	Егоров С.П.	10.03
Заклучить договор с ООО "ТехИндастриал" и выдать аванс	Носов Ю.А.	14.03
Обновить данные по анализу эффективности персонала	Коновалов С.Р.	17.03

[подробнее...](#)



СМИ

GAZETA.RU	
RG.RU	
WWW.NG.RU	
NOVAYAGAZETA.RU	
RIA.RU	

[подробнее...](#)

Структура системы

Разнородные источники информации



Учетные системы



Статистические данные



Анализ СМИ



ГИС



Системы мониторинга



Экспертиза, аналитика

Аналитическая платформа Visiology



Аналитические отчеты и прогнозы

для стратегического планирования и оперативного управления

Комплексная визуализация

оптимизированная под различные дисплейные решения



Видеостена



Ноутбук



Интерактивный дисплей



Мобильные устройства

Что такое **данные** для предприятия?



Терабайты жестких дисков в серверных шкафах?



Миллионы записей и документов в учетных системах и базах данных?



Вся информация, обрабатываемая сотрудниками предприятия?

Данные – это актив предприятия!

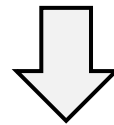
Материальный актив



Стоимость
(приобретения и владения)

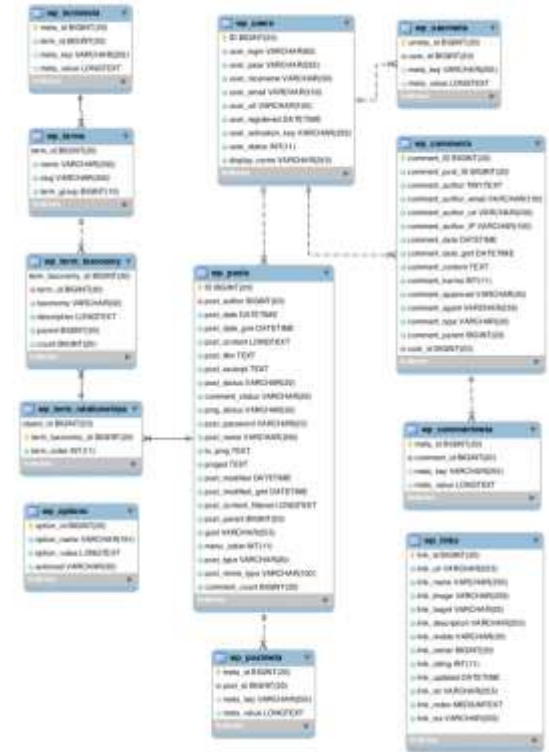
Качество

Жизненный цикл



Приносит прибыль
или убыток...

Данные



Извлечение прибыли из данных

Управленческая аналитика

- Аналитические системы позволяют руководителю понять, в каком состоянии находится организация в целом, какие направления прибыльные, какие нет, в каких подразделениях ситуация ухудшается, а в каких - наоборот. Эта информация позволяет принимать управленческие решения тогда, когда это необходимо, а не тогда, когда уже случился кризис.

Операционная аналитика

- В эту группу попадают все аналитические активности, связанные с процессами и проектами основной и обеспечивающей деятельности. Соответственно, цель такой аналитики – оптимизация соответствующих процессов, снижение издержек, улучшение качества и результата.

Прямая монетизация данных

- Например, интернет-компании, продающие информацию о трафике клиентов банкам, что те в свою очередь могли сделать им более таргетированные предложения. Другой пример, агрегаторы данных из различных источников (например, данные по юрлицам, госзакупкам или судебным делам) продающие доступ к своим данным, как сервис.



Уровни зрелости организации в области управления данными

Начальный

- Не оценивается качество данных
- Отчетность строится вручную
- Много «скрытых» знаний

Развивающийся

- Внедрены системы бизнес-аналитики
- Менеджмент использует анализ данных как вспомогательный инструмент для принятия решений
- На уровне отделов внедряются решения для обеспечения качества данных

Зрелый

- Выделено подразделение, отвечающее за управление данными
- Качество данных и процессов управления ими имеет измеримые метрики и непрерывно улучшается
- Решения для управления качеством данных и метаданных внедрены и являются стратегическими для компании
- Менеджмент принимает data-driven решения
- Проводятся исследования в области применения продвинутой аналитики к бизнесу компании

Продвинутый

- Методы продвинутой аналитики и машинного обучения широко применяются во всех сферах деятельности компании
- Все сотрудники компании обеспечены удобными инструментами аналитики и опираются в своей работе на данные



Big Data

Большие данные

Predictive Analytics

Предсказательная аналитика

Business Intelligence

~~Бизнес-интеллигенция~~

Бизнес-аналитика

Machine Learning

Машинное обучение



Descriptive Analytics

Описательная аналитика

➡ Что произошло?

Predictive Analytics

Предсказательная аналитика

➡ Что произойдет?

Prescriptive Analytics

Предписывающая аналитика

➡ Что нужно делать?



Информация о прошлом опыте



Predictive Analytics

Предсказательная аналитика



Прогнозирование будущего



Пример. Target (американский «Ашан»)

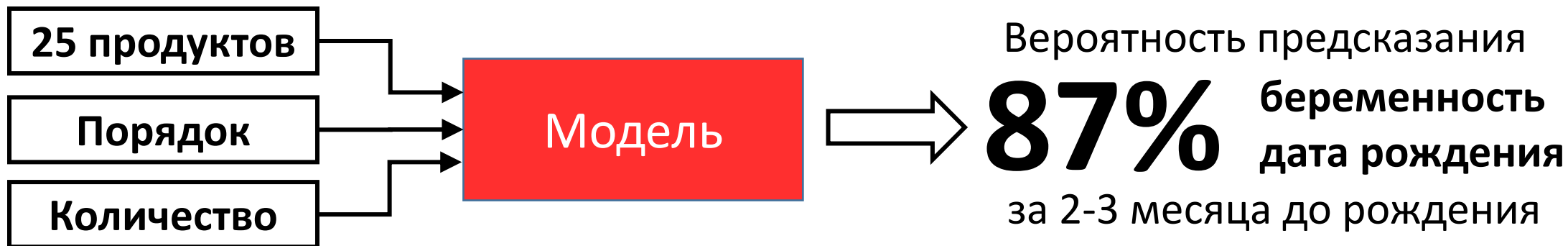


1.



2. Идентификация момента рождения ребенка у покупательниц

3.



4. Отправка спецпредложений



VISIOLOGY

$$\sigma^2 \left(\sum_{i=1}^n a_i X_i \right)$$

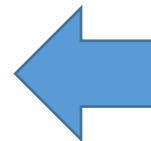
Математические методы

В основном – теория вероятностей
и математическая статистика



Predictive Analytics

Предсказательная аналитика



Данные

Количество данных
должно быть
значительным



Компьютерные технологии

Вычислительная мощность и
современные алгоритмы



Что такое тервер и матстат?

Теория вероятности

Условия эксперимента

↓ рассчитываем

Вероятностная модель

↓ предсказываем

Результаты эксперимента

Матстатистика

Условия эксперимента

↑ выявляем

Вероятностная модель

↑ подбираем

Результаты эксперимента

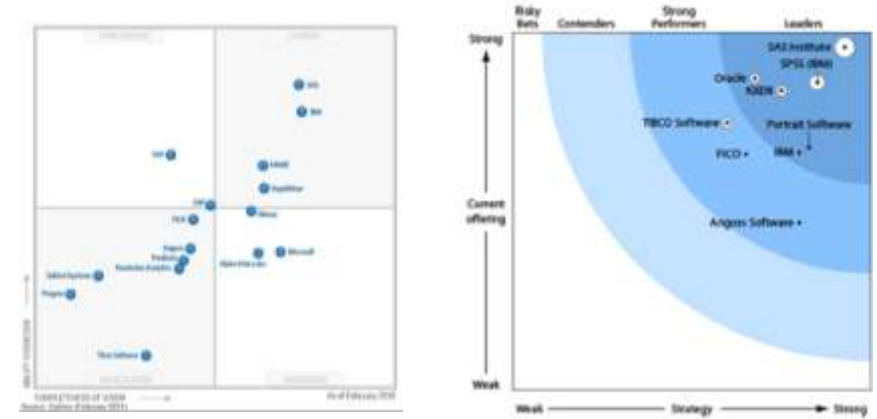
Как это выглядит...

$$\begin{bmatrix} \sum x_i^4 & \sum x_i^3 & \sum x_i^2 \\ \sum x_i^3 & \sum x_i^2 & \sum x_i \\ \sum x_i^2 & \sum x_i & n \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \sum x_i^2 y_i \\ \sum x_i y_i \\ \sum y_i \end{bmatrix}$$

$$\begin{aligned} \hat{\beta} &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n x_i y_i - \frac{1}{n} \sum_{i=1}^n x_i \sum_{j=1}^n y_j}{\sum_{i=1}^n (x_i^2) - \frac{1}{n} (\sum_{i=1}^n x_i)^2} \\ &= \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{\text{Cov}[x, y]}{\text{Var}[x]} = r_{xy} \frac{s_y}{s_x}, \\ \hat{\alpha} &= \bar{y} - \hat{\beta} \bar{x}, \end{aligned}$$

Что изменилось за последние 15 лет?

Gartner и Forrester создали отдельные направления по Advanced Analytics



Объем рынка только ПО и услуг
в области предсказательной аналитики

\$2B

= объем рынка Enterprise Videoconferencing



Data Scientist: The Sexiest Job of the 21st Century



VISIOLOGY

$$\sigma^2 \left(\sum_{i=1}^n a_i X_i \right)$$

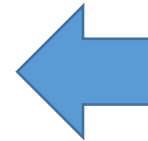
Математические методы

В основном – теория вероятностей
и математическая статистика



Predictive Analytics

Предсказательная аналитика



Big Data

Большие данные



Machine Learning

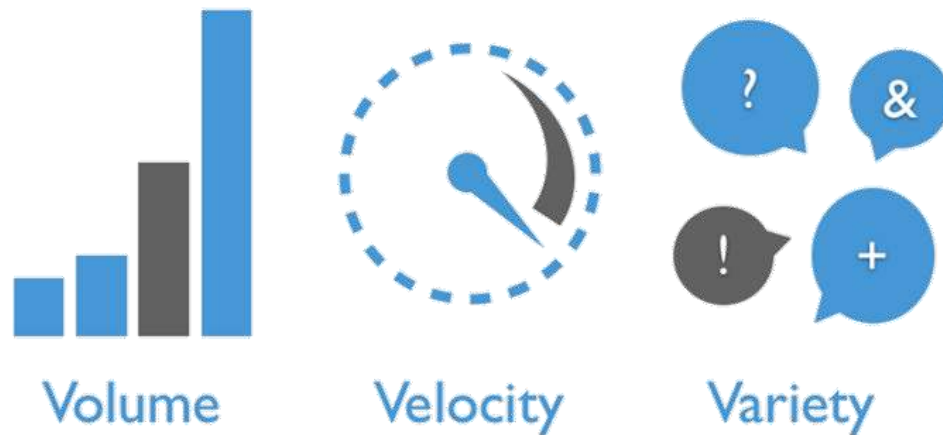
Машинное обучение



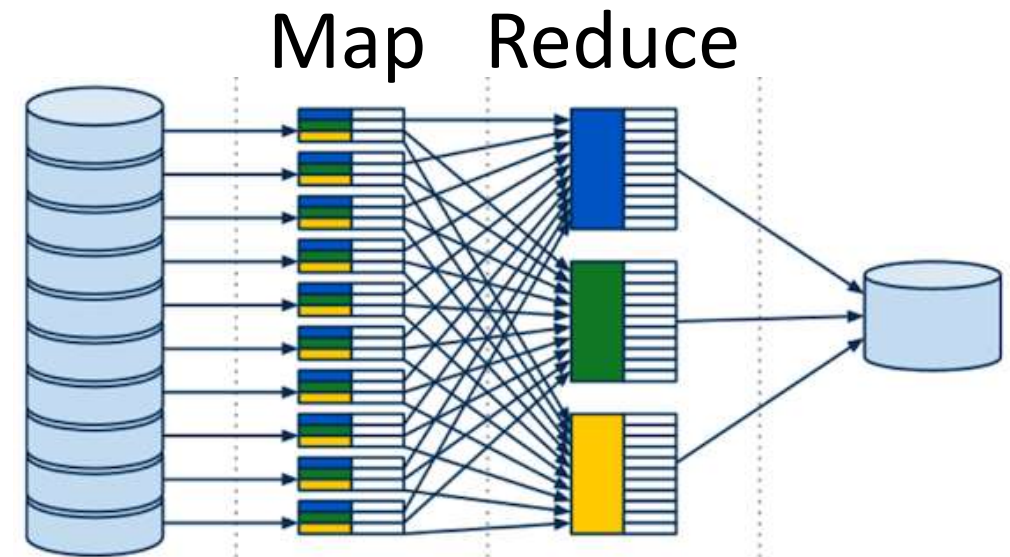
Big Data

Большие данные

- ЭТО



Данные, которые слишком сложны
для традиционных средств
аналитики



Инструменты для анализа данных,
распределенных по кластерам
серверов



Machine Learning - ЭТО

Машинное обучение

**компьютерные методы и алгоритмы, позволяющие применить
разделы математической статистики, теории вероятностей,
численных методов оптимизации дискретного анализа
для выделения знаний из данных**

Метрические методы классификации и регрессии

Логические методы классификации

Критерии выбора моделей и методы отбора признаков

Градиентные методы обучения

Метод опорных векторов

Многомерная линейная регрессия

Нелинейная регрессия

Прогнозирование временных рядов

Байесовская теория классификации

Логистическая регрессия. Разделение смеси распределений

Кластеризация

Нейронные сети

Линейные композиции, бустинг

Эвристические, стохастические, нелинейные композиции

Ранжирование

Поиск ассоциативных правил

Задачи с частичным обучением

Коллаборативная фильтрация

Тематическое моделирование

Обучение с подкреплением



Как повысить зрелость управления данными?

Центр компетенций по управлению данным

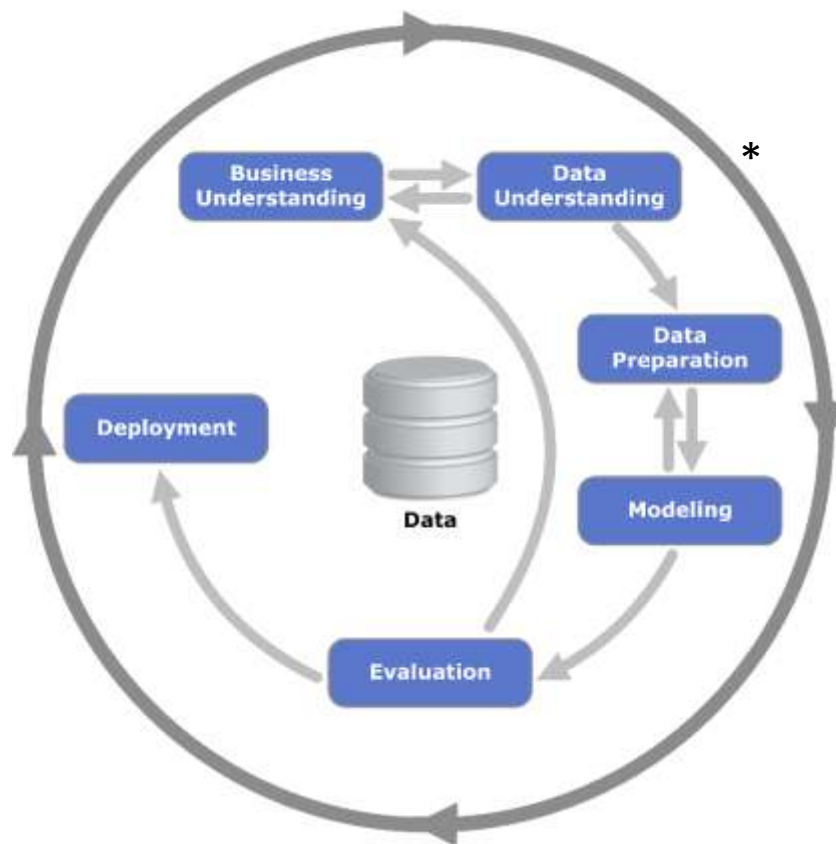


CDO



Data Management Office

Привязка к бизнес-целям



Вовлечение топ-менеджмента



* - Методология CRISP-DM





www.polymedia.ru



VISIOLOGY

www.visiology.ru

Вахмянин Иван
ivan@visiology.ru



VISIOLOGY