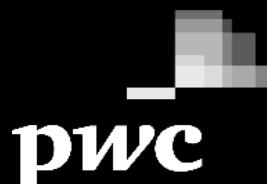


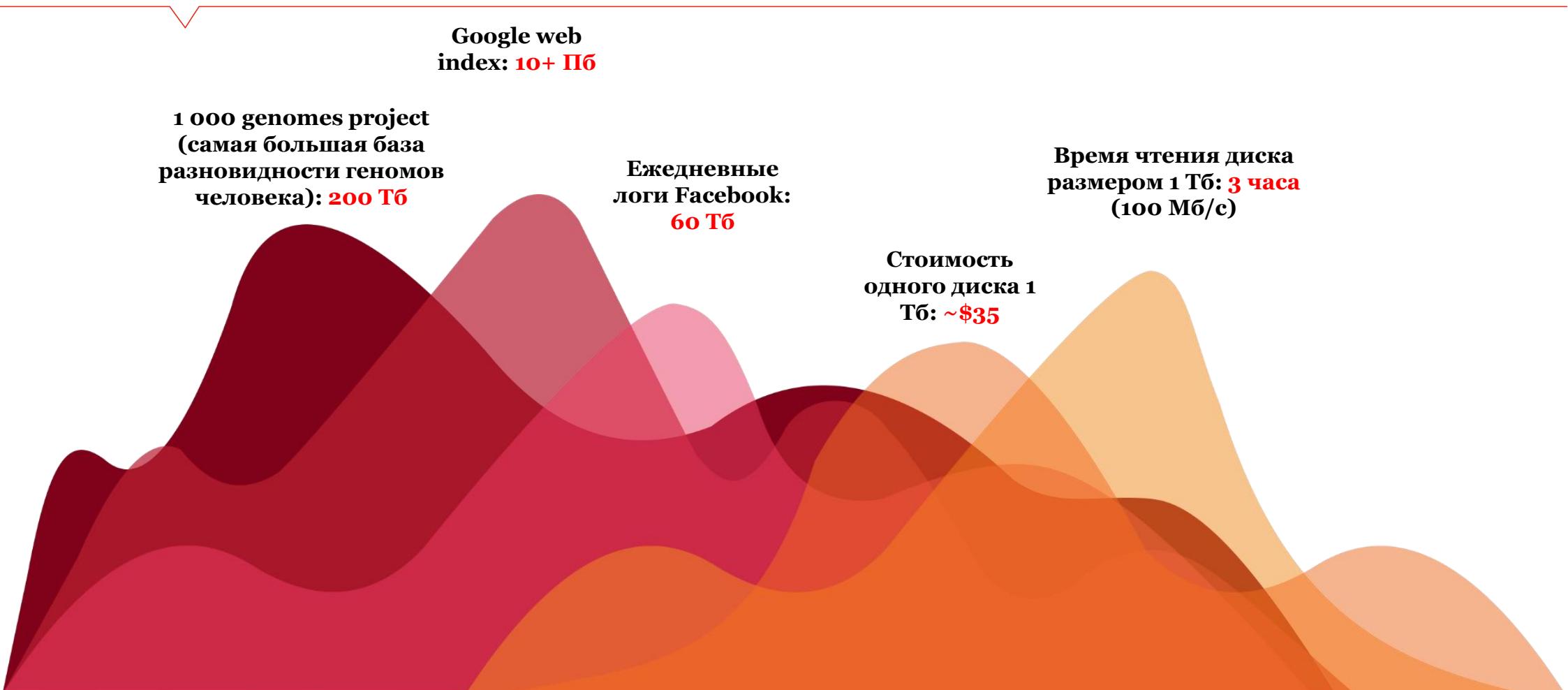
# Big Data & Machine Learning

What's that. Why and what for?



Январь, 2017

# Данных много и их объем растет с каждым днем..



# Технологии Больших данных и машинного обучения вокруг нас...

---



...помогают нам бороться со спамом..

---



..определяют наши предпочтения и вкусы..

The screenshot shows a 'Recommended for You' pop-up window from Amazon.com. At the top left is the Amazon logo, and at the top right is the text 'Recommended for You'. Below this, a message reads: 'Amazon.com has new recommendations for you based on items you purchased or viewed recently, and items you own.' A 'Close' button is located in the top right corner of the main content area.

**Other Movies You Might Enjoy**

Movie Title	Rating	Action Buttons
Amelie	★★★★★	Add   Not Interested
Y Tu Mama Tambien	★★★★★	Add   Not Interested
Eiken	★★★★★	Add
Guys & Dolls	★★★★★	Add   Not Interested
Mostly Martha	★★★★★	Add   Not Interested
Only Human	★★★★★	Add   Not Interested
Russian Dolls	★★★★★	Add   Not Interested

**Eiken has been added to your Queue at position 2.**

This movie is available now.  
[Move To Top Of My Queue](#)

[Continue Browsing](#)   [Visit your Queue >](#)

**Recommended for You**

Item Title	Image	Action Buttons
LOOK INSIDE! <b>BIG</b> by TOM PETERS		
Fascinate: Your 7 Triggers to Persuasion and Captivation		
Sherlock Holmes [Blu-ray]		
Alice in Wonderland [Blu-ray]		

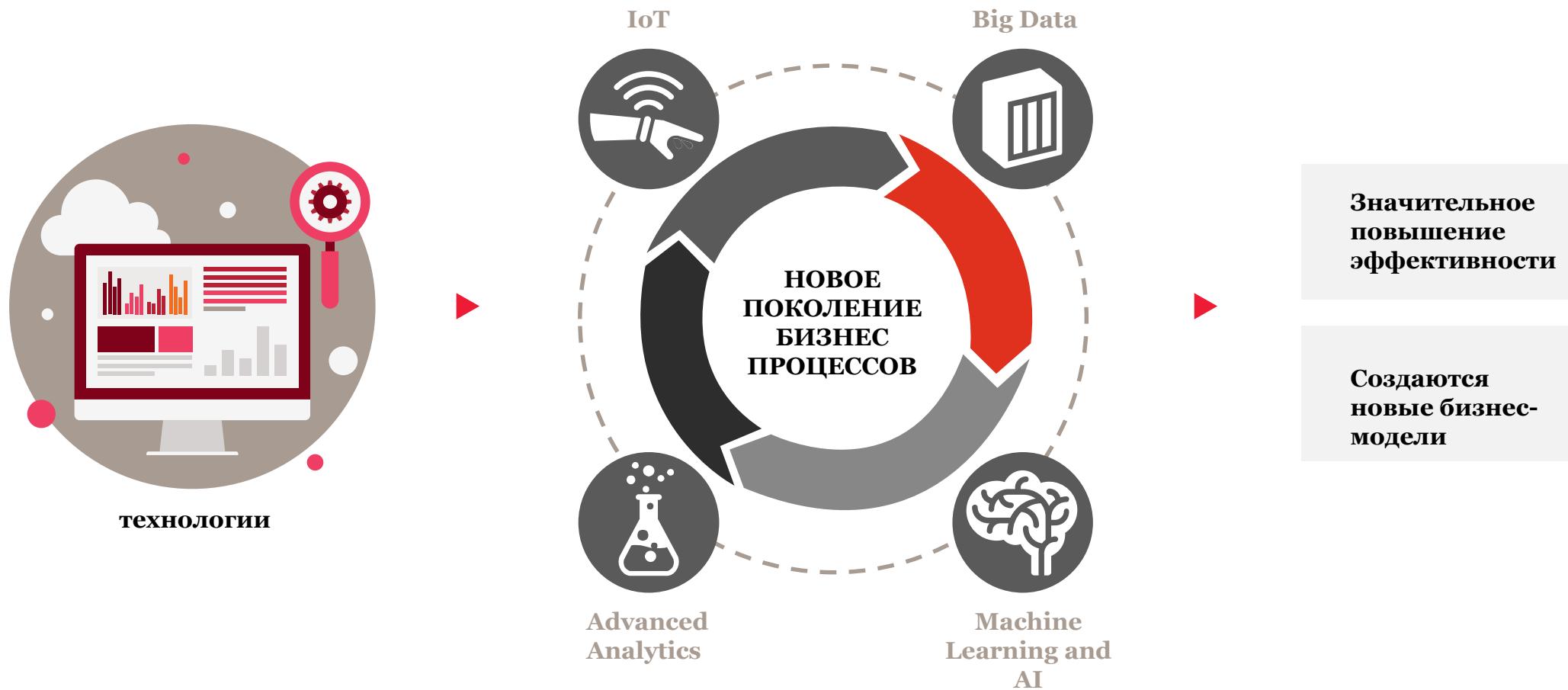
Слышат нас и помогают нам в повседневной активности..

---



# Существующие технологии меняют бизнес.

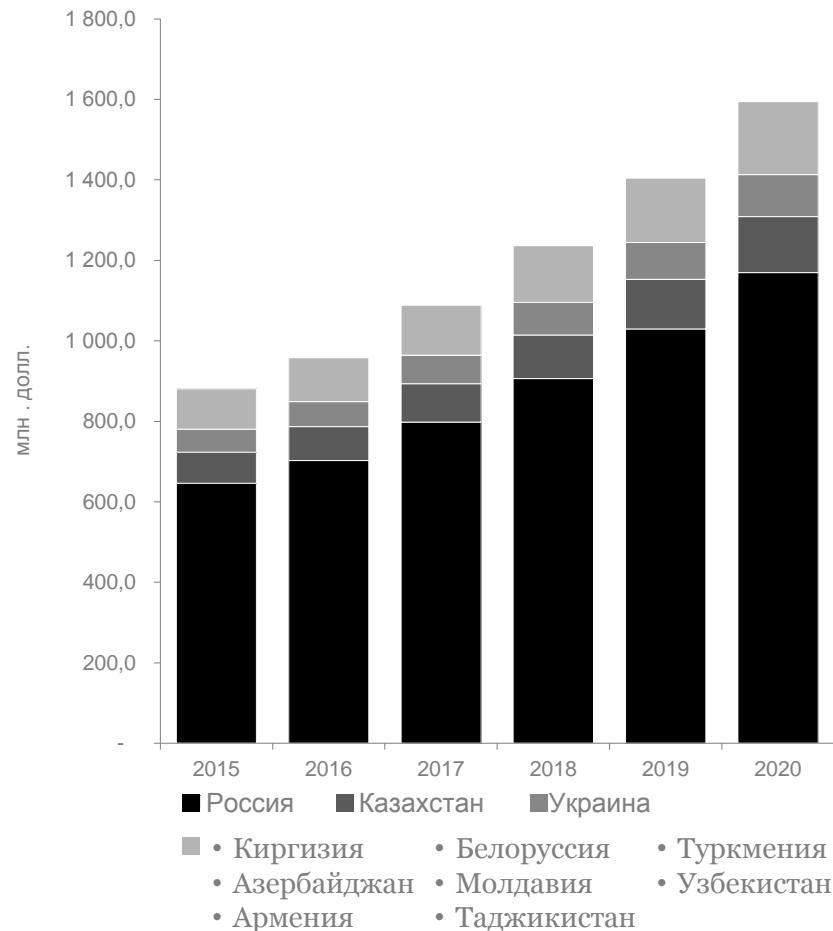
Технологии анализа Больших данных, Машинное обучение и  
Интернет вещей.



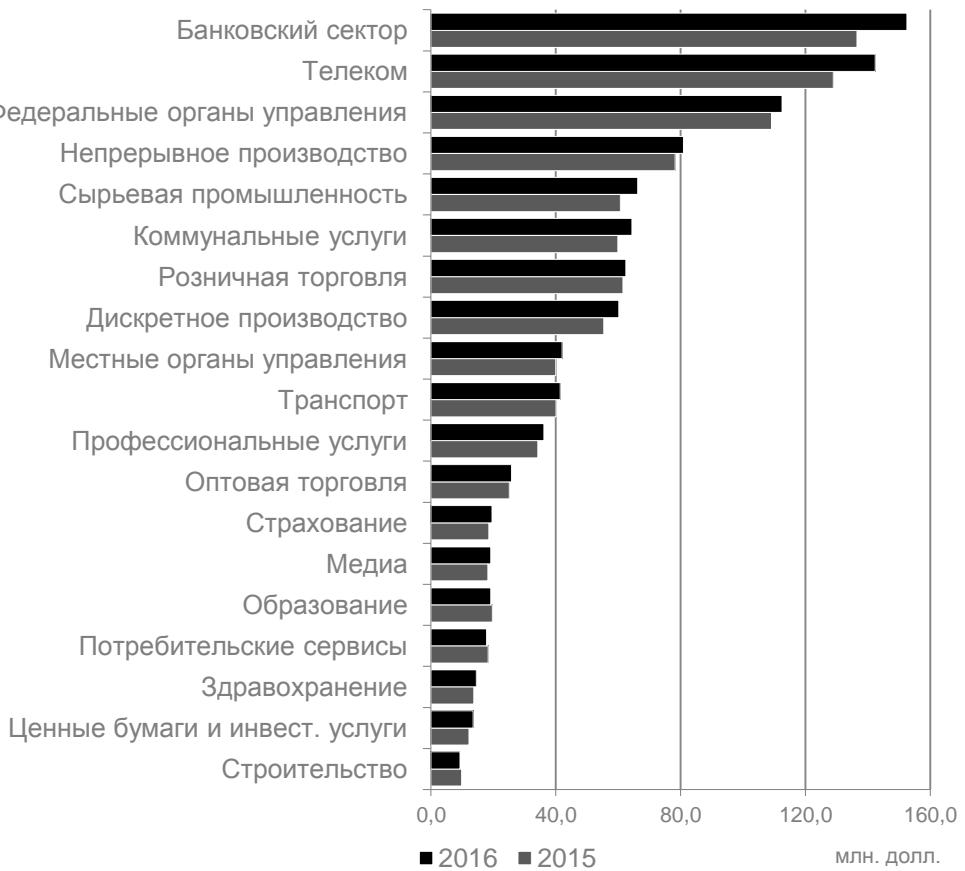
# Big Data в России и СНГ – это растущий рынок

**Область Big Data и машинного обучения только начала развиваться, но уже сейчас прогнозируется двузначный рост данной индустрии в ближайшие 5 лет**

Оценка рынка Big Data в СНГ, млн. долл.



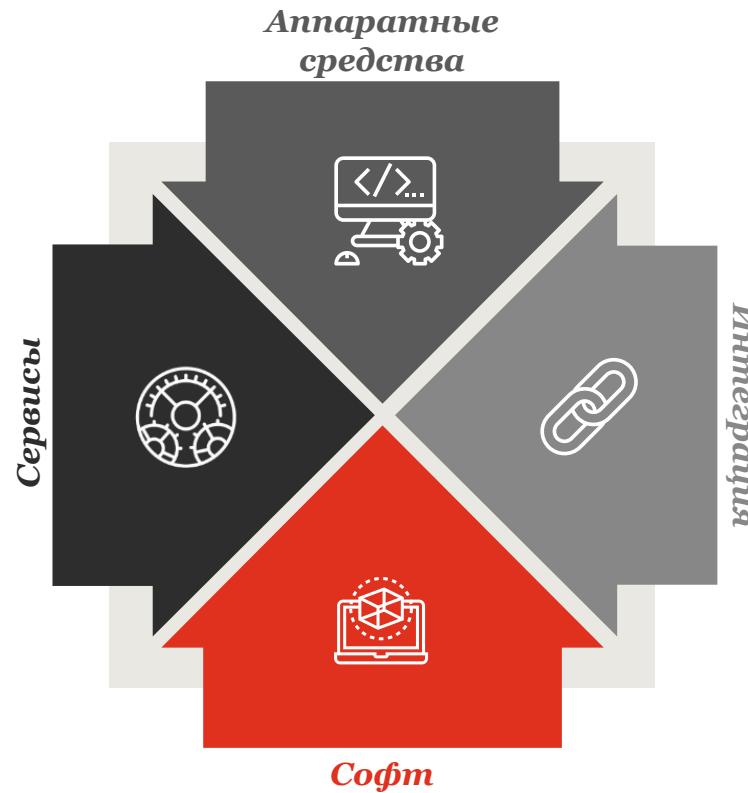
Области внедрения и применения решений в России, млн. долл.



# Глобальный рынок IoT растет двузначными темпами

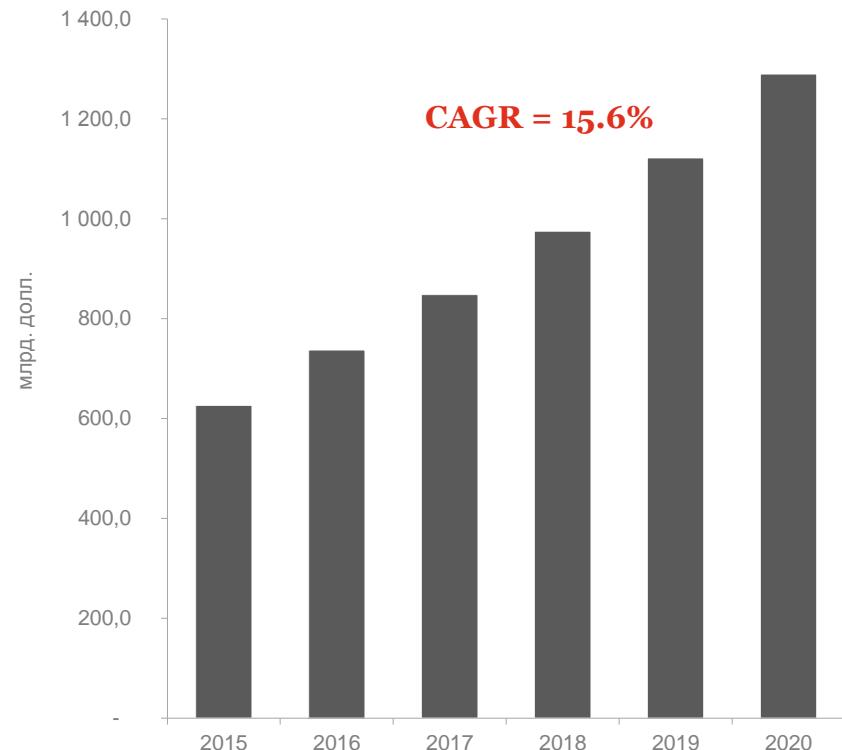


## ИНТЕРНЕТ ВЕЩЕЙ



Средний ежегодный рост  
рынка IoT **15.6%**

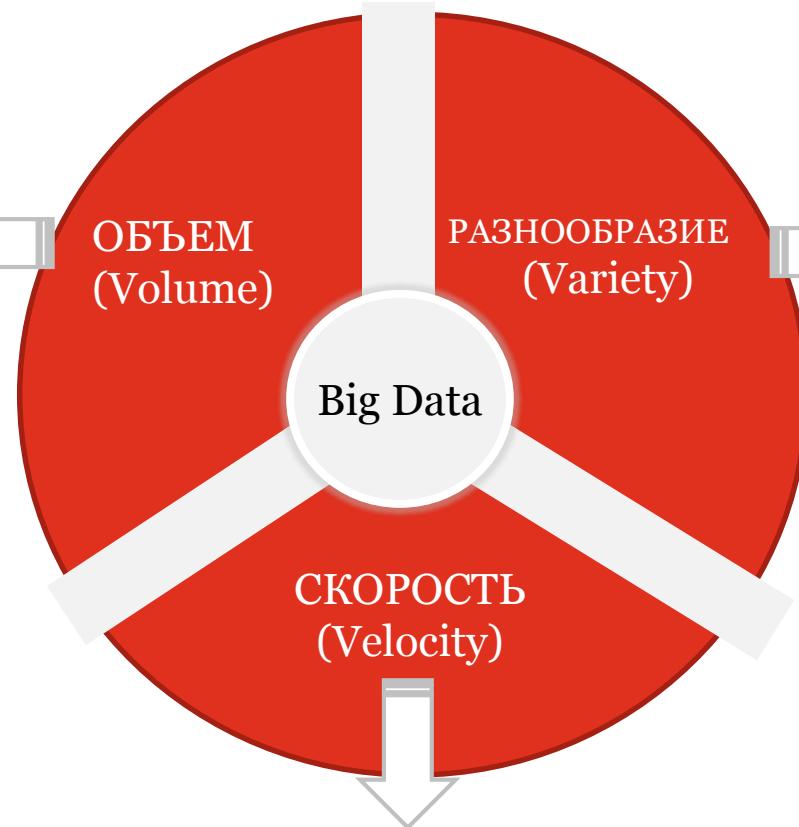
Оценка рынка IoT в мире, млрд. долл.



Источник: IDC

# Что такое Большие данные?

- Обработка информации огромных объемов



- Работа со структурированными и неструктурированными данными
- Оценка достоверности и релевантности информации

- Работа с потоковыми данными в режиме реального времени

*Big Data - серия подходов, инструментов и методов обработки структурированных и неструктурированных данных больших объемов и значительного многообразия для получения человеко-читаемых результатов, эффективных в условиях непрерывного прироста.*

# Технологии анализа больших данных применяются ведущими компаниями для решения сложных и комплексных задач



**Big Data – это инструмент, который может применяться для решения широкого класса задач:**



Повышение эффективности и мониторинг процессов



Предсказание событий, финансовых и нефинансовых показателей



Поиск и анализ коренных причин и драйверов возникновения событий и трендов



**Необходимость использования новых технологий возникает у компаний по следующим причинам:**



Проблемы бизнеса или технологические проблемы невозможно решить классическими консалтинговыми или аналитическими методами



Специалисты компаний понимают, что они обладают всеми необходимыми данными для решения таких задач



Конкуренты применяют передовые методы и за счет этого обходят остальных участников рынка

Доступность и мощность аппаратного обеспечения  
позволяет реализовывать самые амбициозные проекты  
уже сегодня.

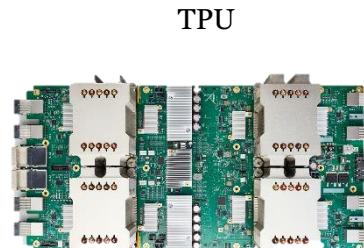
**СЕРВЕРА ДЛЯ  
ХРАНЕНИЯ И  
ОБРАБОТКИ ДАННЫХ**



**ВИДЕОКАРТЫ ДЛЯ  
РАБОТЫ С НЕЙРОННЫМИ  
СЕТЬЯМИ**



GPU



TPU

**ОБЛАЧНЫЕ  
РЕШЕНИЯ**



*арендуемые облака:*



Microsoft  
Azure



amazon  
web services

# Что такое Интернет Вещей (IoT)?

**IoT – концепция пространства, в котором все из аналогового и цифрового миров может быть совмещено.**  
**«Вещи» оцифровываются и становятся частью цифрового мира**

## ГЕНЕРАЦИЯ ДАННЫХ



состояние



данные о внешней среде



параметры оборудования



позиционирование



данные об использовании

## ВОЗМОЖНОСТЬ ПЕРЕДАЧИ (ПРИЕМА) ДАННЫХ



LAN WAN



штрихкод



RFID NFC



ZigBee



Bluetooth



WiFi WiMax



2G 3G 4G 5G



спутниковая связь

## МОДЕЛИ ВЗАИМОДЕЙСТВИЯ



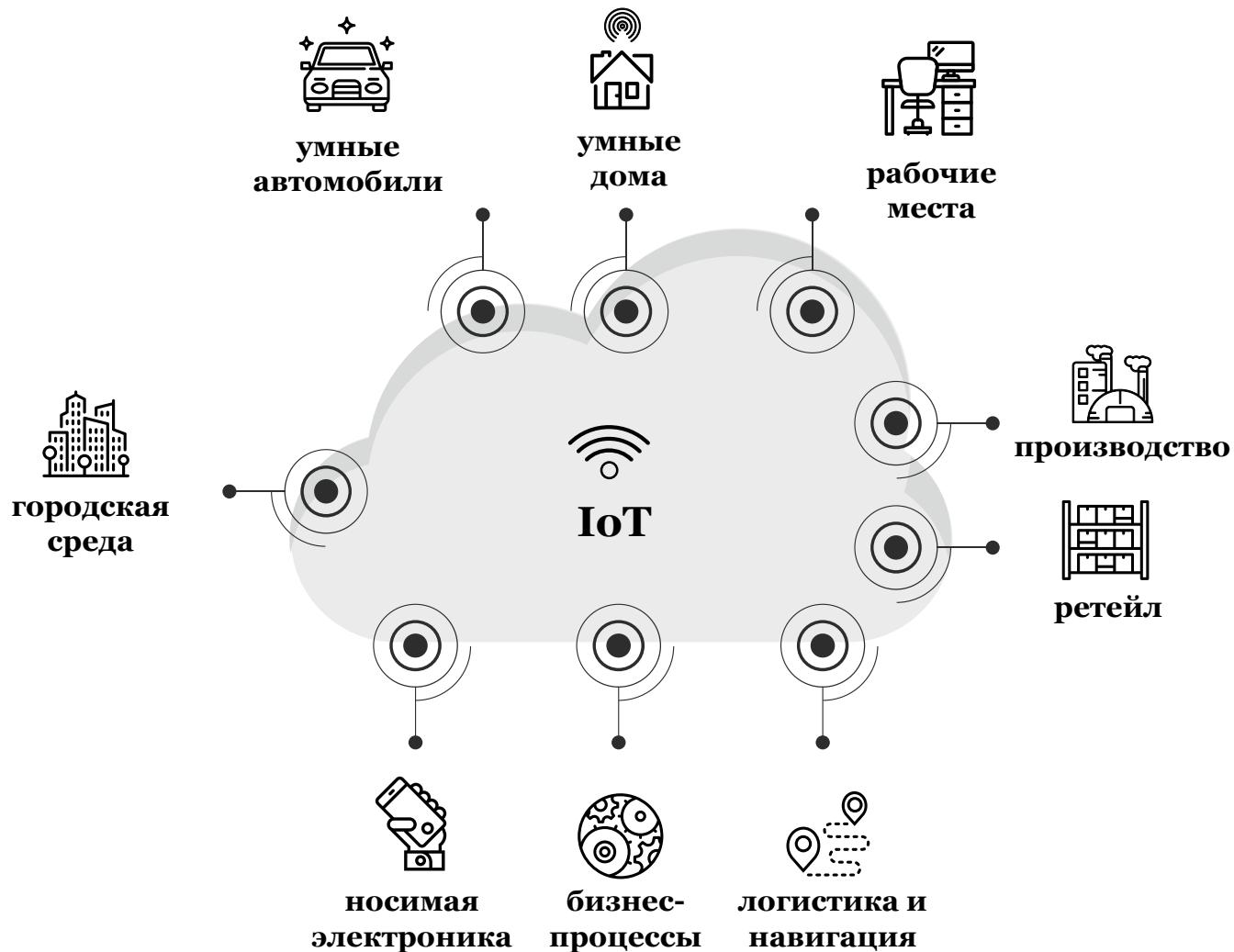
Human machine interface



Machine-to-Machine



# Где применяется технологии анализа больших данных и машинного обучения ?

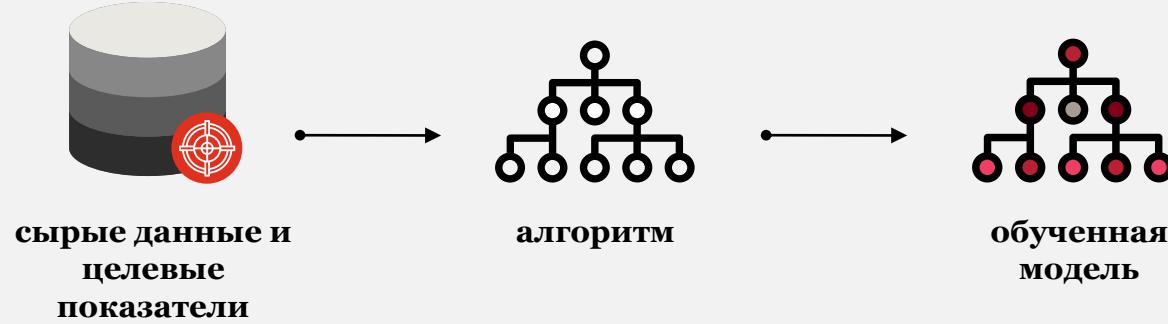


# Что такое Машинное обучение?

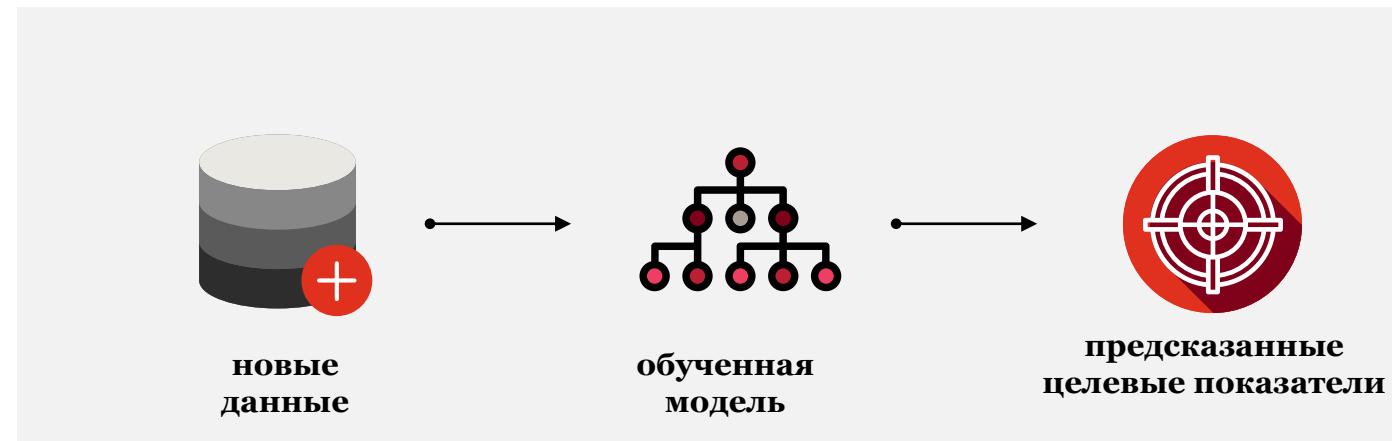
**Машинное обучение** – класс методов, характерной чертой которых является не прямое решение задачи, а обучение в процессе применения решению множества сходных задач



ОБУЧЕНИЕ

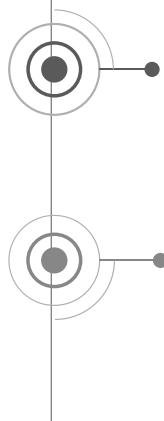


ПРЕДСКАЗАНИЕ



# Пример использования технологий машинного обучения для мониторинга и контроля состояния

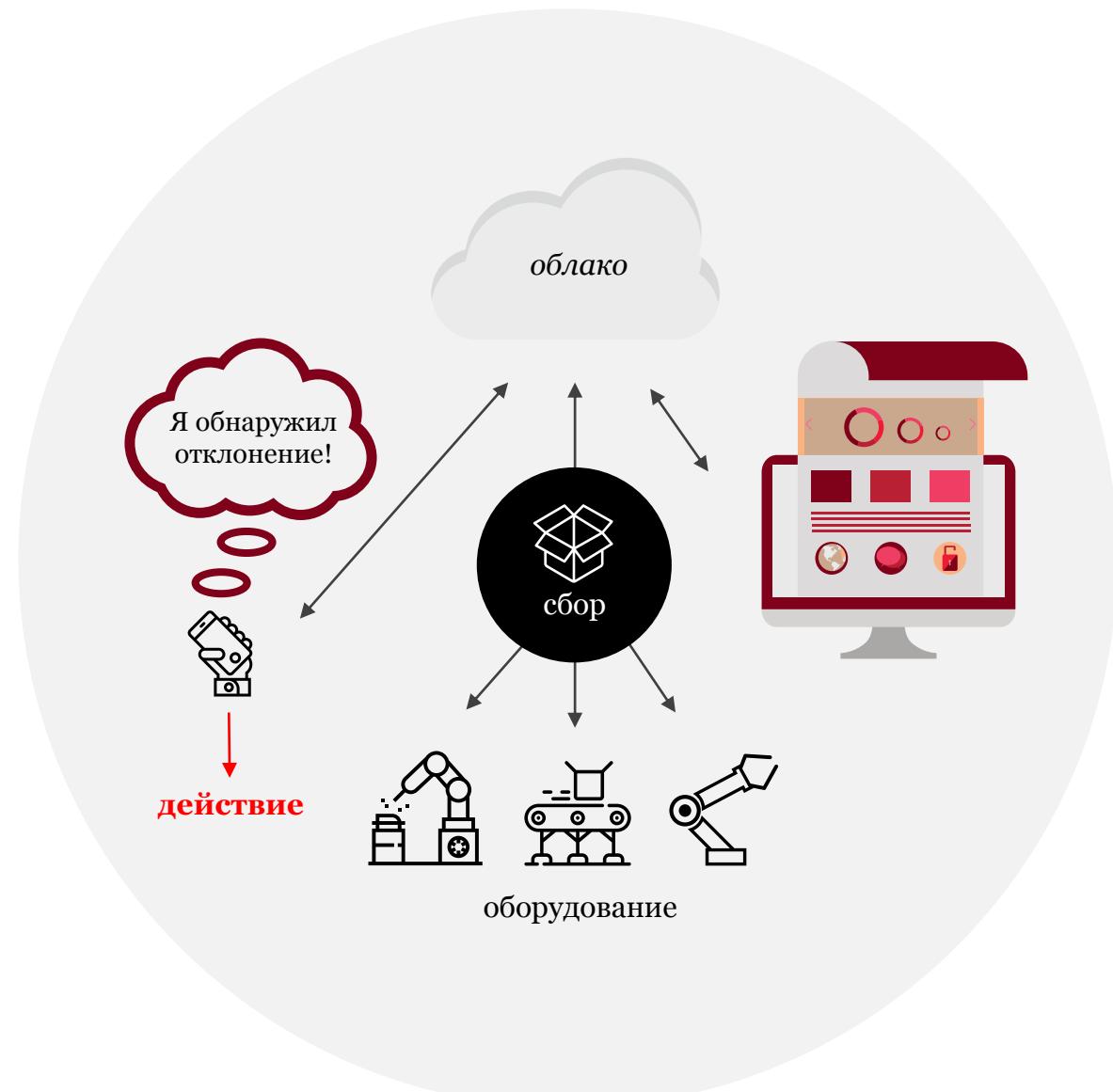
## ЖКХ: водоснабжение



- Необходимо снижение капитальных и операционных затрат
- Использование алгоритмов машинного обучения и данных с датчиков сети



- **снижение на 10% количества прорывов трубопроводов и на 2% снижение объема утечек**
- **восстановление до 50% недостающей информации об отдельных элементах 7000 км сети трубопроводов**

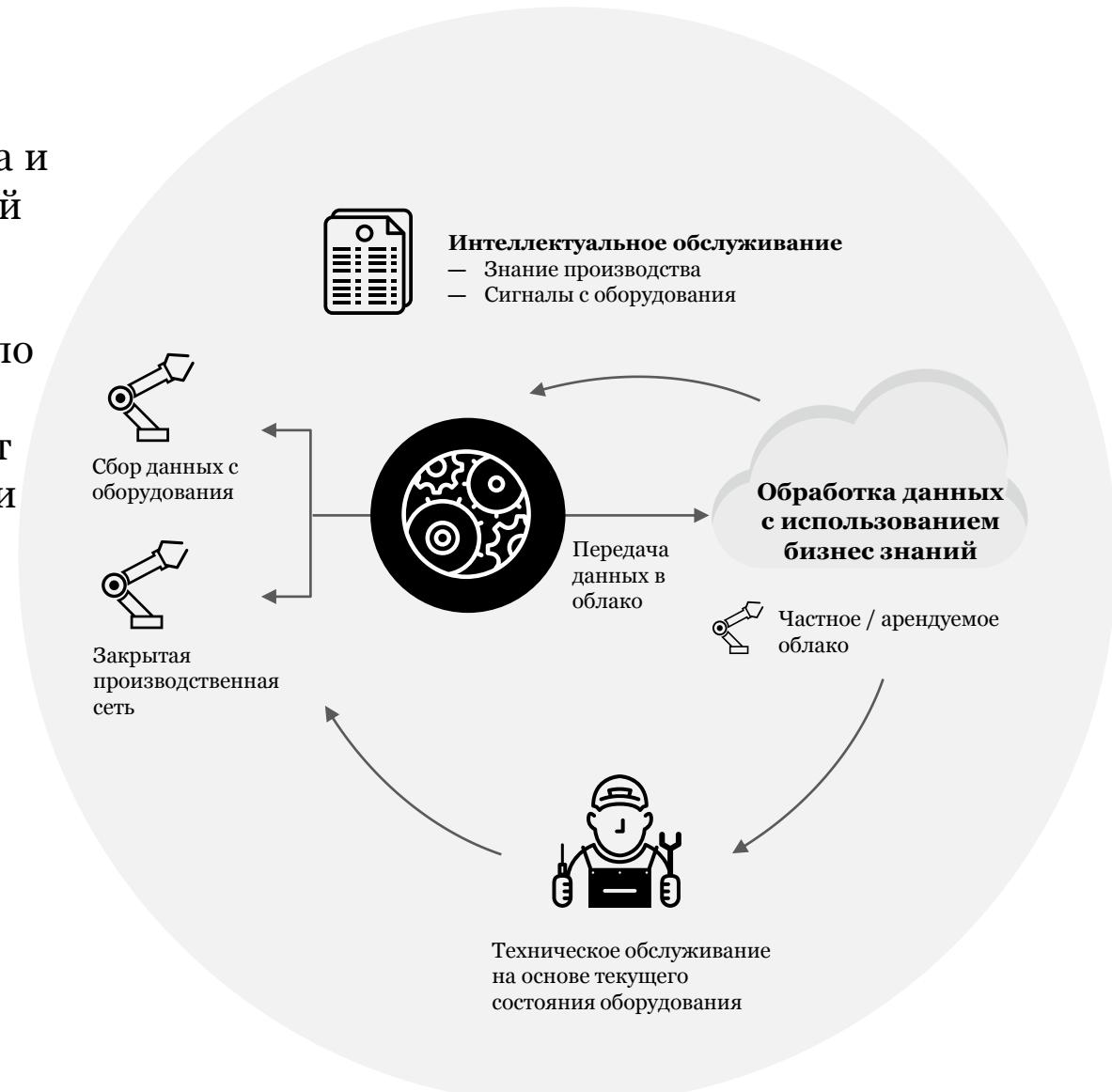


# Пример использования технологий машинного обучения для предиктивного обслуживания

Западная государственная компания, обслуживающая поезда и инфраструктуру железнодорожной сети

Трансформирует бизнес-процесс по техническому обслуживанию и ремонтам. Компания переходит от планового обслуживания к модели предиктивного обслуживания

**Система предсказала необходимость обслуживания для **61%** железнодорожных переключателей, которые позднее вышли из строя**



# Пример использования технологий машинного обучения для мониторинга транспортного парка

## Оптимизация маршрутов

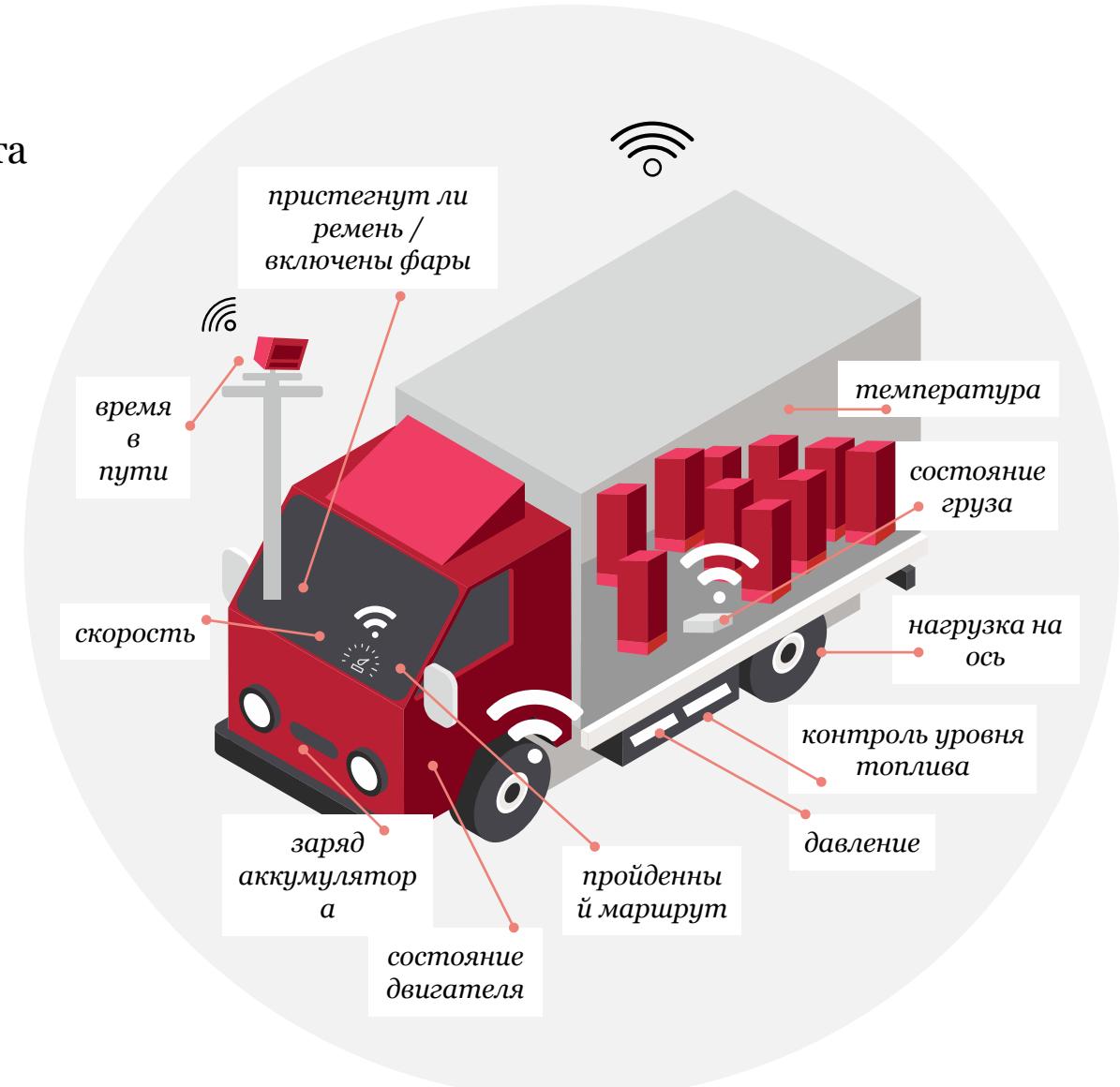
- Нахождение оптимального пути
- Наблюдение за отклонениями от маршрута
- Снижение транспортных расходов

## Удаленный контроль состояния транспорта

- Повышение безопасности
- Снижение рисков мошенничества
- Онлайн-мониторинг заказа клиентами

## Прогнозное обслуживание транспортного парка

- Обслуживание по состоянию вместо планового
- Снижение непредвиденных поломок и аварий
- Снижение порожнего пробега для проведения ремонтов



# Пример использования технологий машинного обучения

## Предиктивное техническое обслуживание

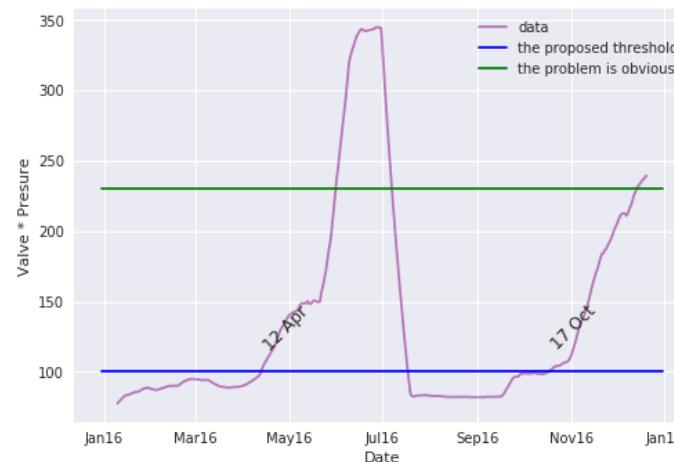
### Химический завод непрерывного цикла производства

#### Постановка задачи

Засор дистилляционной колонны одна из проблем производства

- Непредвиденный засор колонки останавливает производство на **2 недели**
- Плановая чистка колонки требует остановки производства на **3 дня**

На основе данных из производственных систем и создан алгоритм, который предсказывает событие наступление засора



# Пример использования технологий машинного обучения Снижение брака металлопроката

Металлопрокатный завод при прокатке металла получает **1% брака**

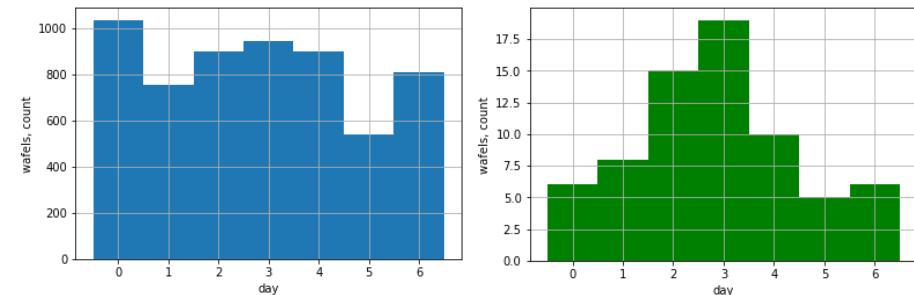
**Технологические параметры прокатного стана настраиваются некорректно в 1% случаев, в результате чего, прокатанные листы подвергаются деформациям**

**Задачи, стоящие перед нами:**

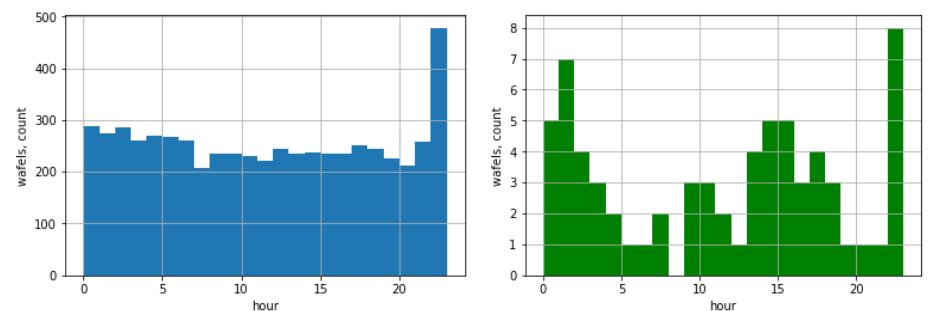
- Определить причины возникновения брака
- Обучить предиктивный алгоритм обнаружения таких событий
- Предотвратить возникновение брака в будущем путем внедрения решения

**Результаты:**

- Выявили сезонность возникновения брака
- Восстановили логику работы прокатного стана по логам
- Построенная модель прогнозирования брака показывала результаты более 98%

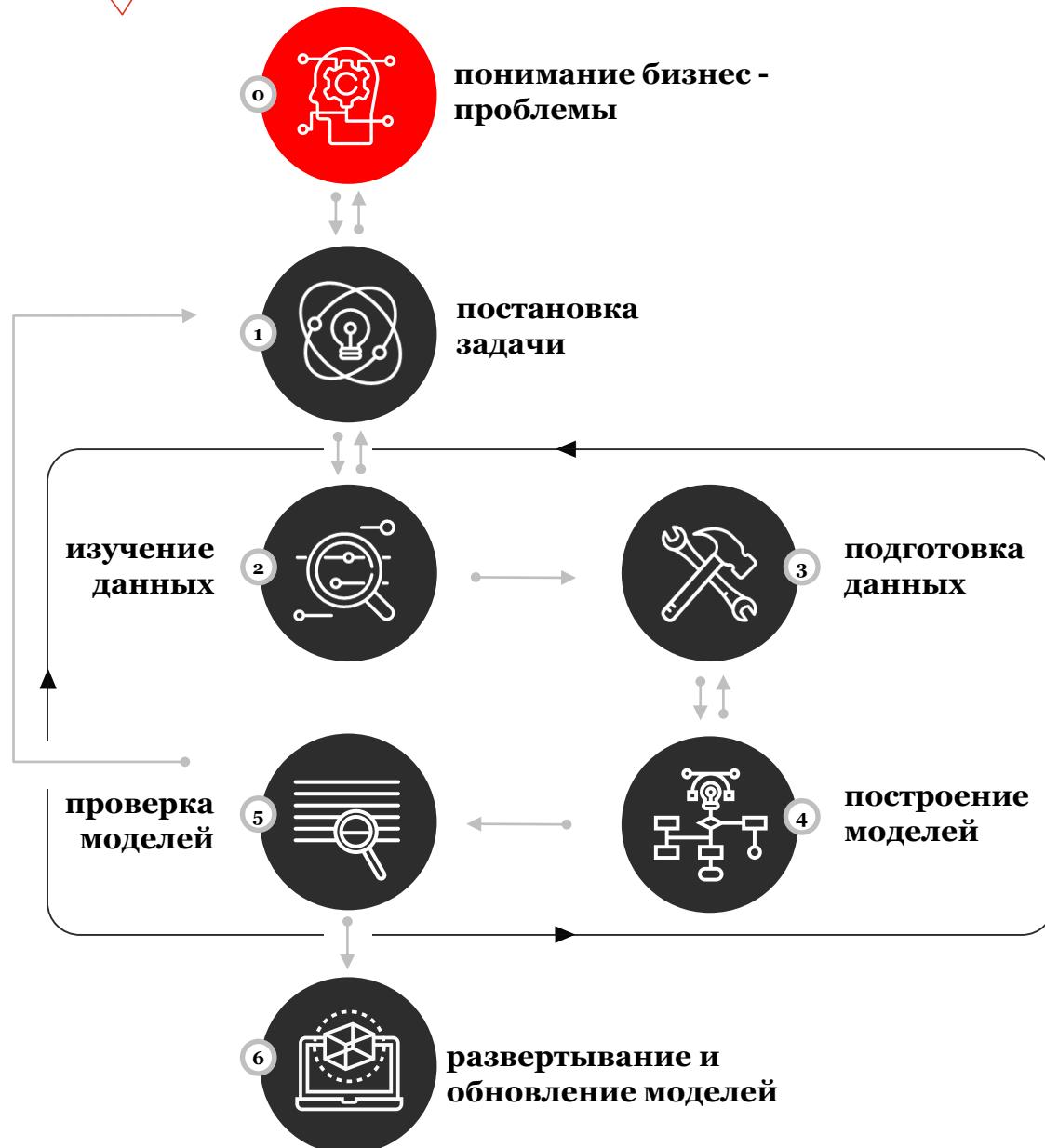


число брака увеличивается с понедельника по четверг



число брака увеличивается и коррелирует со сменой рабочих

# Ключевые этапы при внедрении технологий анализа больших данных и машинного обучения



## ЭТАПЫ:

- 0 —
  - Понимание бизнес-целей клиента
  - Анализ бизнес-процессов
  - Сбор бизнес-требований
- 1 —
  - Выдвижение гипотез
  - Формирование технической задачи
- 2 —
  - Описать исходные данные
  - Исследовать данные
  - Проверить качество данных
- 3 —
  - Выгрузка данных для моделей
  - Построение ETL
  - Стриминговая обработка данных
- 4 —
  - Proof of Concept
  - Выбор алгоритмов
  - Применение ML, DeepLearning
- 5 —
  - Проверка качества
  - Представление результатов
- 6 —
  - Непрерывное улучшение
  - Управление проектом
  - Тестирование



---

Области применения технологий анализа  
Больших данных и машинного обучения

## Подход

Обработка в  
режиме  
реального  
времени

Машинное  
обучение

Имитационное  
моделирование

Text mining

## Данные

 Веб-ресурсы,  
продукт,  
датчики,  
приложения

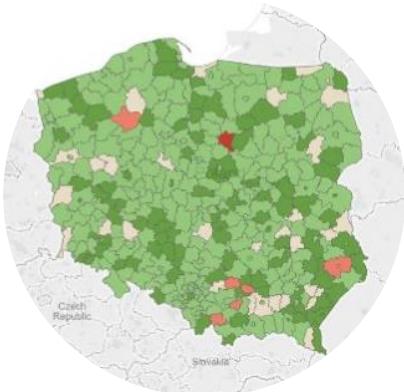
 Внешние  
данные

 Рыночные  
данные

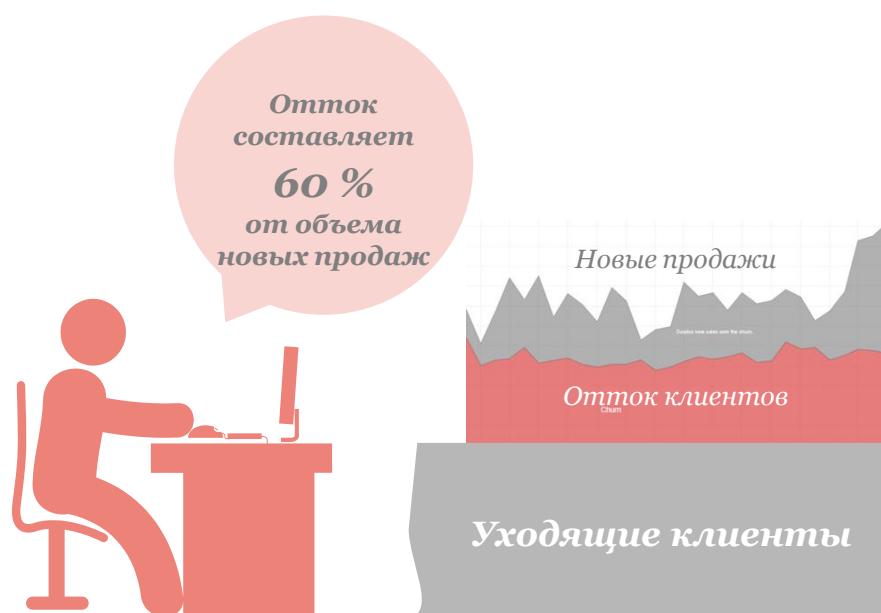
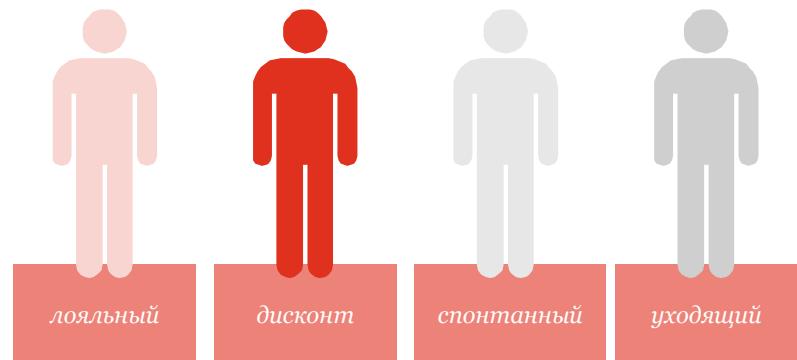
 Бизнес-  
партнеры

# Выявить отток клиентов и повысить лояльность клиентов

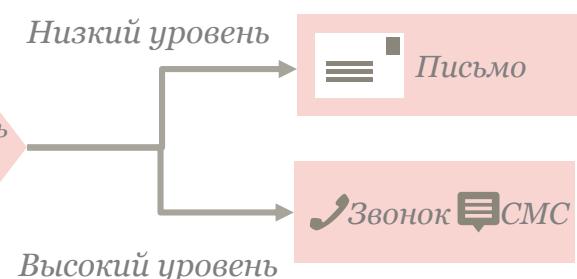
Качественная  
аналитика всего  
спектра  
информации о  
клиенте позволяют  
заблаговременно  
выявлять  
отток клиентов...



Повышение лояльности и доходности от текущей клиентской базы достигается путем разработки маркетинговых кампаний, направленных на каждого клиента, с учетом особенностей его поведения.



Анализ всего перечня информации о клиенте, включая его поведение в социальных медиа, помогает выявлять различные причины ухода клиентов



**Подход**

Машинное обучение

Имитационное моделирование

Text mining

Анализ социального графа клиентов

**Данные**

Социальные сети

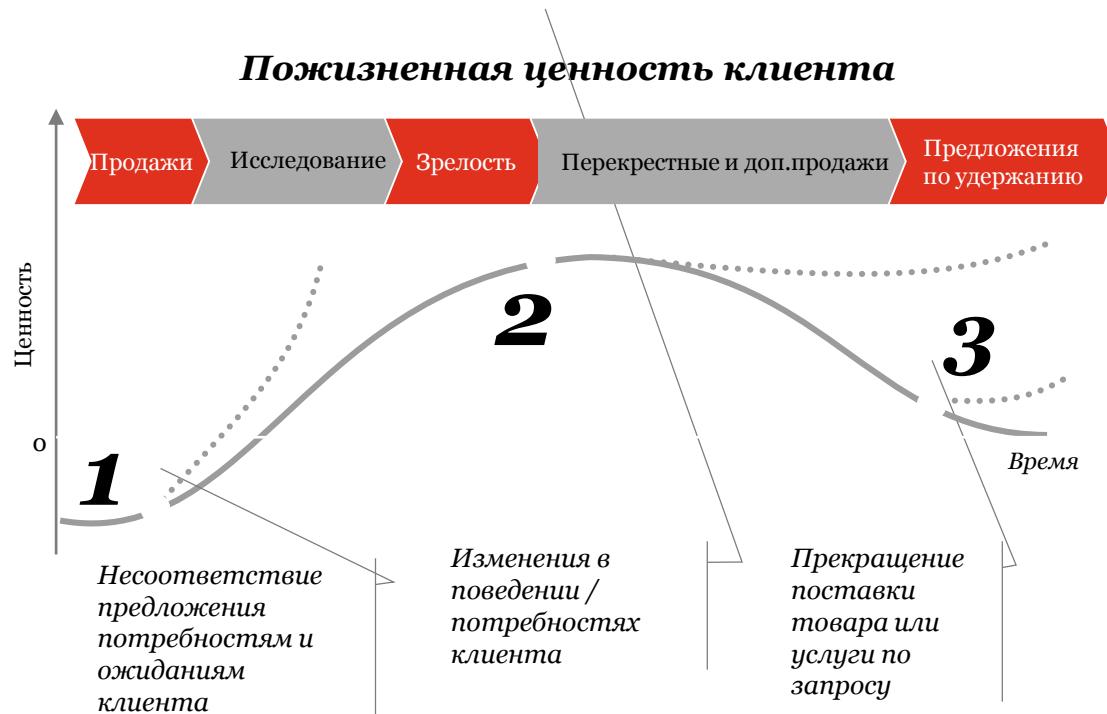
Внешние данные

Рыночные данные

Веб-ресурсы, продукт, датчики, приложения

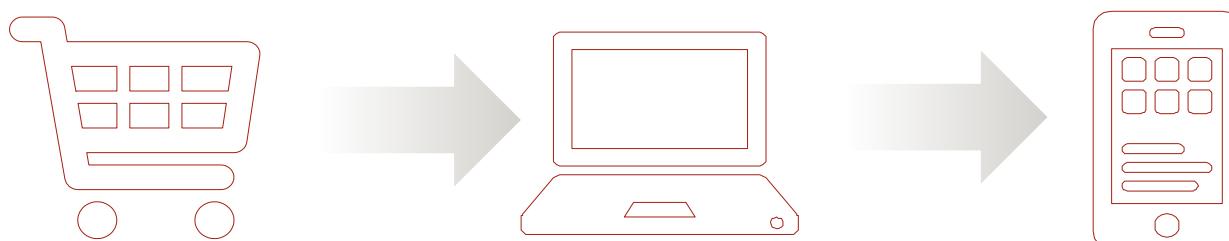
# Расчет ценности клиента и оптимизация расходов на маркетинг

*Ожидания и приоритеты клиентов постоянно меняются*



*Маркетинговые мероприятия, в основе которых лежит расчет ожидаемой выручки от клиентов в течение всего периода сотрудничества клиента с компанией, способствуют усилению связи (ассоциации) с компанией*

*Оптимизация маркетинговых расходов и предоставление возможности миграции клиентов между каналами в соответствии с их ценностью*



Подход

Обработка в  
режиме  
реального  
времени

Изучение  
ассоциативных  
правил

A/B-  
тестирование

Данные

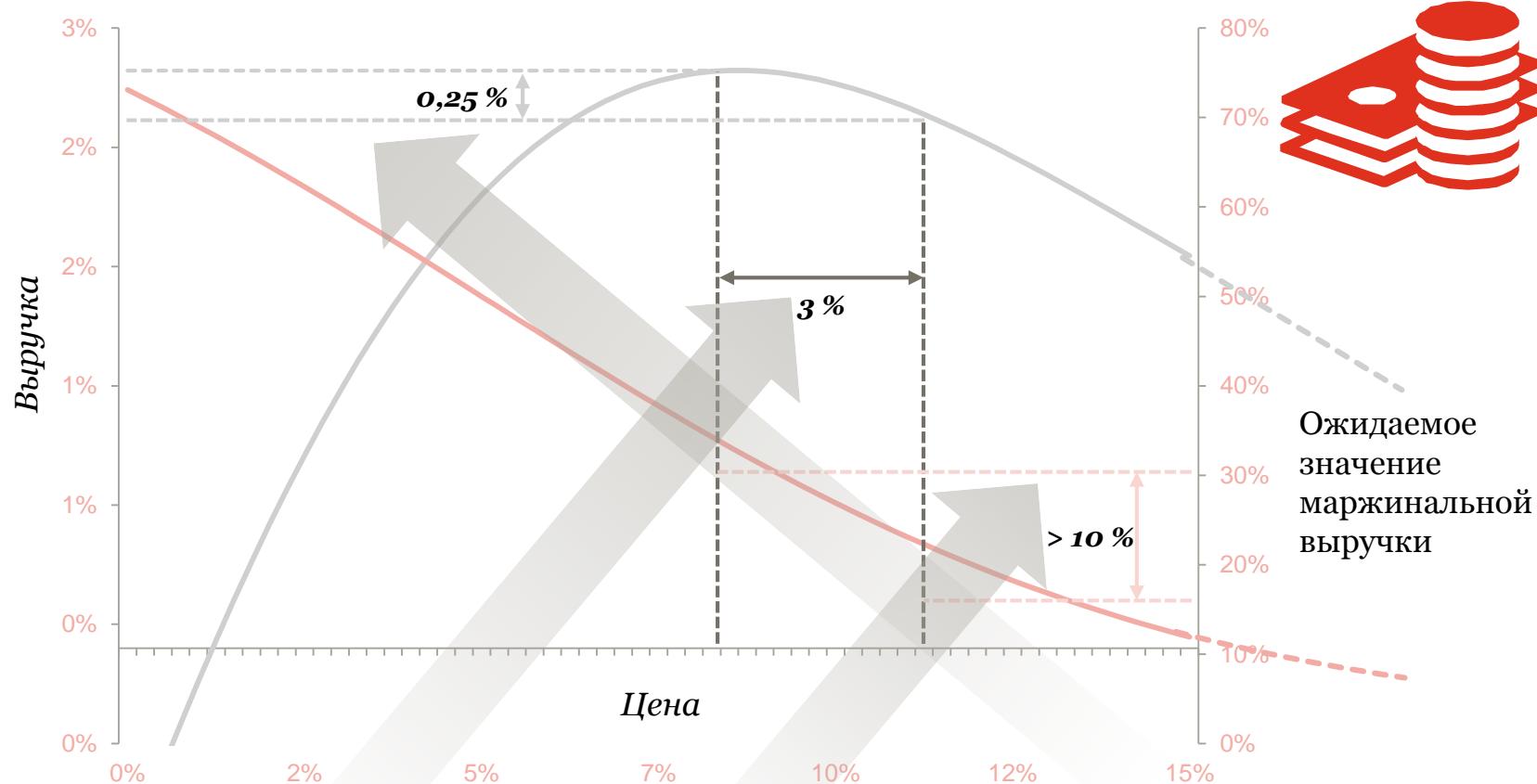
Web-ресурсы,  
продукт,  
датчики,  
приложения

Внешние  
данные

Рыночные  
данные

Бизнес-  
партнеры

# Определение идеальной цены для каждого клиента с целью максимального увеличения выручки



Внедрение гибкого  
ценообразования на базе  
модели эластичности..

...может привести  
к привлечению  
новых клиентов

Мы можем найти цену, которая  
обеспечит максимальную  
ожидаемую выручку

Подход

Обработка в  
режиме  
реального  
времени

Машинное  
обучение

Изучение  
ассоциативных  
правил

Сегментация

Данные

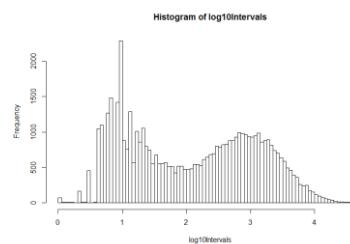
Социальные  
сети

Веб-ресурсы,  
продукт,  
датчики,  
приложения

Внешние  
данные

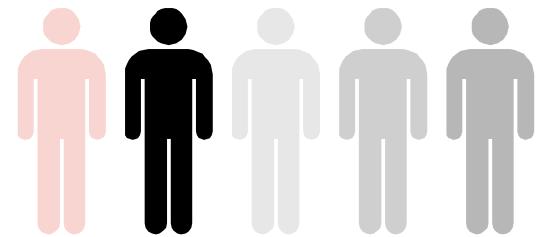
Рыночные  
данные

# Персонализация клиентского опыта, создание возможностей для перекрестных и дополнительных продаж



Информация о  
покупках |  
транзакциях  
клиента

Анализ рыночной  
корзины позволяет  
лучше разобраться  
в шаблонах  
покупок...



... что улучшает  
персонализацию вашего  
предложения

Рекомендательные системы позволяют уловить потребности клиента и  
тем самым повысить текущие продажи

Связь товаров



Корзина



Покупатели этого товара  
также купили



**Подход**

Машинное  
обучение

Имитационное  
моделирование

Text mining

Анализ  
социальных  
медиа

**Данные**


Социальные  
сети



Внешние  
данные



Бизнес-  
партнеры



Веб-ресурсы,  
продукт,  
датчики,  
приложения

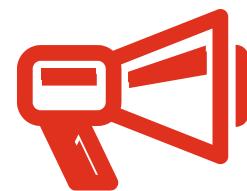
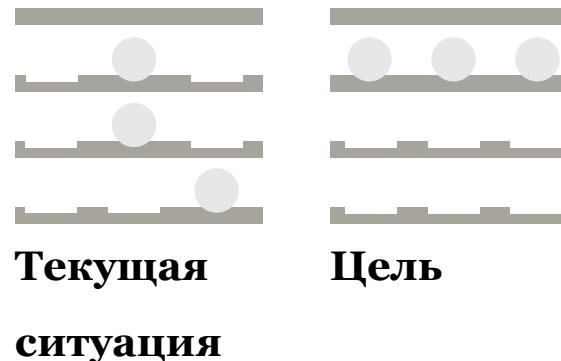
# Моделирование поведения потребителей...

*... и  
окружающих  
условий для  
повышения  
эффективност  
и процессов*

*... или точного  
выстраивания  
продуктового  
маркетинга в  
магазине*



Оптимальный  
ассортимент продукции



Маркетинговые  
мероприятия

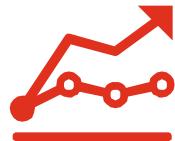
*При помощи  
анализа планировки  
в магазине можно  
оптимизировать  
распределение  
товаров и  
увеличить объемы  
продаж*



*... и на рынке*

# Непрерывный контроль за маркетинговыми расходами

Получение представления о результатах эффективных кампаний...



Тщательный анализ прошлых кампаний



Разработка маркетингового плана



Контроль эффективности продвижения товаров в режиме реального времени

## Определение ключевых факторов



День недели



Каналы



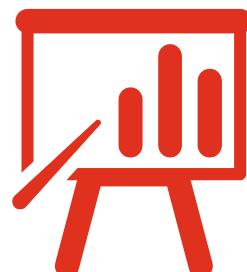
Маркетинговые расходы



Целевая группа



Действия конкурентов



... выявление связи между рекламой и показателями эффективности бизнеса

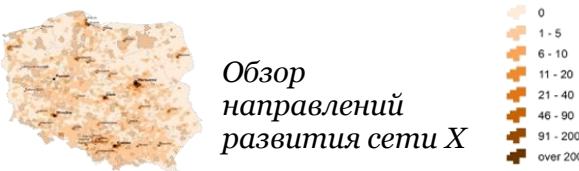
... для повышения рентабельности инвестиций



# Определение наиболее привлекательного местоположения точки продаж

1

Модель функциональных возможностей



Потенциал для открытия новых магазинов  
(максимальное число новых магазинов)

2

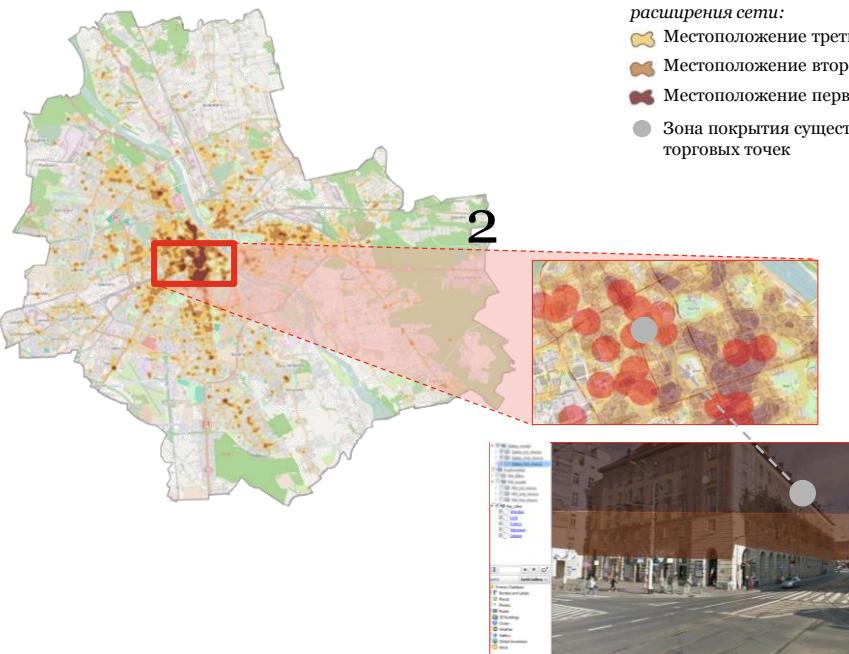
Геопространственный прогнозный анализ повышает точность определения оптимального местоположения магазинов для расширения сети

3

Экономическое обоснование



Пример анализа из нашего проектного опыта

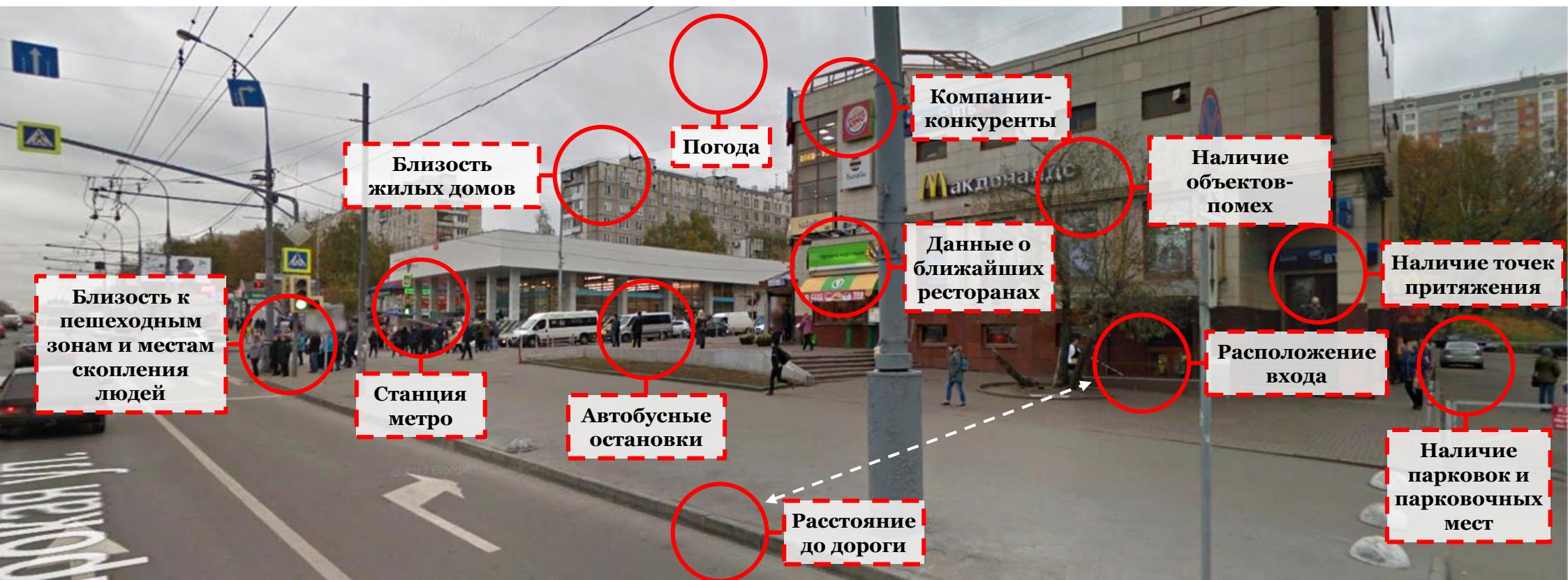


Пример показателей

Приоритетные направления расширения сети:  
● Местоположение третьего выбора  
● Местоположение второго выбора  
● Местоположение первого выбора  
● Зона покрытия существующих торговых точек

# Определение наиболее привлекательного местоположения точки продаж.

Детали.

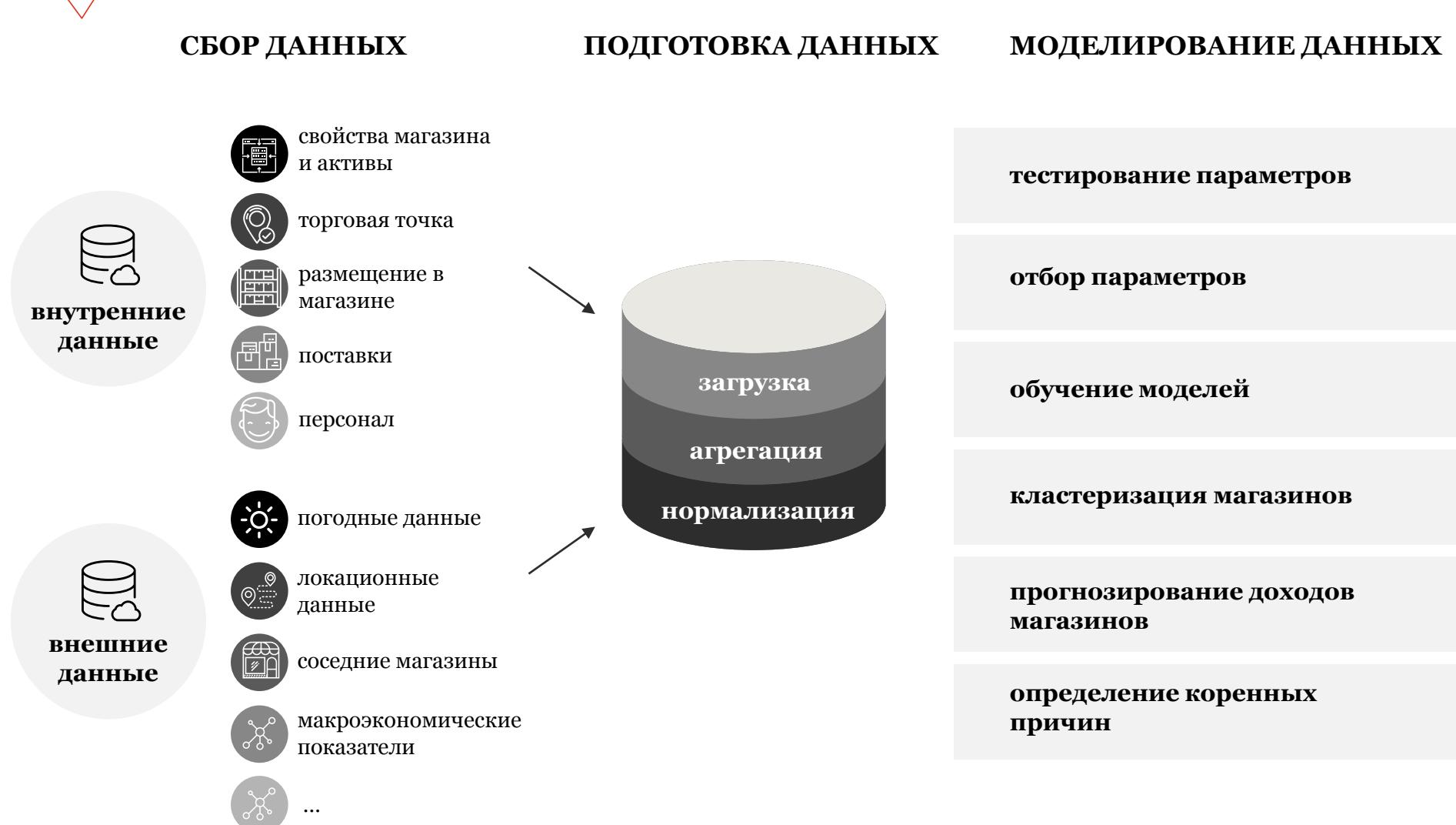


# Определение наиболее привлекательного местоположения точки продаж. Детали.



# Определение наиболее привлекательного местоположения точки продаж.

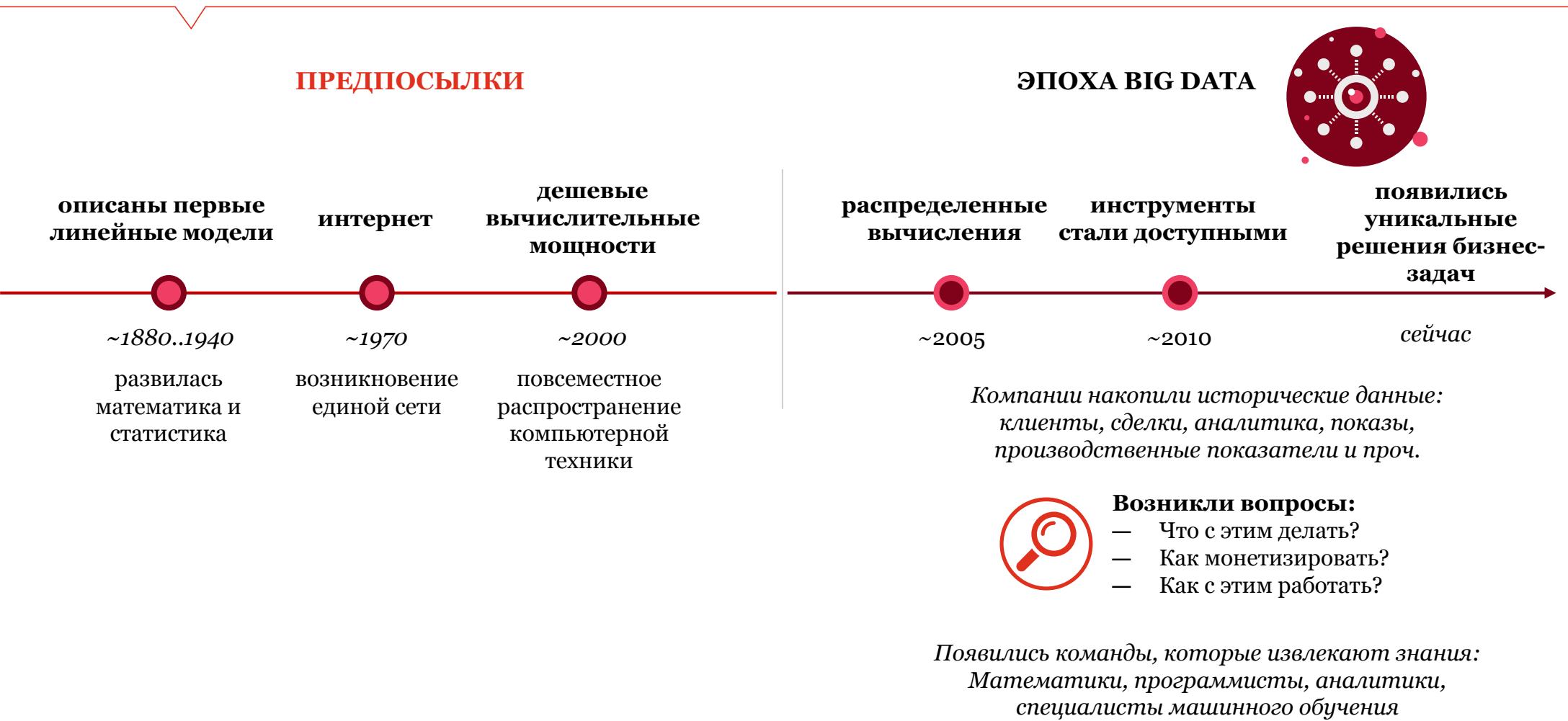
Детали.





Научно-технологические предпосылки развития

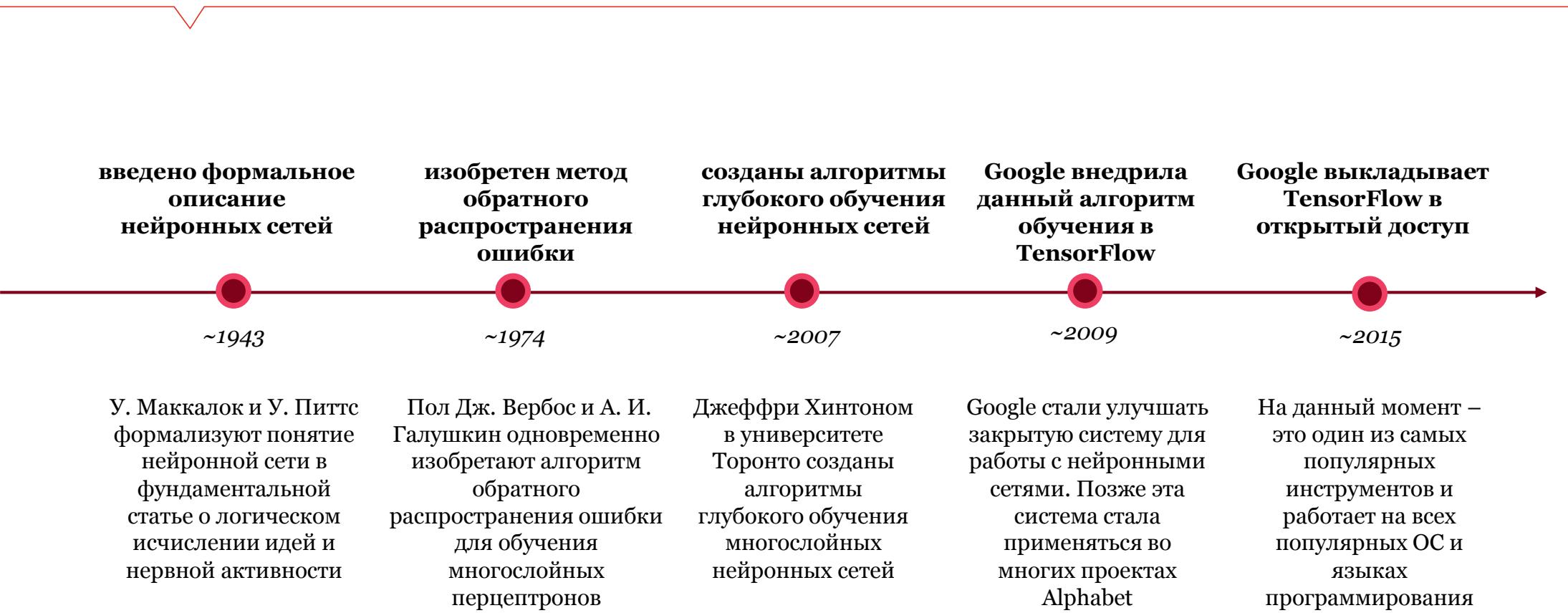
# С чего все начиналось



# Как инструменты машинного обучения стали доступными



# Как нейронные сети стали доступными



**Спасибо  
за внимание!**

**Олег Данильченко**

*Центр компетенции по прикладному  
анализу данных*

Oleg.danilchenko@pwc.com

Тел.: +7 (925) 400 99 13

PwC в России ([www.pwc.ru](http://www.pwc.ru)) предоставляет услуги в области аудита и бизнес-консультирования, а также налоговые и юридические услуги компаниям разных отраслей. В офисах PwC в Москве, Санкт-Петербурге, Екатеринбурге, Казани, Новосибирске, Ростове-на-Дону, Краснодаре, Воронеже, Владикавказе и Уфе работают более 2 500 специалистов. Мы используем свои знания, богатый опыт и творческий подход для разработки практических советов и решений, открывающих новые перспективы для бизнеса. Глобальная сеть фирм PwC объединяет более 223 000 сотрудников в 157 странах.

\* Под «PwC» понимается Общество с ограниченной ответственностью «ПрайсвотерхаусКуперс Консультирование» или, в зависимости от контекста, другие фирмы, входящие в глобальную сеть PricewaterhouseCoopers International Limited (PwCIL). Каждая фирма сети является самостоятельным юридическим лицом.

© 2017 Общество с ограниченной ответственностью «ПрайсвотерхаусКуперс Консультирование». Все права защищены. .