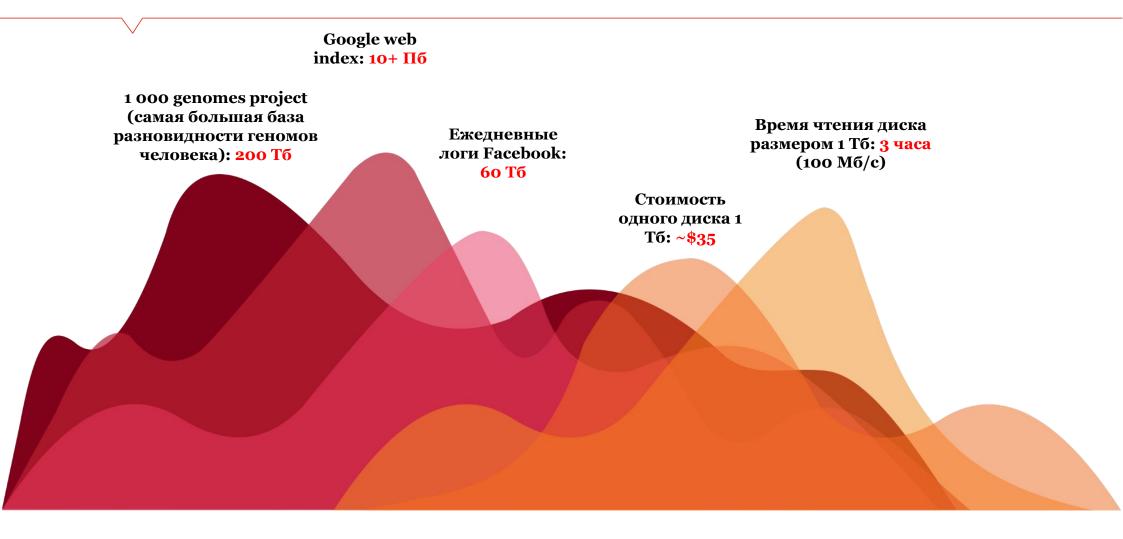
Big Data & Machine Learning

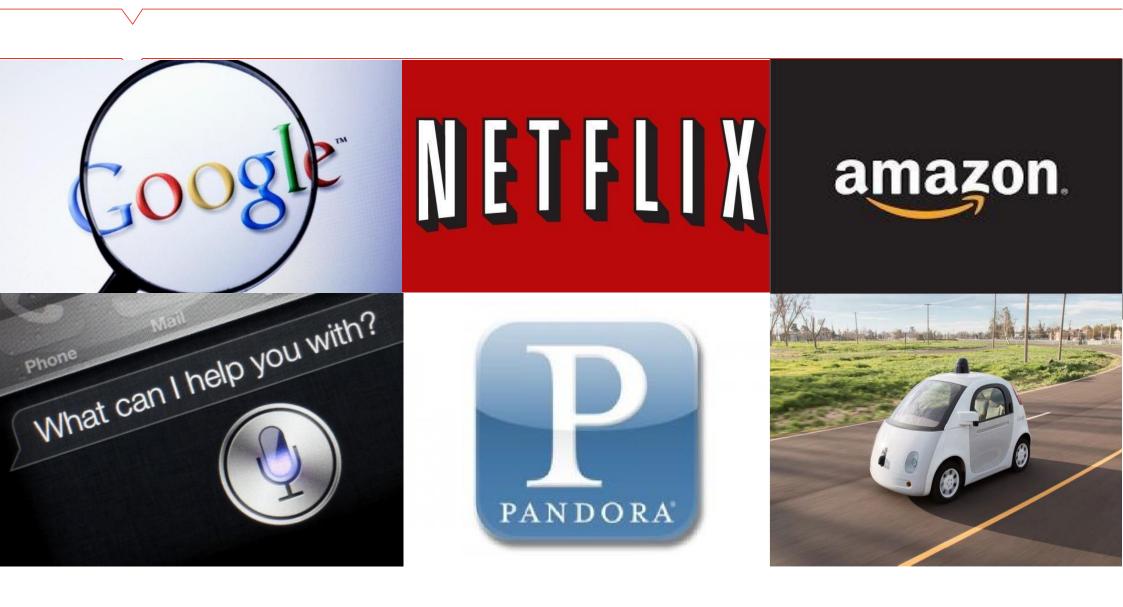
What's that. Why and what for?



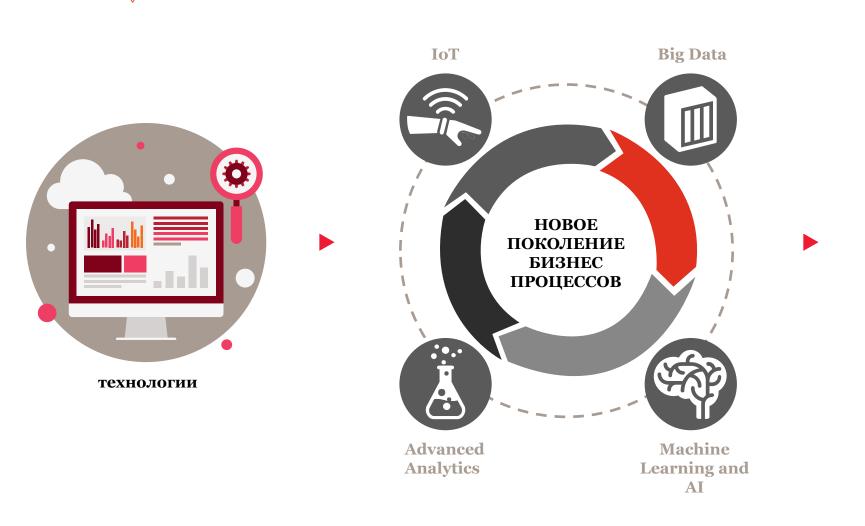
Данных много и их объем растет с каждым днем..



Технологии Больших данных и машинного обучения вокруг нас...



Существующие технологии меняют бизнес. Технологии анализа Больших данных, Машинное обучение и Интернет вещей.

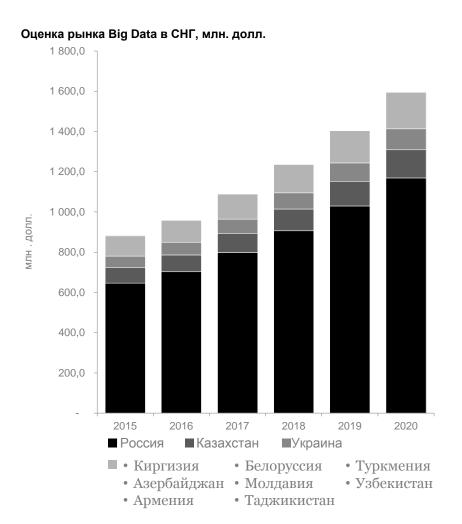


Значительное повышение эффективности

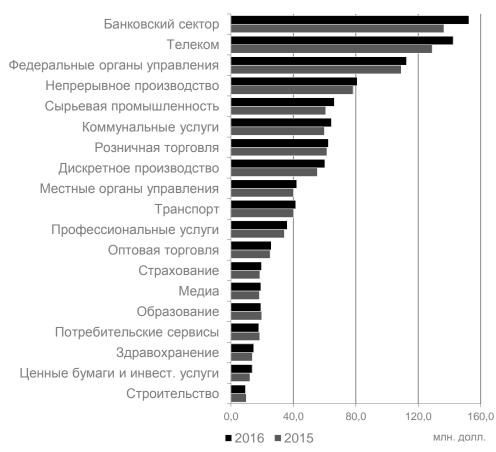
Создаются новые бизнесмодели

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

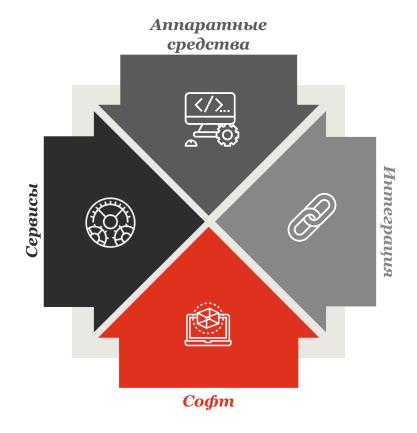


Области внедрения и применения решений в России, млн. долл.



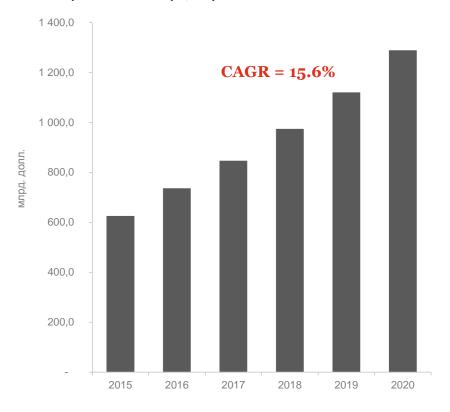
Глобальный рынок ІоТ растет двузначными темпами





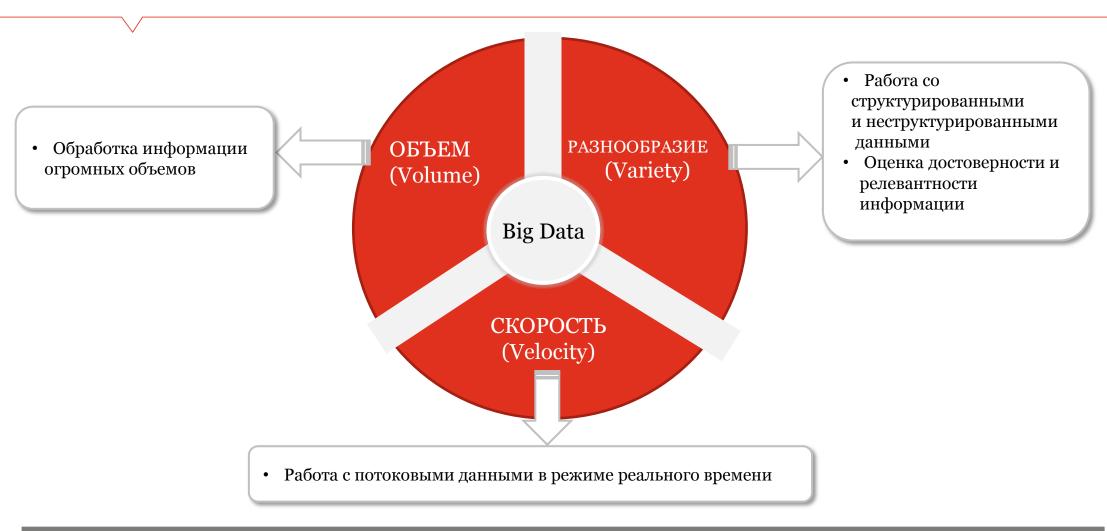
Средний ежегодный рост рынка ІоТ 15.6%

Оценка рынка ІоТ в мире, млрд. долл.



Источник: IDC

Что такое Большие данные?



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 применяется 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 компании 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 СЕТЯМИ

GPU



TPU



ОБЛАЧНЫЕ РЕШЕНИЯ



арендуемые облака:



Что такое Интернет Вещей (ІоТ)?

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

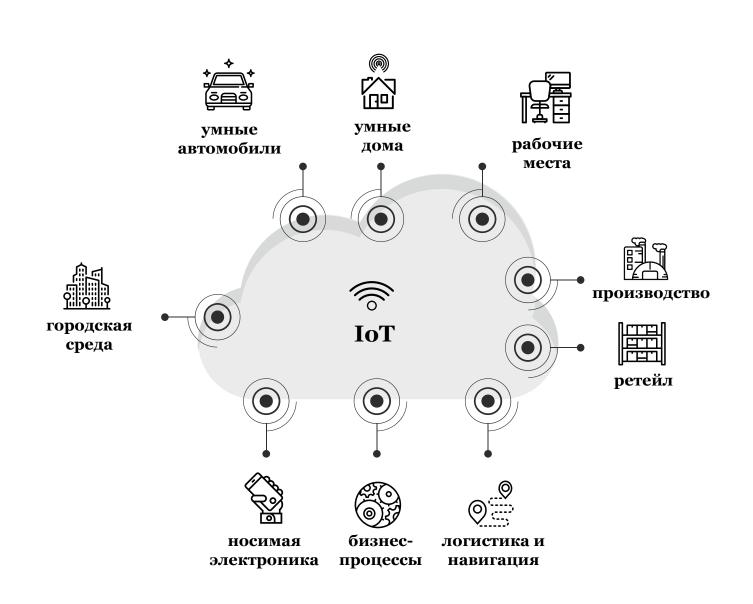


Machine-to-Machine





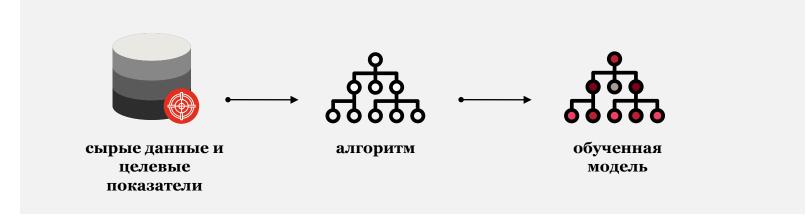
Где применяется технологии анализа больших данных и машинного обучения?



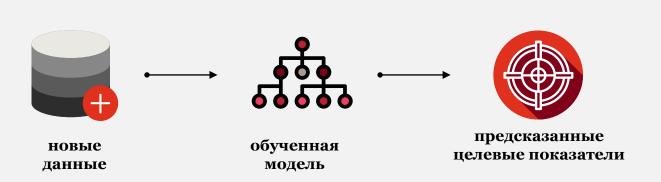
Что такое Машинное обучение?

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



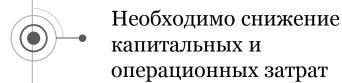






Пример использования технологий машинного обучения для мониторинга и контроля состояния

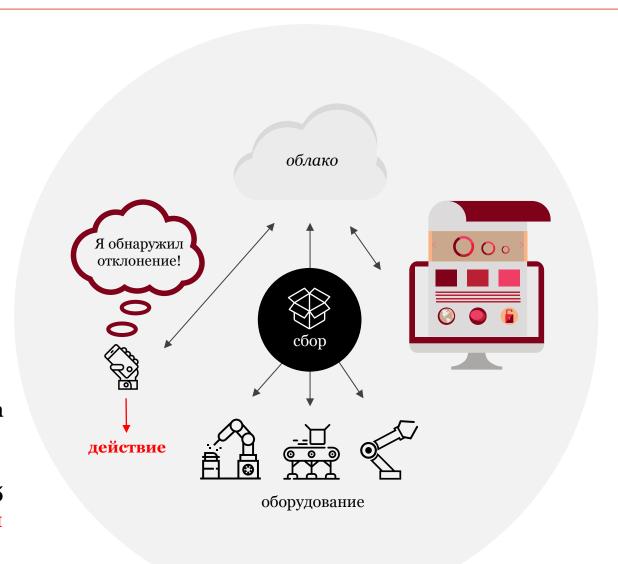
ЖКХ: водоснабжение



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

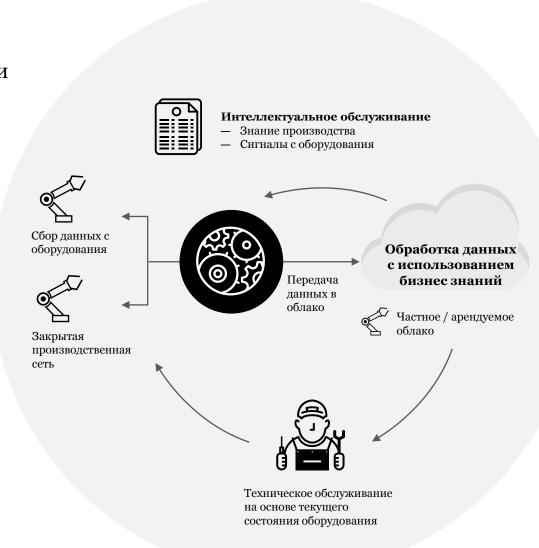


Пример использования технологий машинного обучения для предиктивного обслуживания

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

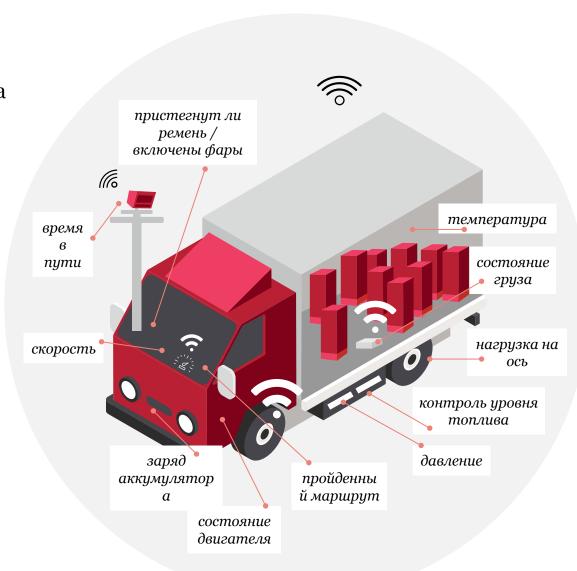
- Нахождение оптимального пути
- Наблюдение за отклонениями от маршрута
- Снижение транспортных расходов

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

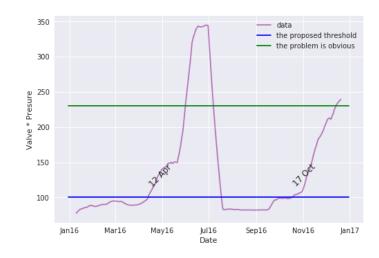


Засор дистилляционной колонны одна из проблем производства

- 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

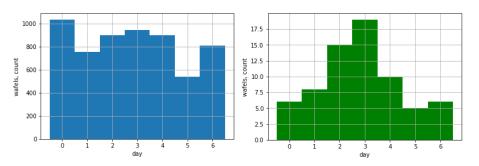
Технологические параметры прокатного стана настраиваются некорректно в 1% случаев, в результате чего, прокатанные листы подвергаются деформациям

Задачи, стоящие перед нами:

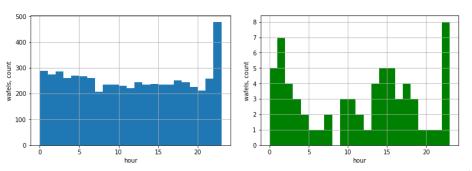
- Определить причины возникновения брака
- Обучить предиктивный алгоритм обнаружения таких событий
- Предотвратить возникновение брака в будущем путем внедрения решения

Результаты:

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

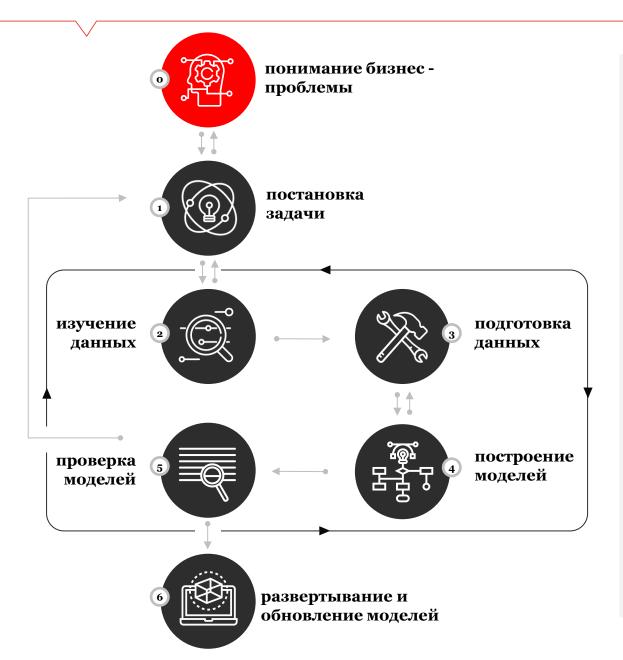


число брака увеличивается с понедельника по четверг



число брака увеличивается и коррелирует со сменой рабочих

Ключевые этапы при внедрении технологий анализа больших данных и машинного обучения



ЭТАПЫ:

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



Области применения технологий анализа Больших данных и машинного обучения

Обработка в режиме реального времени

Машинное обучение

Имитационное моделирование

Text mining

Данные









Бизнеспартнеры

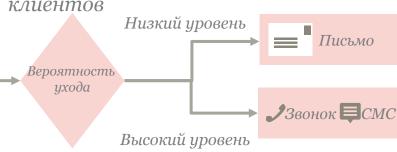
Выявить отток клиентов и повысить лояльность клиентов

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





Машинное обучение

Имитационное моделирование

Text mining

Анализ социального графа клиентов

Данные





Внешние данные

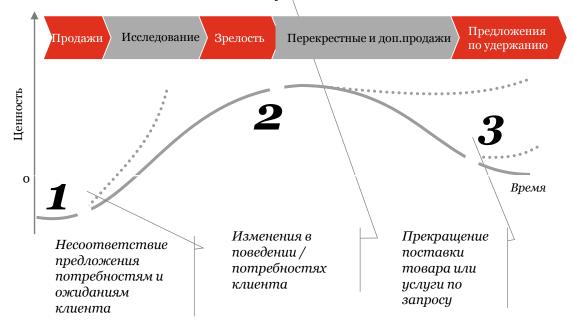




Расчет ценности клиента и оптимизация расходов на маркетинг

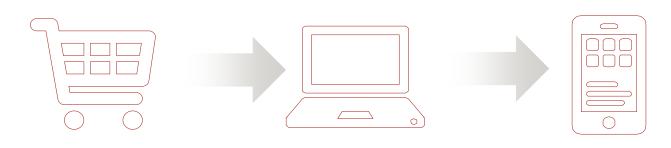
Ожидания и приоритеты клиентов постоянно меняются

Пожизненная ценность клиента



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



Обработка в режиме реального времени

Изучение ассоциативных правил

А/Втестирование

Данные



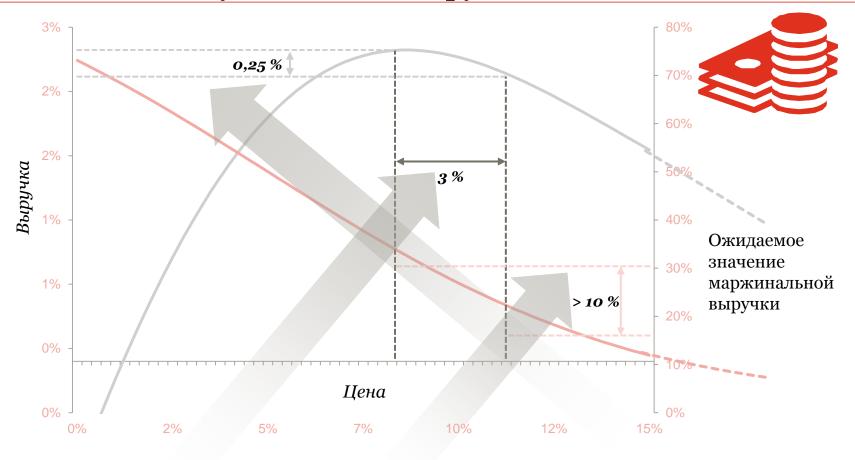






Бизнеспартнеры

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

Обработка в режиме реального времени

Машинное обучение

Изучение ассоциативных правил

Сегментация

Данные









Персонализация клиентского опыта, создание возможностей для перекрестных и дополнительных продаж



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







Машинное обучение

Имитационное моделирование Text mining

Анализ социальных медиа

Данные







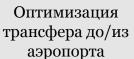
Бизнеспартнеры



Моделирование поведения потребителей...

... и окружающих условий для повышения эффективност и процессов







Сеть общественного транспорта

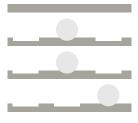


Торговый центр

... или точного выстраивания продуктового маркетинга в магазине

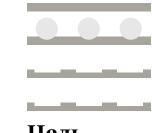


Оптимальный ассортимент продукции



ситуация

Текущая



Цель



Маркетинговые мероприятия

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

продаж

... и на рынке

Машинное обучение

Геоаналитика

А/Втестирование

Разбивка на кластеры



Данные









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





... выявление связи между рекламой и показателями эффективности бизнеса

... для повышения рентабельности инвестиций

Машинное обучение

Геопространс твенный анализ

Данные



Веб-ресурсы, продукт, датчики, приложения



Статистическ ие бюро





Внешние данные

Определение наиболее привлекательного местоположения точки продаж

Модель функциональных возможностей

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

Экономическое обоснование







Потенциал для открытия новых магазинов (максимальное число новых магазинов)



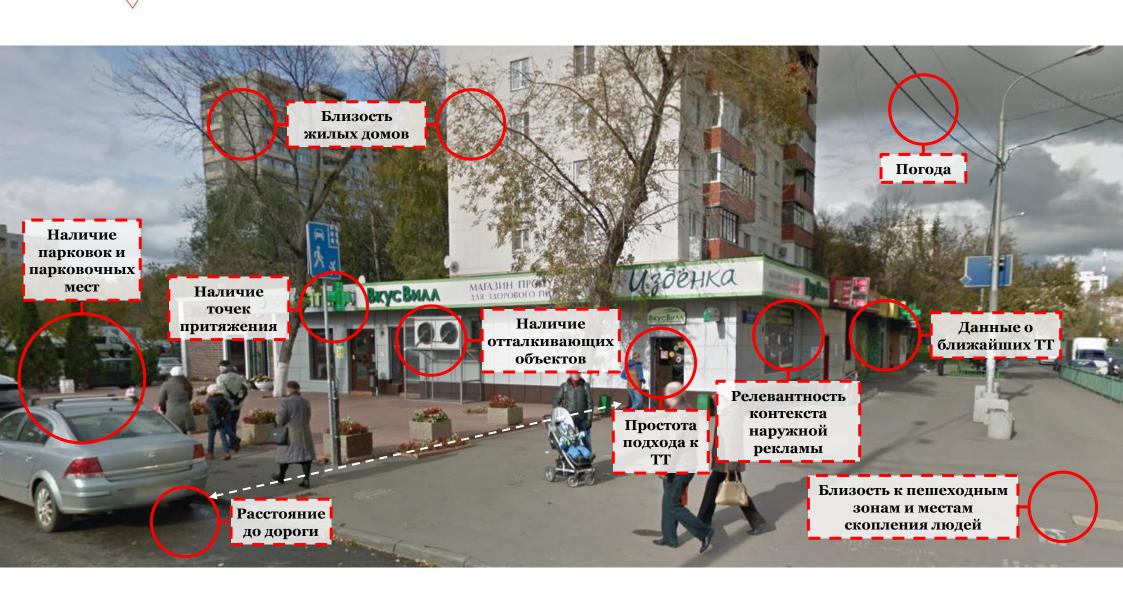
Определение наиболее привлекательного местоположения точки продаж.

Детали.



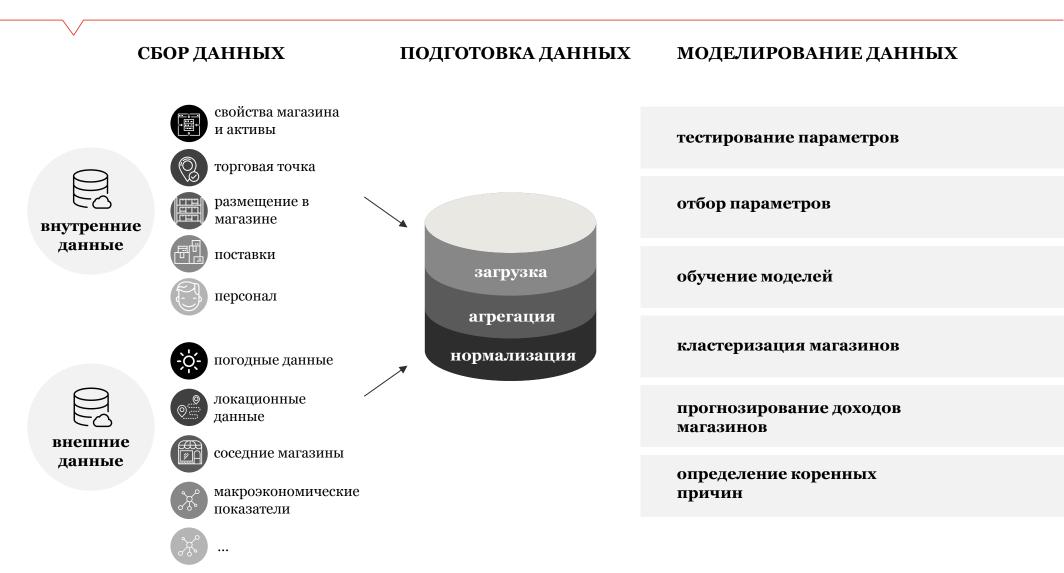
Определение наиболее привлекательного местоположения точки продаж.

Детали.



Определение наиболее привлекательного местоположения точки продаж.

Детали.





Научно-технологические предпосылки развития

С чего все начиналось

ПРЕДПОСЫЛКИ

дешевые

техники



ЭПОХА BIG DATA

описаны первые интернет вычислительные линейные модели мощности ~1880..1940 ~1970 ~2000 развилась возникновение повсеместное единой сети математика и распространение компьютерной статистика

Компании накопили исторические данные: клиенты, сделки, аналитика, показы, производственные показатели и проч.

~2010



~2005

Возникли вопросы:

- Что с этим делать?
- Как монетизировать?

сейчас

Как с этим работать?

Появились команды, которые извлекают знания: Математики, программисты, аналитики, специалисты машинного обучения

Как инструменты машинного обучения стали доступными

Salford Systems Яндекс копирует Открытый аналог XGboost решает Яндекс создали TreeNet **TreeNet TreeNet - XGboost** предоставляет свой задачи **CatBoost** ранжирования сейчас ~2007-2008 ~2002 ~2013 ~2015 библиотека компания построила компания копирует группа специалистов Яндекс открыл код становится одной из своей библиотеки, бизнес на продаже библиотеку для создает аналог библиотеки с более

высокой точностью и

открытым исходным

кодом

самых популярных,

переписана на многие

языки

программирования

которая по качеству

превосходит все

предыдущие

библиотеки,

реализующей алгоритм

градиентного бустинга

решения задач

ранжирования, что

позволило выжить и

успешно провести ІРО

Как нейронные сети стали доступными

введено формальное описание нейронных сетей изобретен метод обратного распространения ошибки созданы алгоритмы глубокого обучения нейронных сетей Google внедрила данный алгоритм обучения в TensorFlow

Google выкладывает TensorFlow в открытый доступ



~1943

~1974

~2007

~2009

~2015

У. Маккалок и У. Питтс формализуют понятие нейронной сети в фундаментальной статье о логическом исчислении идей и нервной активности

Пол Дж. Вербос и А. И. Галушкин одновременно изобретают алгоритм обратного распространения ошибки для обучения многослойных перцептронов

Джеффри Хинтоном в университете Торонто созданы алгоритмы глубокого обучения многослойных нейронных сетей

Google стали улучшать закрытую систему для работы с нейронными сетями. Позже эта система стала применяться во многих проектах Alphabet На данный момент — это один из самых популярных инструментов и работает на всех популярных ОС и языках программирования

Спасибо за внимание!

Олег Данильченко

Центр компетенции по прикладному анализу данных

Oleg.danilchenko@pwc.com Тел.: +7 (925) 400 99 13

РwC в **России** (www.pwc.ru) предоставляет услуги в области аудита и бизнесконсультирования, 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. .