

Большие данные

- **Большие данные (Big Data)** — совокупность подходов, инструментов и методов, предназначенных для обработки структурированных и неструктурированных данных (в т.ч. из разных независимых источников) с целью получения воспринимаемых человеком результатов. Большие данные характеризуются значительным объемом, разнообразием и скоростью обновления, что делает стандартные методы и инструменты работы с информацией недостаточно эффективными.

**Технология Больших данных — это инструмент принятия решений на основе больших объемов информации.**

Примерами источников информации, для которых необходимы методы работы с большими данными могут служить:

- логи поведения пользователей в Интернете;
- GPS-сигналы от автомобилей для транспортной компании;
- информация о транзакциях всех клиентов банка;
- информация о всех покупках в крупной розничной сети;
- информация с многочисленных городских IP-видеокамер;
- информация с датчиков большого производства, оборудованного технологией промышленного Интернета и т.д.

**Большие данные объединяют  
техники и технологии, которые  
извлекают смысл из данных на  
экстремальном пределе  
практичности.**

# Использование технологии больших данных позволяют: **Лучше понимать клиентов**

Большие данные могут рассказать много нового о том, что заставляет ваших клиентов покупать предлагаемые им продукты и услуги.

## **Разработать целевые маркетинговые сообщения**

Ключ к пониманию ваших клиентов – ответ на вопрос «Как и когда они делают покупки?», «Что является для них ценным?». На основе подобной информации вы можете разработать более эффективные маркетинговые сообщения для рынка и снизить затраты на привлечение клиентов. Большие данные могут помочь вашей компании (вне зависимости от ее размера) решить, когда, где и как ориентироваться на потенциальных клиентов и какой контент будет работать лучше всего в настоящий период времени.

## **Ускорить вывод на рынок новых продуктов**

Данная причина особенно важна для онлайн-продуктов и услуг, поскольку пользовательские данные клиентов могут помочь вам понять, что и как работает, а что нет. Сколько времени ваши клиенты пользуются продуктом? Что мотивирует их продолжать пользоваться продуктом или наоборот - отказаться? Благодаря этой информации, вы можете увеличить целевые показатели существующих продуктов и сделать новые продукты с более высокими результатами

## **Быстрая проверка идей**

Особенность малого бизнеса заключается в большей гибкости по сравнению с корпорациями. С помощью больших данных вы можете быстро извлечь выгоду из полученных идей, проверить их в практике и вывести продукты на рынок, прежде, чем ваши более крупные конкуренты начнут действовать

# Оптимизация путешествия клиента

Путешествие клиента – это маркетинговый термин. Большие данные как раз дают вам представление о путешествии клиента и его пользовательском опыте. Как потребители воспринимают наш бренд? Как они путешествуют через Ваш сайт? Сколько времени требуется, чтобы превратить их в продажу? С помощью этой информации Вы можете обнаружить точки прилипания в пути от впечатления до преобразования и сделать улучшения на этом пути.

## **Улучшение бизнес-процессов внутри организации**

Просматривая данные по существующим внутри компании бизнес-процессам, собранные компанией на протяжении многих лет, а также анализируя их, можно не только найти слабые места в бизнес-процессах, но и оптимизировать их либо переделать с нуля. Большие данные могут помочь вам предсказать, сколько клиентов прогнозируется в следующем квартале и помочь укомплектовать свои магазины.

# **Методы работы с большими данными:**

1. Машинное обучение
2. Анализ настроений
3. Анализ социальной сети
4. Ассоциация правил
5. Анализ дерева классификации
6. Генетические алгоритмы
7. Регрессионный анализ

## **Машинное обучение помогает:**

1. Различать спам и не спам в электронной почте
2. Изучать пользовательские предпочтения и давать рекомендации
3. Определять лучший контент для привлечения потенциальных клиентов
4. Определять вероятность выигрыша дела и устанавливать юридические тарифы

## **Анализ настроений помогает:**

- Улучшать обслуживание в гостиничной сети, анализируя комментарии гостей
- Настраивать стимулы и услуги для удовлетворения потребностей клиента
- Определить по мнениям в социальной сети о чем думают клиенты.

## **Анализ социальных сетей используют чтобы:**

1. Увидеть, как люди из разных групп населения формируют связи с посторонними лицами
2. Выяснить важность и влияние конкретного человека в группе
3. Найти минимальное количество прямых связей для соединения двух людей
4. Понять социальную структуру клиентской базы

## **С помощью правил ассоциации:**

1. Размещают продукты в большей близости друг к другу, чтобы увеличились продажи
2. Извлекают информацию о посетителях веб-сайтов из журналов веб-сервера
3. Анализируют биологические данные
4. Отслеживают системные журналы для обнаружения злоумышленников
5. Определяют чаще ли покупатели чая берут газированные напитки

## **Классификация** используется для:

1. Автоматического присвоения документов категориям
2. Классификации организмов по группам
3. Разработки профилей студентов, проходящих онлайн-курсы

# Генетические алгоритмы используют для:

1. Составления расписания врачей для отделений неотложной помощи в больницах
2. Расчет оптимальных материалов для разработки экономичных автомобилей
3. Создания «искусственно творческого» контента, такого как игра слов и шутки

# Регрессионный анализ используют для определения:

1. Уровней удовлетворенности клиентов
2. Как прогноз погоды за предыдущий день влияет на количество полученных звонков в службу поддержки
3. Как район и размер домов влияют на цену жилья

# Применение Больших данных (Big Data)

Самый быстрый рост расходов на технологии больших данных происходит в банковской сфере, здравоохранении, страховании, ценных бумагах и инвестиционных услугах, а также в области телекоммуникаций.

Банки и компании, выпускающие кредитные карты, используют большие данные, чтобы выявлять закономерности, которые указывают на преступную деятельность.

Снижение финансовых рисков – до 60% в банковском секторе;

- Больница может использовать его для обеспечения безопасности, актуальности данных пациента и полного сохранения их качества. Размещая базы данных о здоровье в блокчейн, больница обеспечивает всем своим сотрудникам доступ к единому, неизменяемому источнику информации.

**Реклама, маркетинг и торговля.** Не имеет значения, что вы предлагаете потребителю, качественные товары, услуги или яркие обещания – все это должно находить отклик. И тут снова приходят на помощь озера данных. Компании получили возможность предугадывать желания – предлагать продукт, который вы подсознательно хотите купить.

- Промышленность в целом и энергетический сектор в частности генерируют огромный объем информации, которая при применении традиционных методов аналитики используется не в полной мере. ВД позволяют извлекать, преобразовывать, обрабатывать и хранить данные, которые не всегда учитываются, но могут косвенно влиять на работу производства.

Большие данные открывают перед энергетикой и другие возможности. Они позволяют тщательнее контролировать динамику потребления электроэнергии, точно отслеживать дебиторскую задолженность, дают возможность качественнее управлять стоимостью контрактов и прогнозировать возникновение спроса на дополнительные услуги среди клиентов отрасли.

# Проблемы внедрения

**Первая проблема** — дефицит кадров: сектору не хватает специалистов по всем направлениям, связанным с обработкой и внедрением больших данных. Сейчас многие коммерческие компании готовят кадры самостоятельно, но этого недостаточно, чтобы решить эту проблему.

**Вторая — инфраструктура.** По количеству ЦОД и вычислительных мощностей РФ существенно уступает развитым странам. В США, например, они занимают около 65 млн кв. метров. В РФ — на порядок меньше. Нужно развивать инфраструктуру как для обработки данных, так и для передачи этих данных на высокой скорости.

**Третья — доступность данных.** У нас есть серьезные ограничения по обработке и обмену разными типами данных. Необходимы стандарты по анонимизации и деперсонализации данных, возможность свободного оборота данных, которые перестали быть персональными.

## **Четвертая проблема — исследования и идеи.**

Нет необходимой экспериментальной среды и ресурсов для разработки и внедрения новых технологий работы с Big Data. Для инноваторов необходимо создать облегченный доступ к большим данным, кросс-отраслевые инкубаторы.

**Пятая — невозможность массово имплементировать большие данные в масштабах всей экономики.**

# Проблемы Big Data

1) Самой большой проблемой больших данных являются **затраты на их обработку**. Сюда можно включить как дорогостоящее оборудование, так и расходы на заработную плату квалифицированным специалистам, способным обслуживать огромные массивы информации. Очевидно, что оборудование придется регулярно обновлять, чтобы оно не теряло минимальной работоспособности при увеличении объема данных.

**2) Большое количество информации,** которую необходимо обрабатывать. Если, например, исследование дает не 2-3, а многочисленное количество результатов, очень сложно остаться объективным и выделить из общего потока данных только те, которые окажут реальное влияние на состояние какого-либо явления

**3) Проблема конфиденциальности Big Data.** В связи с тем, что большинство сервисов по обслуживанию клиентов переходят на онлайн-использование данных, очень легко стать очередной мишенью для киберпреступников. Даже простое хранение личной информации без совершения каких-либо интернет-транзакций может быть чревато нежелательными для клиентов облачных хранилищ последствиями.

#### **4) Проблема потери информации.**

Меры предосторожности требуют не ограничиваться простым однократным резервированием данных, а делать хотя бы 2-3 резервных копии хранилища. Однако с увеличением объема растут сложности с резервированием – и IT-специалисты пытаются найти оптимальное решение данной проблемы

# Результаты внедрения технологий больших данных:

1. рост качества клиентского сервиса;
2. оптимизация интеграции в цепи поставок;
3. оптимизация планирования организации;
4. ускорение взаимодействия с клиентами;
5. повышение эффективности обработки запросов клиентов;
6. снижение затрат на сервис;
7. оптимизация обработки клиентских заявок