

«Большие данные»

Характеристики больших данных

- 1) объём (volume) – величина физического объёма данных. Большие данные измеряются в десятках терабайт;
- 2) скорость (velocity) – скорость постоянного прироста данных, а также необходимости высокоскоростной обработки и получения результатов на их основе;
- 3) многообразии (variety) – возможность одновременной обработки различных типов структурированных и полуструктурированных данных.

Сферы применения больших данных

В школе – рекомендательные системы позволяют получать курсы по интересам. На основании того, как школьник учится, можно прогнозировать, как он закончит учебный год, и корректировать его обучение, чтобы он вышел на более высокий результат, уведомлять родителей о росте, увлечениях их ребёнка и рекомендовать им дополнительное обучение.

В медицине – роботы могут распознавать симптомы болезней на ранних стадиях и с большей вероятностью рекомендовать правильное лечение.

В сфере безопасности – с помощью нейросетей происходит обработка видео на камерах в городах, которые позволяют ловить преступников почти сразу после того, как они попадают в их поле зрения.

На транспорте – навигаторы анализируют текущую ситуацию на дорогах и выбирают наиболее короткий путь до нужного места, при этом меняя его в зависимости от ситуации на дорогах.

Сферы применения больших данных

В соцсетях – умная лента рекомендует то, что будет интересно. Музыка в ВКонтакте может рекомендовать других исполнителей, которые могут понравиться пользователям, даже если они никогда раньше их не слышали. Нейросети могут раскрашивать фотографии и видео под картины известных художников.

В играх – тренироваться с ботами уже давно стало привычным делом. Программы могут выиграть у человека даже в таких интеллектуальных играх как шахматы.

В искусстве – нейросети умеют сочинять стихи, которые, практически неотличимы от написанных человеком. Они пишут целые произведения, которые проходят в финалы литературных конкурсов, создают музыку и новые картины.

Профессии в области больших данных

ИТ-медик. Врач с хорошим знанием информационных технологий, который создаёт, обрабатывает и анализирует физиологические данные и индивидуальные показатели здоровья и применяет к ним методики работы с большими данными. На основании проведённого анализа он предлагает индивидуальные медицинские и оздоровительные рекомендации для пациентов.

Архитектор информационных систем. Мало собрать данные. Их нужно упаковать и разместить в некотором месте. Для того, чтобы с данными было удобно работать, крайне важно правильно организовать их хранение и администрирование так, чтобы в любой момент любой пользователь мог получить доступ именно к тем данным и в том виде, в котором они ему необходимы. Этим занимается архитектор информационных систем.

Профессии в области больших данных

Клинический биоинформатик. Биоинформатика – это изучение процессов, связанных с организмом человека, но с помощью компьютерного моделирования. В случае нестандартного течения болезни клинический биоинформатик строит компьютерную модель биохимических процессов болезни, чтобы понять первопричины заболевания, выявляет нарушения на клеточном и субклеточном уровнях.

ИТ-проповедник. Обучение людей новым технологиям, убеждение их, что цифровой мир несёт им благо, а не зло, привлечение их на свою сторону – это задача, которая будет становиться всё важнее по мере того, как диджитализация будет проникать в привычный нам мир вещей. Задачи для такого специалиста две: обучение людей новым технологиям, помощь им в приобретении новых навыков, а также снятие зачастую иррациональных страхов перед цифровым миром.

Профессии в области больших данных

Оператор кросс-логистики. Этот специалист помогает логистической компании избежать опозданий и сэкономить деньги. Планируя грузоперевозки, приходится очень точно выверять необходимое количество транспорта, поскольку переоценка приведёт к ненужным расходам, а недооценка — к нехватке мощностей и длительным задержкам и, как следствие, штрафам.

Системный инженер интеллектуальных энергосистем.

Умные сети постоянно генерируют огромное количество данных на всех своих участках. Использование технологий больших данных в электроэнергетике позволяет не только сократить расходы производителей, транспортеров и конечных потребителей энергии, но и сделать электроэнергию более зеленой, а планету — более чистой. Поэтому специалисты в этой сфере так необходимы.

Профессии в области больших данных

BIM (Building Information Model) — проектировщик. Сейчас никто не строит здания по чертежам на бумаге. На всех этапах строительства используется компьютерное моделирование. Но это не просто чертежи, перенесенные в 3D-формат. Это целая система данных о каждом этапе строительства, то есть весь жизненный цикл здания – в одной системе. Конечно, это огромные объёмы данных, в которых должен разбираться BIM-инженер.

Дата-маркетолог. В задачи дата-маркетологов входит определение оптимального момента, когда клиенту стоит предложить тот или иной товар или услугу, дать дополнительную информацию или оказать иную помощь для повышения лояльности. Тем самым работа с большими данными делает бизнес более клиентоориентированным.

Профессии в области больших данных

Специалист по машинному обучению. Задача специалиста по машинному обучению — автоматизировать те процессы, где невозможно прописать чёткий алгоритм действия, принять решение и получить требуемый результат на основе некоторого набора больших данных. Например, создать программу, которая самостоятельно будет определять, какую рекламу и в какой момент необходимо показывать пользователю в зависимости от его действий.

Утилизатор больших данных. Этот человек проводит опросы, анкетирование, ищет закономерности в данных, выдвигая и проверяя самые безумные гипотезы, и способен найти причины тех или иных действий человека в онлайн, связать их с поступками и даже предложить человеку то или иное решение. При помощи больших данных ему удаётся определить психологические особенности человека и предсказать, как он будет действовать в офлайне.

Профессии в области больших данных

Трендвотчер/форсайтер. Специалист, который отслеживает появление новых тенденций в разных отраслях экономики, общественной жизни, политике или культуре, и составляет отчеты о влиянии этих новшеств на клиентские потребности. Он умеет прогнозировать тренды и улавливать важные изменения до того, как они станут общепринятыми направлениями развития.

Антифейковый менеджер. Этот специалист создаёт уникальный информационный стиль компании или конкретного человека, чтобы его было сложно скопировать при помощи алгоритмов. Он хорошо понимает, как создаются фейки, умеет отличать созданные нейросетью голос, текст и видео от оригинальных и умеет очищать факты в новостях.

Игровой аналитик. Может спрогнозировать, на чём можно увеличить доходы компании, за счёт исследования данных внутри игры и поиска закономерностей. Совместно с командой он подготавливает систему сбора игровых данных для последующего анализа, выявления поведения игроков и влияния внешних факторов на саму игру, анализирует игровую экономику и предлагает пути дальнейшего развития.

Профессии в области больших данных

Аналитик электронной коммерции. Аналитик электронной коммерции — это не только специалист по работе с клиентами и программными продуктами, но ещё и тот, кто умеет хорошо прогнозировать продажи, основываясь на обработке больших данных за предыдущие периоды, делать гипотезы и проводить рыночные тестирования.

Дата-инженер. Суть работы дата-инженера заключается в построении стабильных систем добычи и очистки данных, чтобы процесс последующего анализа полученных больших массивов данных был максимально удобным и доступным для аналитиков компании. Именно от написанных им программ зависит качество «сырых» данных на входе в модели, которые используются для создания новых алгоритмов.

Аналитик эксплуатационных данных. Анализируя текущее состояние сетей, аналитик делает прогноз по их дальнейшему развитию на ближайшие недели, месяцы или даже первые годы и принимает решение относительно необходимости монтажа нового или демонтажа старого оборудования. Задача аналитика — предвидеть оптимальный момент изменения сетей в будущем, чтобы к этому моменту всё было готово для трансформации.

Большие данные — это...?

«Большие данные — это данные, которые невозможно обработать на одном компьютере».

«Большими данными называют различные инструменты, подходы и методы обработки как структурированных, так и неструктурированных данных для того, чтобы их использовать для конкретных задач и целей».

Вопросы

«Почему сфера больших данных стала так популярна в последнее время?»

«Где применяется анализ больших данных?»

Итог урока

«Какая информация была для вас новой?»

«С какими источниками больших данных мы сталкиваемся в повседневной жизни?»

«Каким образом сами участвуем в генерации больших данных?»

«Для чего нужен анализ больших данных?»