

Урок Цифры

Цели и задачи урока.

Цель урока:

Сформировать у учеников представление о понятии «большие данные» и разобрать основные профессии, связанные с работой в данной области. Также поговорить о кибербезопасности.

Задачи урока:

1. Обсудить понятие «большие данные».
2. Разобрать примеры применения концепции больших данных в современном мире.
3. На основе игры смоделировать процесс обработки данных.
4. С помощью игры разобрать процесс обучение нейросети.
5. Обсудить полученный опыт, сформулировать выводы.

1. Вступление

«Сегодня у нас пройдет нестандартное занятие в рамках акции «Урок Цифры», которая проводится по всей России. Урок будет посвящен теме: большие данные. Акция направлена на развитие цифровых знаний и навыков, которые необходимы в настоящее время. Мы живем в цифровом мире и каждый день сталкиваемся с большими данными, даже если не подозреваем об этом. Например, пользуясь телефоном, компьютером вы приобщаетесь к теме больших данных. Как вы думаете, каким образом это происходит?» (Выслушайте ответы детей).

2. Обсуждение с учениками терминологии: «данные», «большие данные», области применения концепции больших данных

Обсудите с учениками понятие «данные»

Что такое «данные». (Выслушайте ответы детей, обсудите).

Где можно встретиться с ними, в чем их измеряют (байты, мегабайты, гигабайты и др.). А сколько их, как вы думаете?

«Сейчас в мире примерно 40 зеттабайтов информации».

В Облаке на 100 Гб может поместиться около 100 фильмов или 10000 фотографий. Зеттабайт это 10 миллиардов по 100 фильмов. *И с каждым днем данных становится все больше и больше.*



Выделяют три основные 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 необходимость высокоскоростной обработки и получения результатов на их основе;
- 3) многообразие (variety) – возможность одновременной обработки различных типов структурированных и полуструктурированных данных.

К первым трем добавляют:

- 4) достоверность (veracity) – данные должны быть репрезентативны и непротиворечивы;
- 5) ценность (value) – данные должны обладать полезностью или потенциальной ценностью (многие данные, собираемые сегодня, не всегда подлежат обработке, большинство просто хранится до нужного времени).

Приведите примеры больших данных.

1. 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 ребенка и рекомендовать им дополнительное обучение.
2. В медицине – роботы могут распознавать симптомы болезней на ранних стадиях и с большей вероятностью рекомендовать правильное лечение.
3. На транспорте – навигаторы анализируют текущую ситуацию на дорогах и выбирают наиболее короткий путь до нужного места, при этом изменяя его в зависимости от ситуации на дорогах.

4. 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.

3. Игра «Обработка данных»

«Пора немного размяться, согласны? Для игры понадобится 5 человек разного роста. Вам нужно молча построиться по росту как можно быстрее. Теперь нужно построиться по датам рождения. На это задание уже ушло побольше времени.»

Только что мы с вами обработали данные. На это ушло 5-7 минут. А если взять 10 человек или всю школу, 150 человек? Уйдет намного больше времени. Мы будем долго заниматься обработкой данных. А компьютер? Быстро, правильно. Именно поэтому мы передаем наши данные, чтобы он обработал их. И более того, еще и получает то, о чем даже не смог предположить вначале. Для простоты вся работа с данными объединена в три этапа – генерацию, предобработку и анализ.

В момент генерации новых данных используются различные устройства, которые фиксируют происходящее вокруг, например, в этом классе – это датчик пожарной безопасности. В вашем телефоне – это GPS который считывает ваше положение и так далее. Фиксируется все, что вы делаете в приложениях, соцсетях, играх и веб-сайтах – это может быть собрано и в будущем проанализировано. Будьте осторожны, не выкладывайте много информации о себе, много мошенников, которые могут воспользоваться вашими данными. Все эти данные сохраняются в больших хранилищах – ЦОДах – Центрах обработки данных. И на сегодняшний день ЦОДы – это одни из самых защищенных мест в любых ИТ-компаниях и государствах.

После этого начинается процесс анализа данных. Дата-аналитики и исследователи (это такая профессия) строят прогнозные модели, проверяют гипотезы на данных и, если все работает, то запускают их в жизнь.

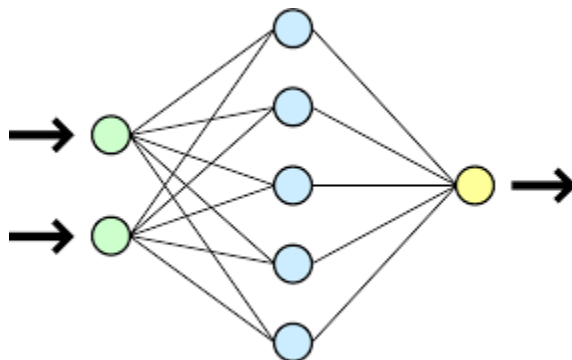
4. Игра «машинное обучение»

Предложите ребятам сыграть в еще одну игру, а затем на ее примере обсудите принцип машинного обучения.

«Давайте попробуем разобраться, почему компьютер такой умный. Кто из вас знает, что такое нейросеть? (Ответы детей). Ученые создали мозг компьютеру как у человека, и назвали его «нейронной сетью». Система обрабатывает входящую информацию через входной нейрон, пропускает через несколько слоев внутри и через выходной нейрон дает нам ответ на поставленную задачу. Точно также обучаемся и мы. Когда мы были маленькими и впервые видели какие-то предметы, то пытались выделить в них какие-то черты. И потом уже сами определяли, что это. Пока мы росли, все больше и больше вещей узнавали, процесс не прекращался. Также обучается и нейросеть».

«Нейросети показывают фото собаки. Далее каждый нейрон из первого слоя (зеленый кружок) определяет характерные черты. И передает информацию последующим слоям для анализа. На выходе последнему нейрону говорят, что данный объект – собака. Система запоминает набор характеристик. Например, она может

выделять определенные области и анализировать наличие тех или иных объектов на фото, соотношение размеров и т.д.



Эту процедуру мы повторяем многократно. Показываем много пород собак, все он запоминает.

А теперь давайте немного поиграем. Я буду предлагать вам разные предметы, а вы мне будете говорить, что для них характерно, и чем отличаются.

5. Дискуссия на тему профессий в области больших данных

Расскажите ученикам про новые профессии в области больших данных

«Кто же создает все эти алгоритмы, нейросети, кто с ними работает? На самом деле — очень много разных специалистов.

«ИТ-медик

Врач с хорошим знанием информационных технологий, который создает, обрабатывает и анализирует физиологические данные и индивидуальные показатели состояния здоровья пациентов и применяет к ним методики работы с большими данными. На основании проведенного анализа он предлагает индивидуальные рекомендации для выздоровления 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.

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 все 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.

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.

6. Рефлексия

Что вы сегодня узнали нового, для чего нужны большие данные, и что полезного можно сделать, используя ее. Захотел ли кто-то из вас стать в будущем ИТ-специалистом.