

**5-18 ноября, 2019**

**МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ**

**по организации и проведению в школах Российской Федерации тематических уроков информатики
в рамках Всероссийской образовательной акции «Урок цифры».**

**Стартовый урок: «Большие данные».**

**(сценарий для проведения урока без Интернета)**

**Москва**

**2019**

**СОДЕРЖАНИЕ**

|  |  |
| --- | --- |
| Цели и задачи урока. План урока. |  3 |
| Анонс занятия. |  4 |
| Обсуждение с учениками терминологии: «данные», «большие данные», области применения концепции больших данных.  |  4 |
| Игра «обработка данных». |  9 |
| Игра «машинное обучение».Дискуссия на тему профессий в области больших данных. |  11  14  |
| Рефлексия. |  16 |
|  |  |
|  |  |
|  |  |

**Цели и задачи урока. План урока.**

**Цель урока:**

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

**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.

**В качестве подготовки к уроку** предлагается изучить данный документ и сформулировать собственный план занятия на основе предложенного.

Предлагаемый План занятия:

|  |  |  |
| --- | --- | --- |
| **Этап** | **Содержание этапа** | **Время этапа** |
| 1. Анонс занятия | – Формулируем задачу на урок. | 5 мин. |
| 2. Обсуждение с учениками новой терминологии | – Обсуждаем термины «данные», «большие данные», области применения концепции больших данных. | 10 мин. |
| 3. Игра «Обработка данных» | – Проводим подвижную игру. – На примере данной игры дискутируем о процессе обработки данных. | 10 мин. |
| 4. Игра «Машинное обучение» | – Проводим игру.– На примере игры рассказываем о машинном обучении.  | 10 мин. |
| 5. Дискуссия на тему «Профессии в области больших данных» | – Обсуждение новых профессий, которые существуют в области больших данных. | 5 мин |
| 6. Рефлексия | – Фиксируем результаты урока. | 5 мин. |

***\*Далее в методических рекомендациях предлагаемый для преподавателя текст выделен курсивом.***

**1. Анонс занятия (5 мин.)**

*«Сегодня у нас пройдет нестандартное занятие в рамках акции «Урок Цифры», которая проводится по всей России. В ее рамках будет проведен цикл необычных уроков информатики, которые пройдут в большинстве школ страны и охватят миллионы учеников с 1-го по 11-й классы. Каждый из них будет посвящен определенной теме: большие данные, сети и облачные технологии, безопасность будущего, персональные помощники, искусственный интеллект. Акция направлена на развитие цифровых компетенций, знаний и навыков, востребованных в современном мире».[[1]](#footnote-1)*

**Подвести ребят к теме «Большие данные».**

*«Мы живем в цифровом мире и каждый день сталкиваемся с большими данными, даже если не подозреваем об этом. Например, пользуясь смартфоном, 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 (10 мин.)**

**Обсудите с учениками понятие «данные»** Задайте ученикам вопрос, как они могут объяснить, что такое «данные». Выслушайте варианты, предложенные ребятами. Обсудите общие характеристики предложенных определений.

Спросите ребят, где они встречаются с разными данными, в чем измеряются данные (байты, мегабайты, гигабайты и др.). Затем предложите ученикам примерно оценить, сколько сейчас существует данных. После нескольких ответов раскройте реальную оценку – зеттабайты данных.

*«Сейчас в мире примерно 40 зеттабайтов информации. А вы можете себе представить, что такое «зеттабайт»?»*

Выслушайте ответы учеников. Для сравнения расскажите, что в Облаке на 100 Гб может поместиться около 100 фильмов в HD или 10000 фотографий, снятых на смартфон. Зеттабайт это 10 миллиардов по 100 фильмов.

Покажите ученикам график со слайда 4. Отметьте неуклонный рост количества информации.



*«Какие данные можно считать большими? Какие данные маленькие и что им нужно сделать, чтобы стать большими?»*

Выслушайте ответы из зала.Обсудите, что четкого разделения между «малыми» и «большими» данными нет. Термину «большие данные» всего 15 лет, и окончательного определения еще не сложилось.  Большими данными называют:

– различные инструменты, подходы и методы обработки как структурированных, так и неструктурированных данных для того, чтобы использовать их для конкретных задач и целей.

– данные, которые невозможно обработать на одном компьютере.

На слайде 6 указаны характеристики больших данных:



Выделяют три основные характеристики больших данных («три V»):

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).



Перейдите к слайду 7 и попросите ребят привести примеры больших данных. Когда кто-то предлагает свой вариант, то вы задаете вопросы на проверку факторам с предыдущего слайда: подтверждается ли объем, скорость, многообразие, достоверность и ценность.

Похвалите ребят за их примеры. В качестве дополнительных примеров можете использовать приведенные ниже.

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. 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.
4. 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.
5. 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.
6. В играх – тренироваться с ботами уже давно стало привычным делом. Программы могут выиграть у человека даже в таких интеллектуальных играх как шахматы или го.
7. 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.

Далее предложите ученикам самим поразмыслить над тем, какие данные они сами генерируют каждый день, и потом эти данные переходят в разряд «больших»? Обсудите предложенные учениками варианты.

На слайде 8 приведены варианты.



**3. Игра «Обработка данных» (10 мин.)**

 **Предложите ребятам сыграть в подвижную игру, и потом на ее примере разберите основные этапы обработки данных.**

*«Пора немного размяться, согласны? Для игры понадобится 5 человек разного роста. Поднимите руку, кто хочет поучаствовать в небольшой игре».*

Помогите ребятам выстроиться в линейку лицом к классу в свободном порядке. Далее дайте задание.

*«Итак, для вас есть задание: вам нужно молча (!) построиться по росту по моей команде, на счет «три». А я засеку время. Итак, вы готовы? Раз. Два. Три!* (Это задание дети должны выполнить очень быстро*). Итак, я вижу, что вы уже готовы*. (Внимательно проверьте и спросите класс – согласны ли они с тем, что те выполнили задание. Все отвечают «да»). *Молодцы, быстро справились: всего 30 секунд!* (Поясните, сколько времени это заняло, после чего дайте второе задание).

*Теперь, для вас задание посложнее. Так сказать, со звездочкой. Вам нужно так же молча (!) построиться по датам рождения, и так же по команде «три» вы должны начать его выполнять, а я засеку, сколько это займет времени на этот раз. Итак, вы готовы? Раз. Два. Три!* (Это задание занимает больше времени). *Кажется, вы готовы.* (Подходите по очереди и спрашивайте полную дату рождения). *Молодцы!* (Даже если дети не справились, надо их похвалить)*. Всего 2 минуты.* (Назовите точное время выполнения задания*).*

*Только что мы смоделировали процесс обработки данных. И мы сделали это за 5 минут. А если бы вас было не 5 человек* (назовите реальное число детей)*, а 105? А в программе Excele не важно, 5 строк отсортировать или 1005 – этот процесс займет доли секунды. При этом программа сможет это сделать даже сразу по 2 факторам – рост и дата – одновременно. Именно поэтому человек осознанно передает на машину такие сложные задачи по расчетам разных действий и процессов. И более того, еще и получает то, о чем даже не смог предположить вначале. Давайте узнаем, как это происходит?»*

Перейдите к следующему слайду. Объясните процесс работы с данными согласно слайду 12.

**

*«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 вашу геолокацию вплоть до полуметра, гироскоп и акселерометр определяют положение вашего телефона в пространстве и так далее. Фиксируется все, что вы делаете в приложениях, соцсетях (мы говорили уже про сообщения, лайки и т. п.), играх и веб-сайтах –это может быть собрано и в будущем проанализировано.*

*Данные могут быть получены сразу большим объемом от других организаций. Так, нередко многие магазины делятся своими данными между собой.*

*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.*

*Далее происходит первичная обработка данных, так как их надо структурировать, очистить от ненужных или являющихся выбросами, которые могут смазать картинку. Данные могут между собой соединяться, обогащаться за счет друг друга. Конечно же, для этого необходима специальная инфраструктура (базы данных для умеренно большого объема данных и MapReduce для настоящей Big Data), которая требует поддержки со стороны дата-инженеров.*

*После этого начинается процесс анализа данных. Дата-аналитики и исследователи строят прогнозные модели, проверяют гипотезы на данных и, если все работает, то запускают их в жизнь. Для этого они используют методы визуализации данных, статистической проверки гипотез и технологии машинного обучения, в частности нейросети.*

*Сегодня не всегда понятно, как обрабатывать многие данные, еще не придуманы все алгоритмы на свете, потому многое пока просто хранится. Это можно сравнить с айсбергом – на поверхности мы видим только 1/9 часть применения данных, а через годы станет ясно, как использовать и оставшуюся часть, что может значительно изменить нашу жизнь. 15 лет назад еще не было специалистов по машинному обучению, как и специалистов по мобильной разработке, а сегодня и тех, и других требуются десятки и сотни тысяч. И это, кстати, очень увлекательная тема – научить машину думать и делать то, что ты хочешь».*

**4. Игра «машинное обучение» (10 мин.)**

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

*«Давайте попробуем разобраться, как машина учится. Кто из вас знает, что такое нейросеть?* (Ответы детей). *Ученые изучали работу мозга, и потом, по аналогии создали систему, которая позволяет машине обучаться, и назвали ее «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 те или иные объекты, то пытались выделить в них общие и отличительные черты. Раз за разом, мы запоминали те черты, которые характерны, например, для кошки. И потом уже сами стали определять без подсказок, что данное животное – это кошка. Этот процесс никогда не прекращается. Мы сталкиваемся с новыми объектами и учимся их распознавать. Аналогичным образом обучается и нейросеть».*

Продемонстрируйте слайд 12.

**

*«Нейросети показывают фото кошки. Далее каждый нейрон из первого слоя (квадратики) определяет характерные черты. И передает информацию последующим слоям для анализа. На выходе последнему нейрону говорят, что данный объект – это кошка. Система запоминает набор характеристик. Например, она может выделять определенные области и анализировать наличие тех или иных объектов на фото, соотношение размеров и т.д.* (слайд 13).



*Эту процедуру мы повторяем многократно* (слайд 14).

**

*И тогда система запоминает, что есть обязательные черты, которые есть у всех кошек, а есть менее значимые, которые могут отличаться у кошек разной породы. На фото мы видим, например, что кошки персидской породы отличаются формой носа.*

*Как вы думаете, если при обучении не предъявить системе фото с персидскими кошками, а потом показать такую кошку, то сможет ли система опознать ее?* (Выслушайте ответы*). На самом деле, может и не узнать, поскольку запомнит, что для всех кошек характерно определенное строение носа.*

*А теперь давайте немного поиграем* (слайд 15). *Я буду предлагать вам разные объекты, а вы мне будете говорить, что для них характерно, и чем отличаются.*

*Что это за объект и какие у него есть характерные (важные) черты?* (Слайд 16). *А так?* (По щелчку мыши покажите следующий стол). *Что осталось и что поменялось? А так?* (Следующее изображение).

*Хорошо.* (Аналогично обсудите слайд 17).

*А как теперь нейросети различить стол и стул? Что у них будет одинаковым? И что разным?»*

Покажите картинку со стулом, у которого спинка имеет форму головы кошки.

*«Как вы думаете, если наша нейросеть знает, что такое кошки, стулья и столы, то, что она нам ответит? Вопрос вообще-то интересный, здесь могут быть разные варианты ответов. Если нейросеть видела только настоящих кошек, то скорее склонится в сторону стула. Ответ – «нужно пробовать».*

*Такой способ обучения называется «обучение с учителем». Но есть и другие варианты.*

*Ну что, стало понятнее? Теперь вы знаете, что такое нейросеть и сможете рассказать об этом своим родителям и друзьям, которые не смогли быть на этом уроке. А мы пойдем дальше. (Переходите на следующий слайд)».*

**5. Дискуссия на тему профессий в области больших данных (10 мин.)**

**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. На слайде – лишь представлены лишь некоторые из них. Давайте попробуем разобраться, кто есть кто и чем занимается. Как вы думаете, какие из них существуют уже сейчас, а какие появятся в ближайшем будущем? А как вы думаете, чем занимаются остальные люди на картинках?»*

Примерный текст для рассказа про некоторые новые профессии:

***«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.*

***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 объемы данных, в которых должен разбираться BIM-инженер.*

***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.*

***Антифейковый менеджер***

*Этот специалист создает 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.*

**Профориентационный тест**

**(Если у учеников есть возможность пройти со своих мобильных телефонов или из дома).**

В тесте 12 вопросов. Можете предложить ребятам пройти тест после урока, и потом поделиться результатами в социальных сетях.

*«Хотите узнать на кого из них вы похожи больше всего?*

*Предлагаю вам сейчас пройти интересный тест – ответить на несколько вопросов и определить, какая профессия, связанная с большими данными, вам подходит больше всего. Для этого считайте QR-код, это легко сделать с помощью камеры ВКонтакте. Далее начинайте отвечать на вопросы и делитесь своими результатами с друзьями».*

1. **Рефлексия**

Спросите у учеников что они сегодня узнали нового, для чего нужны большие данные, и что полезного можно сделать, используя эту концепцию. Поинтересуйтесь, захотел ли кто-то стать в будущем ИТ-специалистом.

1. Здесь и далее в методических рекомендациях предлагаемая прямая речь преподавателя выделена курсивом. [↑](#footnote-ref-1)