



7 класс · Урок 24 из 34 · 45 минут

Что такое данные и датасет?

3.8. Мир данных

Результат: составляет паспорт набора данных, различает объект, признак и метку; проводит короткий обезличенный опрос и записывает результат таблицей.

Листы ученика

01. Проблемная ситуация

В таблице для выбора школьного кружка есть числа 10, 15, 20. Помощник объявил: «Это возраст участников». Но заголовка и единиц нет. Какие ещё значения могли иметь эти числа? Почему таблица без описания может дать неверный вывод?

0-5 минут: минуту прочитайте ситуацию, за две запишите своё решение и основание, за две сравните с соседом и сформулируйте вопрос проверки.

Место для ответа



02. Объяснение и разобранный пример

5-14 минут: 4 минуты - понятия; 3 - совместный разбор; 2 - проверка ограничений примера.

Данные - записанные сведения: измерения, ответы, тексты, изображения или звук. Они становятся полезны в конкретной задаче при известном происхождении и смысле. Знания человека включают опыт и объяснения; файл с числами сам по себе не равен пониманию. **Датасет** - организованный набор примеров, предназначенный для исследования или работы модели. В таблице строка обычно описывает объект или наблюдение, столбец - его признак. Важно заранее определить, что считается одним объектом.

Признак - характеристика, доступная для решения задачи. **Метка** - известный ответ для обучения или проверки, например «подходит объявлению». Не любая колонка является меткой: номер записи лишь помогает её найти. Для прогноза результата нельзя подать модели будущий правильный ответ как уже известный признак. Табличный набор может содержать категории и числа; набор изображений - файлы с описанием, текстовый - документы, аудио - записи. Состав и правила связи с метками нужно документировать.

Паспорт датасета отвечает: задача, объект строки, смысл и единицы полей, источник, способ сбора, пропуски, ограничения и разрешённое использование. **Большие данные** связаны не только с большим количеством строк, но и со скоростью поступления, разнообразием и сложностью обработки. Миллионы повторов одной ошибки не превращаются в надёжные знания. Наш набор из четырёх строк маленький; мы учимся описывать его, а не заявляем работу с big data.

Разобранный пример. Для выбора формата встречи строка означает один вымышленный ответ, поля: код, интерес, доступное время в минутах. Запись «А, техника, 15» не описывает возраст. Если задача - прогнозировать, понравилась ли встреча после её посещения, будущая оценка не должна попадать во входные данные при прогнозе до встречи.



03. Практика 1: Опишите набор и проведите мини-опрос

Код	Интерес	Время, мин
А	Природа	15
Б	Техника	20
В	Природа	10
Г	Искусство	15

Таблица - резервные вымышленные ответы. Основной вариант: за 3 минуты группа задаёт одноклассникам два вопроса: «Какую учебную тему выберешь: природа, техника, искусство?» и «Какой формат удобнее: 10, 15 или 20 минут?» Участие добровольное. Не спрашивайте имя, оценки, контакты. Записывайте ответы без возможности связать строку с человеком; на публику выводите только число ответов по категориям. При отказе или отсутствии группы используйте резерв и укажите это.

- За 2 минуты определите задачу и единицу наблюдения; подготовьте поля без личных сведений.
- За 3 минуты проведите опрос или обработайте резерв; честно укажите происхождение строк.
- За 4 минуты составьте паспорт и сводку частот интересов; отделите код от признаков.
- За 2 минуты объясните ограничение выборки и представьте общий результат без индивидуальных записей.

Форма результата: Паспорт датасета, рабочая таблица и агрегированная сводка.

Критерии: По 1 баллу: верные поля и единицы; честный источник; сводка без личных сведений; ограничение выводов. Всего 4.

Место для ответа



04. Практика 2: Подберите вид данных под задачу

Задача	Предложенный материал
Найти повреждённый угол книги	Фотографии обложек
Распознать слова объявления	Звуковые записи и проверенные расшифровки
Предсказать длительность выдачи	Таблица числа посетителей и времени
Определить тему новости	Тексты новостей с тематическими метками

Условный помощник утверждает: «Любой датасет - таблица, любой номер - ответ, а данные опроса доказывают вкусы всех детей России». Это учебная реплика.

- За 2 минуты назовите объекты и виды данных четырёх задач.
- За 3 минуты укажите, где нужны эталонные ответы и какие они.
- За 2 минуты исправьте три утверждения помощника.
- За минуту предложите поле паспорта, без которого один из наборов опасно неправильно истолковать.

Форма результата: Четыре карточки «задача → данные → ответ → паспорт».

Критерии: По 1 баллу за виды, эталоны, исправления и нужное описание. Всего 4.

Место для ответа



05. Обсуждение: аргумент и возражение

33-40 минут. Тезис: «Чем больше строк, тем больше знаний». За 2 минуты назовите пользу разнообразных примеров; за 3 возразите тысячей копий строки без единиц; за 2 согласуйте условия полезного расширения.

Место для ответа

06. Индивидуальная проверка понимания

В журнале одна строка: «датчик С, 20, 08:00». Нужно сравнить температуры двух кабинетов. Назовите два недостающих поля, объект записи, подходящий тип набора и границу вывода.

40-45 минут: 3 минуты - самостоятельная запись; 1 - сверка по ключу и исправление другим цветом; 1 - сдача ответа и общий вывод.

Место для ответа



Для учителя: ответы и критерии

7 класс · Урок 24. Что такое данные и датасет?

Подготовьте: экран или распечатки, тетради, ручки; калькулятор по необходимости. Все таблицы, тексты и ключи находятся в статье и PDF. Ключи открывайте после самостоятельной попытки; ученические листы печатайте отдельно от раздела учителя.

Вариант без регистрации: работайте по готовым материалам и учебным журналам на экране или бумаге. Ученики не создают аккаунтов и не передают личных данных. Учебная имитация позволяет проверить рассуждение, но не заменяет настоящий эксперимент с сервисом.

Время	Этап	Результат
0-5	Проблемная ситуация	Первое решение и основания
5-14	Теория и разобранный пример	Понятия, доказательства и ограничения
14-25	Практика 1	Проверяемый результат по критериям
25-33	Практика 2	Перенос в новые условия
33-40	Обсуждение	Аргумент, возражение, общее решение
40-45	Индивидуальная проверка	Самостоятельный ответ и сверка

Что собрать: результаты двух практик, основания вывода и индивидуальную проверку. Минуты включают чтение, обсуждение и запись; раздел соответствия программе предназначен учителю вне этих 45 минут. При 0-1 балле в проверке вернитесь к разобранным примерам; при 2-3 уточните пропущенное звено; при 4 предложите новый случай. Критерии нужны для обратной связи, не задают обязательную школьную отметку.

01-02. Разбор проблемной ситуации и примера

Разобранный пример. Для выбора формата встречи строка означает один вымышленный ответ, поля: код, интерес, доступное время в минутах. Запись «А, техника, 15» не описывает возраст. Если задача - прогнозировать, понравилась ли встреча после её посещения, будущая оценка не должна попадать во входные данные при прогнозе до встречи.

03. Опишите набор и проведите мини-опрос

В резерве 4 ответа: природа - 2, техника - 1, искусство - 1. Объект - один ответ; время измерено в минутах; код - идентификатор учебной записи, не возраст и не метка. Источник - вымышленный резерв, если опрос не состоялся. Для настоящего опроса числа могут быть иными, оценивается процедура и верная сводка. Нельзя объявлять предпочтения четырёх человек предпочтениями всей школы. Даже без имени редкий ответ может быть узнаваем; не публикуйте индивидуальные строки и не сохраняйте соответствие с участниками.

Критерии: По 1 баллу: верные поля и единицы; честный источник; сводка без личных сведений; ограничение выводов. Всего 4.



04. Подберите вид данных под задачу

Фотография - изображение, эталон «повреждение есть/нет» или отмеченная область; запись - аудио, эталон точный текст; наблюдение выдачи - таблица, ответ длительность; новость - текст, ответ тема. Номер не раскрывает нужный ответ автоматически. Наборы бывают не только табличными. Опрос небольшой группы не описывает всю страну. Полезные поля: способ съёмки, язык записи, единица времени, инструкция тем; возможны другие обоснованные варианты.

Критерии: По 1 баллу за виды, эталоны, исправления и нужное описание. Всего 4.

05. Обсуждение: ожидаемый результат

33-40 минут. Тезис: «Чем больше строк, тем больше знаний». За 2 минуты назовите пользу разнообразных примеров; за 3 возразите тысячей копий строки без единиц; за 2 согласуйте условия полезного расширения.

Нужны разные уместные наблюдения, верные подписи и проверка. Объём не исправляет неверное происхождение. Человеческий смысл задачи и опыт не сводятся к размеру файла.

06. Индивидуальная проверка: ключ

Не хватает единицы температуры и места/связи датчика с кабинетом; также полезны дата и условия измерения. Объект - измерение датчика в момент времени. Тип - таблица временных наблюдений. По одной неописанной строке сравнить кабинеты нельзя. По 1 баллу за недостающее, объект, тип, ограничение.

07. Связь с программой

Пункт программы	Действие	Проверяемый результат
Данные, датасет, виды и структура	Описывает строки и поля четырёх видов наборов	Отличает код, признак и метку
Создавать датасет; групповой опрос	Собирает добровольные ответы или использует отмеченный резерв	Представляет таблицу и агрегированную сводку
Большие данные; данные и знания	Объясняет пользу и предел объёма	Не выдаёт маленькую выборку за данные всей школы

Источники и права

Основа содержания и результатов - [исходная программа класса](#). Поурочное название и учебные случаи - авторская методическая развёртка, не дословные названия модулей программы. Числовые наборы, реплики условных помощников и ситуации специально подготовлены для урока. Исторические карточки пересказывают указанные документы; это не дословные цитаты. Учебные журналы не являются результатами испытания реального ИИ.

[Google: задачи и виды машинного обучения](#) - источник для понятий и проверки; учебные задания разработаны отдельно.

Логотип «Школы России» используется в материалах портала; авторские схемы предназначены для учебной работы. Внешние материалы сохраняют указанные условия использования.