**МЕТОДИКА ИЗУЧЕНИЯ ПОИСКА АНОМАЛИЙ В 8-9 КЛАССЕ**

**Ковалева Н.А. (**[**natakovaleva2010@gmail.com**](mailto:natakovaleva2010@gmail.com)**)**

*Муниципальное учреждение дополнительного образования Центр дополнительного образования «Истоки» Павлово-Посадского городского округа Московской области, г. Электрогорск*

**Аннотация**

В работе рассмотрена методика проведения учебного занятия по изучению инструментов искусственного интеллекта. На примере темы «Поиск аномалий» автором предлагается освоить основные приемы работы с библиотеками Python. Предложенная методика позволит сформировать интерес школьников к практическому применению анализа данных и аномалий в данных.

Одним из самых востребованных навыков будущего, станет высокий уровень владения инструментами искусственного интеллекта (далее ИИ). Уже сегодня есть ряд образовательных программ и курсов, направленных на полученные знаний и практических навыков в этой области. Большое внимание уделяется раннему знакомству с технологией ИИ. Так, например, курс на платформе Stepik «Курс по машинному обучению. Проект «ИИ Старт» был специально создан для школьников старших классов и всех, тех кто стремится получить профессию в Data Science. Все желающие могут получить компетенции в области ИИ, изучив этот курс. Несмотря на то, что составители курса приводят много практических примеров у педагогов всегда есть желание адаптировать некоторые задания под уровень знания группы или возможности конкретного обучающегося.

Рассмотрим на примере темы: «Поиск аномалий» методику построения занятия. Тип занятия – объяснение нового материала. Занятие имеет стандартную структуру и включает вводную беседу, знакомство с новым материалом, практическую работу и рефлексию.

Для проведения занятия понадобится оборудование и материалы: компьютеры, подключенные к сети Интернет, проектор, опорная презентация, файл с описанием задания и инструкцией, файл «Аномалии.ipynb» с рабочим кодом для практической работы.

Учебные задачи занятия:

* познакомить обучающихся на простых примерах с понятиями «Аномалия», «Аномалии в реальном мире», «Аномалии в данных», «Выброс»;
* использовать методы библиотек Python для построения графиков и диаграмм в среде Google CoLab;
* делать выводы о наличие или об отсутствии аномалий в данных на основе графиков и диаграмм.

Опорная презентация должна содержать повторение ранее изученного материала, например, работу с библиотеками, фрагментами кода и теоретический материал по новой теме. В рамках одного занятия достаточно сложно охватить весь теоретический материал, поэтому следует выбрать понятные для подростков примеры из окружающего мира, например, «Поющие пески» у деревни Атары, Озеро Баскунчак, в живой природе – альбинизм. Почему же аномалии так важны? Обнаружение аномалий или обнаружение выбросов – это идентификация редких элементов, событий или наблюдений, вызывающих подозрения, поскольку они значительно отличаются от большинства данных. Как правило, аномальные данные могут быть связаны с какой-либо проблемой или редким событием. Метод нахождения аномалий широко применяется в цифровом и реальном мире при обнаружении: неполадок в механизмах по показаниям датчиков, вторжений, подозрительных банковских операций, нестандартных игроков на бирже, этим методом пользуется медицинская диагностика, сейсмология и др.

Основные поисковые вопросы для обучения на занятии: «Из-за чего появляются аномалии?», «Когда их надо учитывать при разработке моделей машинного обучения?», «Как их обнаружить?».

Файл с описанием задания и инструкцией должен содержать подробную информацию о действиях, которые выполняют обучающиеся. Практическая работа выполняется на примере «Зарплата сотрудников». Цель практической работы:на основе ранее сгенерированных данных выявить аномалии по признаку «Зарплата».Построить графики и на их основе сделать выводы.

Синтетический набор данных содержит два столбца: Работник (для 100 человек) и Зарплата (в руб.) в диапазоне от 18000 до 30000.

Предварительно нужно загрузить файл «Аномалии.ipynb» с рабочим кодом в Colab.

Проверим, подгружены ли необходимые для работы библиотеки (Pandas, NumPy, Matplotlib, Faker), сгенерированы ли тестовые данные. Визуально оценим набор данных и посмотрим, можно ли найти аномалии. На рисунке 1 представлен фрагмент таблицы с данными о сотрудниках и их заработной плате. Вручную изменим зарплату нескольких человек, чтобы создать выбросы. В реальности это может произойти по ряду причин, например, из-за невнимательности бухгалтера или сбоя программного обеспечения.

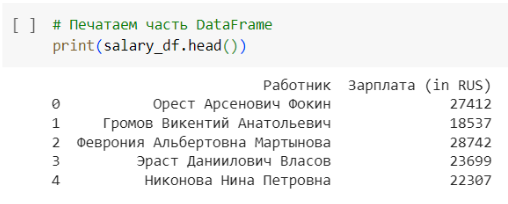


Рис. 1. Фрагмент таблицы с данными

**Проведение экспериментов по поиску аномалий.**

Построим диаграмму размаха или ящик с усами – это быстрый способ изучения одного или нескольких наборов данных в графическом виде. График в удобной форме показывает медиану, нижний и верхний квартили, минимальное и максимальное значение выборки и выбросы. На рисунке 2 представлен результат.

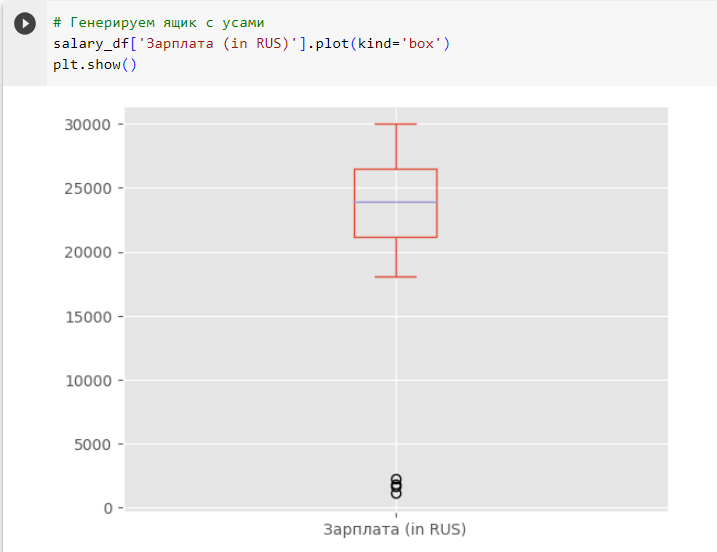


Рис. 2. Диаграмма размаха

Следует обратить внимание на маленькие кружочки в самом низу. Они сразу дают понять, что что-то не так, потому что сильно отличается от остальных данных. Это и есть аномалии в данных. Это выбросы, которые специалист должен увидеть и изучить причину их возникновения. Далее можно построить и другие виды диаграмм, например, гистограмму. Подтвердит наличие аномалий в наборе данных смогут функции, выводящие минимальное и максимальное значение столбца «Зарплата» рисунок 3. Минимальное значение 1112 рублей явно отклоняется от того, что было задано раньше 18000 рублей. Следовательно, это действительно аномалия.

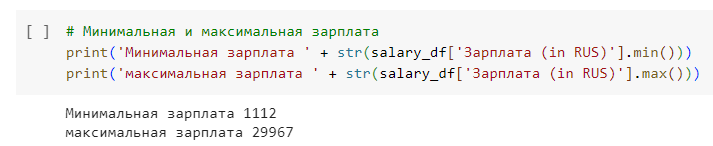


Рис. 3. Использование функций min(), max()

Наш набор данных содержит только один признак (зарплату), в настоящих датасетах аномалии могут встречаться в разных признаках, но даже там такие визуализации помогут их обнаружить.

После выполнения работы обучающиеся будут знать: этапы решения задач по выявлению аномалий в данных, методы построения графиков и диаграмм в среде Colab; будут уметь: работать с библиотеками Python, находить аномалии в данных.

Интеллектуальный анализ данных используется для выделения новой значимой информации из большого объема данных. Учебные занятия, выстроенные по предложенной методике, могут сформировать интерес и привлечь школьников в сферу искусственного интеллекта, помогут показать практическое применение полученным знаниям.

**Литература**

1. 9 подходов для выявления аномалий URL: <https://habr.com/ru/articles/477450/> (Дата обращения 15.05.2024)
2. Курс по машинному обучению. Проект «ИИ Старт». URL: <https://stepik.org/course/125587/promo> (Дата обращения 15.05.2024)
3. Методы обнаружения аномалий в Python URL: <https://evogeek.ru/articles/350871/> (Дата обращения 18.05.2024)