

Контрольные и тестовые вопросы по ПР14

«Анализ и обнаружение аномалий в IoT-данных»

по вариантам с ответами

Вариант 1

1. Почему Isolation Forest эффективно применяется для выявления аномалий в высокоразмерных данных IoT?

- A) Не требует предварительного анализа распределения признаков
- B) Работает только с бинарными данными
- C) Не чувствителен к шуму
- D) Всегда показывает точные результаты

Ответ: А

2. На какой основе метод Isolation Forest выявляет аномалии?

- A) Расстояние между объектами
- B) Частота появления значений
- C) Глубина размещения точек в деревьях
- D) Средние значения признаков

Ответ: С

3. Какая комбинация boolean-признаков IoT-данных является аномальной ночью?

- A) Свет выключен, движения нет
- B) Свет включен, движения нет
- C) Свет выключен, движение есть
- D) Свет включен, движение есть

Ответ: С

4. Что произойдет при завышении параметра contamination в Generate Outlier Flag?

- A) Повысится точность выявления аномалий
- B) Увеличится число ложных аномалий
- C) Снизится чувствительность алгоритма
- D) Данные не будут обработаны

Ответ: В

5. Какая визуализация показывает распределение комбинаций значений датчиков?

- A) Scatter Plot
- B) Histogram
- C) Line Plot
- D) Bar Chart

Ответ: D

6. Какое назначение у параметра use feature heuristic в Isolation Forest?

- A) Управляет глубиной деревьев
- B) Устанавливает число деревьев
- C) Использует случайный набор признаков для каждого дерева
- D) Регулирует долю выборки

Ответ: C

7. Какой график лучше использовать для выявления периодических аномалий?

- A) Pie Chart
- B) Line Plot
- C) Histogram
- D) Box Plot

Ответ: B

8. Назовите два типа экстремальных аномалий по концентрации газа CO и объясните причины их возникновения.

Ответ:

- Положительные экстремумы (высокая концентрация CO): возможна утечка газа.
- Отрицательные экстремумы (низкая концентрация CO): интенсивная вентиляция или неисправность датчика.

9. Почему при детекции аномалий в IoT важно учитывать комбинации признаков, а не только отдельные значения?

Ответ: Потому что некоторые аномалии проявляются только при специфических комбинациях параметров, которые по отдельности могут выглядеть нормальными.

10. Какие действия следует предпринять после обнаружения потенциальной аномалии в IoT-данных?

Ответ: Проверить данные вручную, выяснить причины аномалии и принять меры для предотвращения повторения.

Вариант 2

1. В каком случае нормализация данных особенно критична перед использованием Isolation Forest?

- A) При работе с малым объемом данных
- B) При использовании бинарных признаков
- C) При наличии числовых признаков с разными шкалами измерений
- D) При частых обновлениях данных

Ответ: C

2. Как параметр Max leaf size влияет на результат Isolation Forest?

- A) Определяет число деревьев
- B) Влияет на чувствительность алгоритма к небольшим аномалиям
- C) Регулирует число ложных срабатываний
- D) Устанавливает минимальный порог аномалии

Ответ: B

3. Что позволяет выявить Scatter Plot при анализе IoT-данных?

- A) Частоту аномалий
- B) Взаимодействие двух признаков и аномальные комбинации
- C) Временные тренды
- D) Общие статистические параметры

Ответ: B

4. Что означает Outlier Score в Isolation Forest?

- A) Абсолютное значение аномалии
- B) Вероятность обнаружения аномалии
- C) Степень аномальности точки по глубине дерева
- D) Среднее значение признаков

Ответ: C

5. Почему булевы признаки важны в анализе IoT-данных на аномалии?

- A) Они чаще всего ломаются
- B) Они позволяют выявить неявные аномалии в комбинациях признаков

- C) Ускоряют вычисления
- D) Стабильнее числовых показателей

Ответ: B

6. Какой недостаток характерен для Isolation Forest?

- A) Медленная обработка данных
- B) Необходимость ручного выбора числа кластеров
- C) Чувствительность к параметрам настройки
- D) Неспособность обработки временных данных

Ответ: C

7. Какой тип визуализации подходит для анализа распределения одного признака на аномалии?

- A) Line Plot
- B) Histogram
- C) Scatter Plot
- D) Bar Chart

Ответ: B

8. Какие признаки в IoT-данных требуют особого внимания при выявлении аномалий?

Ответ: Показатели газа (CO, LPG, Smoke), температура и влажность, поскольку их экстремальные значения могут указывать на опасные ситуации.

9. В чем преимущество использования нескольких видов визуализации при анализе аномалий в IoT-данных?

Ответ: Обеспечивает комплексный анализ, позволяя выявить разные типы и причины аномалий.

10. Каким образом ошибки датчиков могут повлиять на результаты анализа аномалий?

Ответ: Ошибки датчиков могут привести к ложным аномалиям или пропускам реальных нештатных ситуаций, снижая качество анализа.

Вариант 3

1. Какой подход позволяет наиболее эффективно обнаруживать аномалии в реальном времени на IoT-устройствах?
A) Пакетная обработка данных
B) Поточковая обработка данных (streaming analytics)
C) Использование статистических таблиц
D) Ручная проверка данных

Ответ: B

2. Какой параметр Isolation Forest управляет количеством объектов, которые попадут в конечный лист дерева?
A) Number of trees
B) Max leaf size
C) Bootstrap ratio
D) Contamination threshold

Ответ: B

3. Почему данные времени (epoch) необходимо преобразовывать в datetime перед анализом аномалий?
A) Ускорение анализа данных
B) Улучшение визуализации и интерпретации временных трендов
C) Уменьшение размера данных
D) Снижение числа ложных аномалий

Ответ: B

4. Каким образом низкая влажность и высокие показатели газа могут быть интерпретированы?
A) Неисправность датчика влажности
B) Вероятность пожара или возгорания
C) Неисправность системы вентиляции
D) Обычная рабочая ситуация

Ответ: B

5. Какая комбинация параметров Isolation Forest позволит наиболее детально выявлять мелкие аномалии?
A) Меньшее количество деревьев, больше Max leaf size
B) Больше количество деревьев, меньше Max leaf size
C) Низкий bootstrap ratio, больше Max leaf size
D) Высокий contamination threshold

Ответ: В

6. Какой тип визуализации лучше всего демонстрирует взаимосвязь двух числовых признаков и аномалий?
- A) Histogram
 - B) Scatter Plot
 - C) Line Plot
 - D) Bar Chart

Ответ: В

7. Какая потенциальная проблема может возникнуть при слишком малом значении contamination threshold?
- A) Слишком много ложных аномалий
 - B) Аномалии не будут обнаруживаться
 - C) Увеличение времени обработки
 - D) Невозможность обработки данных

Ответ: В

8. Какие шаги следует выполнить для предобработки IoT-данных перед использованием алгоритма Isolation Forest?

Ответ: Очистка от пропусков, нормализация признаков и преобразование временных меток из epoch в datetime.

9. Как интерпретировать аномалии, если показатели газа стабильны, но изменяется температура?

Ответ: Возможны проблемы с терморегуляцией помещения или сбой в работе датчика температуры.

10. Какие критерии важно учитывать при настройке алгоритмов выявления аномалий?

Ответ: Чувствительность алгоритма, баланс между ложными и истинными аномалиями, время обработки данных.

Вариант 4

1. Что дает использование случайного подмножества признаков (feature heuristic) при построении деревьев в Isolation Forest?

- A) Увеличивает скорость обработки
- B) Снижает риск переобучения

- C) Увеличивает число ложных аномалий
- D) Уменьшает детализацию

Ответ: В

2. Почему сочетание стабильных значений температуры и газа может быть аномальным?
- A) Это всегда указывает на пожар
 - B) Возможен сбой или «зависание» датчиков
 - C) Это невозможно в реальных условиях
 - D) Это ошибка программного обеспечения

Ответ: В

3. Какой график наиболее эффективен для оценки распределения числового признака на предмет наличия экстремальных значений?
- A) Bar Chart
 - B) Line Plot
 - C) Scatter Plot
 - D) Histogram

Ответ: D

4. В каких случаях высокая концентрация LPG может быть выявлена как аномалия?
- A) При одновременном снижении температуры
 - B) При стабильных показаниях других датчиков
 - C) При внезапном всплеске без видимых причин
 - D) При нормальных колебаниях температуры

Ответ: C

5. Какое последствие будет иметь неправильно выбранный параметр bootstrap ratio в Isolation Forest?
- A) Невозможность обнаружения аномалий
 - B) Потеря важных признаков данных при анализе
 - C) Замедление работы алгоритма
 - D) Снижение точности и стабильности результатов

Ответ: D

6. Какой показатель может косвенно указывать на проблемы с вентиляцией в помещении?
- A) Высокая влажность и низкий уровень газа CO

- В) Стабильно низкая влажность
- С) Резкое снижение концентрации газа CO
- Д) Резкое повышение температуры

Ответ: А

7. Как лучше всего подтвердить, что выявленная аномалия является реальной?
- А) Провести повторный анализ
 - В) Использовать дополнительные источники данных и ручную проверку
 - С) Игнорировать единичные случаи
 - Д) Увеличить contamination threshold

Ответ: В

8. Какие особенности IoT-данных усложняют задачу обнаружения аномалий?

Ответ: Высокая размерность данных, шумы в датчиках, динамичность изменений и необходимость быстрой обработки.

9. Почему сочетание boolean-признаков (например, motion и light) требует отдельного анализа при выявлении аномалий?

Ответ: Некоторые аномалии могут проявляться исключительно в специфических комбинациях бинарных признаков и не быть видны через числовые показатели.

10. Что является причиной ложноположительных срабатываний алгоритмов выявления аномалий?

Ответ: Недостаточная предварительная обработка данных, неправильная настройка алгоритма или временные сбои в работе датчиков.