

# **Контрольные и тестовые вопросы по ПР4**

## **«Анализ данных с использованием ассоциативных правил в RM» по вариантам с ответами**

### **Вариант 1**

1. Какой алгоритм наиболее эффективен для поиска частых наборов элементов без генерации кандидатов?  
A) Apriori  
B) FP-Growth  
C) ECLAT  
D) CART

**Ответ: B**

2. Что означает метрика confidence в ассоциативных правилах?  
A) Частота появления набора в выборке  
B) Вероятность появления всего набора  
C) Вероятность появления правой части при наличии левой  
D) Количество элементов в правой части

**Ответ: C**

3. Какой параметр управления FP-Growth отвечает за игнорирование малозначимых наборов?  
A) Min items per itemset  
B) Max number of itemsets  
C) Min support  
D) Trim item names

**Ответ: C**

4. Что произойдет при установке слишком высокого порога поддержки (support) в FP-Growth?  
A) Увеличится число выявленных наборов  
B) Модель станет точнее  
C) Будут выявлены только самые часто встречающиеся наборы  
D) Увеличится время вычислений

**Ответ: С**

5. Что измеряет метрика Lift в ассоциативных правилах?

- А) Среднюю поддержку
- В) Насколько вероятнее следствие при наличии предпосылки, чем в среднем по выборке
- С) Число итераций в FP-Growth
- Д) Размерность набора

**Ответ: В**

6. Какое преимущество визуализации ассоциативных правил через граф?

- А) Ускорение вычислений
- В) Повышение точности
- С) Упрощение интерпретации связей между элементами
- Д) Замена статистического анализа

**Ответ: С**

7. Что произойдет, если в поле “Item separators” в RapidMiner указан неверный символ?

- А) Алгоритм автоматически определит нужный
- В) Все товары будут считаться как один элемент
- С) Увеличится поддержка правил
- Д) Модель перестанет обучаться

**Ответ: В**

8. Почему важно обрезать пробелы в названиях товаров перед применением FP-Growth?

**Ответ:** Пробелы могут создавать формально разные, но по сути одинаковые элементы, что приводит к снижению поддержки и искажению результатов.

9. Как параметр confidence влияет на итоговое количество правил?

**Ответ:** Чем выше установленный порог confidence, тем меньше правил будет сформировано, так как отбрасываются менее достоверные связи.

10. В каких практических задачах применяются ассоциативные правила, кроме анализа покупок?

**Ответ:** Рекомендательные системы, выявление мошенничества, биоинформатика, анализ клинических данных, оптимизация производственных процессов.

---

## Вариант 2

1. Что из нижеперечисленного верно описывает алгоритм Apriori?

- A) Строит префиксное дерево
- B) Использует список транзакций
- C) Постепенно расширяет частые наборы
- D) Основан на энтропии

**Ответ: C**

2. Что означает поддержка (support) ассоциативного правила  $A \rightarrow B$ ?

- A) Общее количество покупок A
- B) Общее количество покупок B
- C) Количество транзакций, содержащих A и B
- D) Уверенность в предсказании

**Ответ: C**

3. В чем ключевое отличие FP-Growth от Apriori?

- A) Использование деревьев вместо генерации кандидатов
- B) Более высокая чувствительность
- C) Возможность работать с непрерывными данными
- D) Построение SVM-моделей

**Ответ: A**

4. Какой параметр отвечает за фильтрацию редких наборов товаров?

- A) Min items per itemset
- B) Item separators
- C) Max depth
- D) Min support

**Ответ: D**

5. Какое значение параметра min confidence в RapidMiner означает, что только очень надёжные правила будут включены?

- A) 0.2
- B) 0.5
- C) 0.75
- D) 0.95

**Ответ: D**

6. Какое ограничение накладывает установка “Max number of itemsets = 100”?
- A) Отображается максимум 100 правил
  - B) Модель использует только 100 итераций
  - C) Не более 100 уникальных товаров
  - D) Учитываются только 100 частых наборов

**Ответ: D**

7. Почему важно объединять наименования товаров из одной транзакции перед запуском FP-Growth?
- A) Для ускорения сортировки
  - B) Для построения матрицы взаимной информации
  - C) Чтобы каждую транзакцию представить как множество
  - D) Для удаления пропусков

**Ответ: C**

8. Чем отличается confidence от support?

**Ответ:** Support показывает частоту набора в общей выборке, а confidence — условную вероятность появления следствия при наличии предпосылки.

9. Какие ошибки может вызвать отсутствие тримминга строк в данных?

**Ответ:** Повторяющиеся названия товаров с пробелами не будут объединены, что приведет к заниженной частоте и разрозненным правилам.

10. Почему ассоциативный анализ не применим к непрерывным числовым данным без преобразования?

**Ответ:** Алгоритмы, такие как FP-Growth, работают с дискретными наборами, поэтому непрерывные признаки требуют предварительной дискретизации.

---

### Вариант 3

1. Какой тип анализа лежит в основе метода ассоциативных правил?

- A) Классификация
- B) Кластеризация

- C) Поиск закономерностей и зависимостей в транзакциях
- D) Регрессия

**Ответ: C**

2. Что произойдет при снижении значения параметра min support в алгоритме FP-Growth?
- A) Сократится количество найденных правил
  - B) Увеличится глубина дерева
  - C) Увеличится количество частых наборов и, как следствие, правил
  - D) Алгоритм перестанет работать

**Ответ: C**

3. Какой из следующих критериев лучше всего использовать для оценки значимости правила, не учитывая его частоту?
- A) Support
  - B) Confidence
  - C) Lift
  - D) Rule length

**Ответ: C**

4. Какую роль играет параметр “max number of itemsets” при работе с большим объемом транзакций?
- A) Устанавливает число итераций
  - B) Ограничивает количество анализируемых товаров
  - C) Снижает объём выходных данных и контролирует вычислительные ресурсы
  - D) Ограничивает глубину дерева

**Ответ: C**

5. Какое правило будет иметь наибольшую поддержку (support)?
- A) То, которое встречается в большинстве транзакций
  - B) То, которое имеет самую высокую уверенность
  - C) То, которое имеет высокий lift
  - D) То, которое минимально по длине

**Ответ: A**

6. В чем особенность представления транзакционных данных для алгоритма ассоциативных правил в RapidMiner?
- A) Признаки кодируются one-hot

- В) Каждая транзакция представляется строкой с товарами, разделёнными специальным символом
- С) Строится матрица признаков
- Д) Используется бинарная регрессия

**Ответ: В**

7. Что означает значение  $Lift = 1$  для правила  $A \rightarrow B$ ?
- А) Правило имеет идеальную предсказательную силу
  - В) Между А и В нет статистической зависимости
  - С) Правило всегда истинно
  - Д) Правило является избыточным

**Ответ: В**

8. Почему нельзя напрямую применять алгоритм ассоциативных правил к числовым атрибутам без преобразования?

**Ответ:** Числовые данные требуют дискретизации, т.к. FP-Growth работает с категориальными или булевыми признаками, представляющими наличие/отсутствие элементов.

9. Как изменение  $min\ confidence$  может повлиять на качество и количество выявленных правил?

**Ответ:** Повышение порога уменьшает количество правил, но повышает их достоверность; снижение — увеличивает количество, но может включать слабые правила.

10. Чем  $Lift$  отличается от  $Confidence$  при интерпретации правил?

**Ответ:**  $Lift$  показывает, насколько вероятность появления В увеличивается при наличии А по сравнению с её базовой вероятностью, тогда как  $Confidence$  — условная вероятность появления В при А.

---

#### Вариант 4

1. Какое свойство делает FP-Growth более масштабируемым по сравнению с Apriori?
- А) Использование логарифмов
  - В) Отказ от генерации кандидатов и построение FP-дерева
  - С) Параллельная обработка
  - Д) Предварительная нормализация данных

**Ответ: В**

2. Что может стать следствием выбора слишком низкого значения  $\min$  confidence в FP-Growth?
- A) Повышенная точность
  - B) Модель станет интерпретируемой
  - C) Генерация большого числа слабых и нерелевантных правил
  - D) Пропуск редких, но сильных правил

**Ответ: C**

3. Почему важно учитывать метрику Lift при интерпретации ассоциативных правил?
- A) Lift нормирует значения поддержки
  - B) Она измеряет длину правил
  - C) Lift показывает реальную силу связи между товарами по сравнению с их независимым появлением
  - D) Она исключает пересечения между правилами

**Ответ: C**

4. Какое предположение лежит в основе ассоциативного анализа?
- A) Линейная взаимосвязь между переменными
  - B) Независимость атрибутов
  - C) Частое совместное появление элементов указывает на их потенциальную взаимосвязь
  - D) Данные предварительно нормализованы

**Ответ: C**

5. Что произойдет при установке слишком большого значения “max number of itemsets”?
- A) Алгоритм может не завершиться из-за нехватки памяти
  - B) Все возможные правила будут отброшены
  - C) Значение Lift станет равным нулю
  - D) Поддержка будет максимальной

**Ответ: A**

6. Что происходит на первом этапе работы алгоритма FP-Growth?
- A) Создание one-hot матрицы
  - B) Построение дерева частых паттернов (FP-tree)
  - C) Генерация всех возможных правил
  - D) Вычисление lift

**Ответ: B**

7. Какое ограничение имеет метрика Confidence при выборе лучших правил?
- A) Не учитывает базовую вероятность следствия
  - B) Слишком чувствительна к длине правила
  - C) Подходит только для бинарных данных
  - D) Требуется нормализации

**Ответ: А**

8. Как можно адаптировать ассоциативный анализ для выявления паттернов в медицинских записях пациентов?

**Ответ:** Необходимо перекодировать диагнозы, процедуры и лекарства в булев формат (например, «наличие/отсутствие») и применить FP-Growth к полученной матрице.

9. В чём преимущество сокращения длины правила при равных метриках support и confidence?

**Ответ:** Короткие правила проще интерпретировать, они более универсальны и лучше обобщаются на новые данные.

10. Почему важно визуализировать ассоциативные правила в виде графа?

**Ответ:** Граф помогает увидеть связи между наборами товаров или событий, выявить кластеры связанных элементов и упростить интерпретацию сложных взаимосвязей.