

Контрольные и тестовые вопросы по ПРЗ «Визуализации в RapidMiner» по вариантам с ответами

Вариант 1

1. Какой тип визуализации используется для оценки распределения одного числового признака?
A) Pie Chart
B) Box Plot
C) Histogram
D) Scatter Plot

Ответ: С

2. Чем отличается диаграмма рассеяния от пузырьковой диаграммы?
A) Scatter Plot визуализирует категориальные признаки
B) Bubble Plot использует третий признак для размера точек
C) Bubble Plot применяется только к бинарным данным
D) Scatter Plot не отображает взаимосвязи

Ответ: В

3. Для чего используется тепловая карта (Heatmap) в RapidMiner?
A) Для отображения категориальных переменных
B) Для представления плотности текста
C) Для визуализации корреляций между числовыми признаками
D) Для показа распределения пропусков

Ответ: С

4. Почему важно регулировать количество «bins» при построении гистограммы?
A) Это влияет на цветовую гамму
B) Для управления скоростью обучения модели
C) Чтобы точно контролировать разбиение диапазонов значений
D) Чтобы объединять категориальные признаки

Ответ: С

5. Что происходит при неправильном выборе осей в диаграмме рассеяния?
- A) Возникает ошибка загрузки
 - B) График не отображается
 - C) Теряется смысл взаимосвязей
 - D) Значения преобразуются автоматически

Ответ: C

6. Что такое «цветовая кодировка» (Color by) в визуализациях RapidMiner?
- A) Назначение цвета по порядку значений
 - B) Механизм фильтрации
 - C) Способ визуального разделения по значению переменной
 - D) Аналог сводной таблицы

Ответ: C

7. Какой тип графика лучше всего использовать для визуализации долей категорий в целом?
- A) Bar Chart
 - B) Box Plot
 - C) Heatmap
 - D) Pie Chart

Ответ: D

8. Почему в визуализации важно нормализовать числовые признаки?

Ответ: Без нормализации признаки с разными масштабами и диапазонами могут доминировать и исказить восприятие данных на графиках.

9. Как фильтрация данных на визуализациях помогает в анализе?

Ответ: Фильтрация позволяет сосредоточиться на подгруппах, выявить закономерности внутри сегментов, исключить шум и упростить интерпретацию.

10. Назовите минимум два способа повышения наглядности Scatter Plot в RapidMiner.

Ответ: Изменение формы и цвета точек (через Color By), регулировка прозрачности, добавление сетки и фильтров, замена осей.

Вариант 2

1. Что такое Box Plot в RapidMiner?
 - A) Визуализация плотности распределения категорий
 - B) Способ отображения медианы и разброса числового признака
 - C) Диаграмма корреляции
 - D) Метод отображения бинарных связей

Ответ: B

2. Когда лучше использовать столбчатую диаграмму (Bar Chart), а не Pie Chart?
 - A) При отображении числовых признаков
 - B) Для категориальных признаков с небольшим числом значений
 - C) Когда важно увидеть доли
 - D) При отображении плотности распределения

Ответ: B

3. Какой тип визуализации используется для анализа зависимости между двумя числовыми признаками?
 - A) Pie Chart
 - B) Scatter Plot
 - C) Box Plot
 - D) Histogram

Ответ: B

4. Что обозначают «усы» в Box Plot?
 - A) Уровень доверия
 - B) Границы интерквартильного диапазона с выбросами
 - C) Крайние значения категорий
 - D) Визуализацию плотности признака

Ответ: B

5. Для чего в визуализациях RapidMiner изменяют цветовую палитру?
 - A) Для улучшения классификации
 - B) Для ускорения построения моделей
 - C) Для повышения визуального различия между классами
 - D) Для преобразования данных

Ответ: C

6. Почему важно подписывать оси графика?
- А) Чтобы избежать визуального шума
 - В) Для корректного отображения легенды
 - С) Чтобы правильно интерпретировать, какие данные сравниваются
 - Д) Для устранения пропусков

Ответ: С

7. В каком случае нецелесообразно использовать Histogram?
- А) Для анализа числовых данных
 - В) Для анализа категориальных переменных
 - С) При поиске выбросов
 - Д) При анализе плотности значений

Ответ: В

8. В чём отличие между Bar Chart и Histogram?

Ответ: Bar Chart применяется к категориальным данным и имеет отдельные столбцы, а Histogram — к числовым данным, где значения группируются в интервалы.

9. Зачем в визуализациях RapidMiner используют агрегирование (aggregation)?

Ответ: Агрегирование упрощает данные, позволяя отображать суммарные или средние значения, делая визуализацию более компактной и осмысленной.

10. Как визуализация может использоваться для отбора признаков в модели?

Ответ: Анализ диаграмм (например, корреляционной карты или Scatter Plot) позволяет выявить наиболее информативные признаки и исключить избыточные.

Вариант 3

1. В чем преимущество использования диаграммы рассеяния (Scatter Plot) при анализе модели?
- А) Можно визуализировать частоту встречаемости категорий
 - В) Удобно отображать зависимость между двумя числовыми признаками

- C) Легко анализировать размер выборки
- D) Показывает только средние значения

Ответ: В

2. Почему важно учитывать плотность точек на Scatter Plot?
- A) Это влияет на производительность модели
 - B) Чтобы визуализировать выбросы и возможные кластеры
 - C) Для настройки поддержки и доверия
 - D) Для нормализации шкалы

Ответ: В

3. Какую роль играет прозрачность (opacity) в графиках RapidMiner?
- A) Увеличивает разрешение графика
 - B) Показывает группировку признаков
 - C) Позволяет визуализировать плотность наложения точек
 - D) Изменяет размер графика

Ответ: С

4. Какой тип графика наиболее подходит для анализа распределения значений между несколькими категориями?
- A) Heatmap
 - B) Pie Chart
 - C) Box Plot
 - D) Scatter Plot

Ответ: С

5. Что происходит при некорректном выборе осей в визуализации?
- A) RapidMiner автоматически исправляет
 - B) Построение графика отменяется
 - C) Визуализация становится некорректной или вводящей в заблуждение
 - D) Осей становится две

Ответ: С

6. Какое преимущество Heatmap перед другими графиками?
- A) Показывает только медианы
 - B) Выделяет категориальные переменные
 - C) Представляет многомерные зависимости через цветовую

интенсивность

D) Заменяет гистограмму

Ответ: С

7. Когда не рекомендуется использовать Pie Chart?

A) При небольшом числе категорий

B) Когда доли выражаются в процентах

C) При большом количестве категориальных значений

D) Для сравнения бинарных переменных

Ответ: С

8. Почему важно правильно выбрать шкалу при визуализации?

Ответ: Некорректный выбор шкалы может привести к искажению восприятия взаимосвязей и неверным аналитическим выводам.

9. В чем отличие между визуализациями на основе количества (count) и среднего значения (mean)?

Ответ: Визуализация по count показывает частоту появления значений, а по mean — среднюю величину признака в каждой группе, что важно при анализе влияния категорий.

10. Какую информацию может дать Overlay Plot при сравнении моделей?

Ответ: Overlay Plot позволяет наложить несколько кривых или точек, визуализируя разницу между моделями, кривыми ошибок, результатами предсказания.

Вариант 4

1. Какой тип визуализации помогает анализировать взаимосвязи между более чем двумя переменными одновременно?

A) Box Plot

B) Pie Chart

C) Scatter Matrix

D) Histogram

Ответ: С

2. Что произойдёт при наложении фильтра на визуализацию в RapidMiner?

A) Изменится количество итераций

B) Визуализация обновится для отфильтрованных данных

- C) Построится новая модель
- D) Признаки станут категориальными

Ответ: B

3. Почему важно различать типы переменных при выборе способа визуализации?
- A) Это не влияет на выбор
 - B) Для правильной цветовой палитры
 - C) Для корректной агрегации и представления информации
 - D) Для оптимизации скорости обучения

Ответ: C

4. Какое назначение у параметра "Color by" в Scatter Plot?
- A) Фильтрация по цвету
 - B) Добавление переменной в качестве оси
 - C) Назначение переменной для визуального кодирования точек по цвету
 - D) Отображение среднего значения

Ответ: C

5. Что такое агрегация в контексте визуализации в RapidMiner?
- A) Удаление признаков
 - B) Сглаживание функции потерь
 - C) Объединение данных по группам с применением функций (среднее, сумма и др.)
 - D) Разбиение на классы

Ответ: C

6. Какое ограничение у использования Heatmap для визуализации больших матриц признаков?
- A) Не может использовать числовые значения
 - B) Может стать трудно читаемой при большом числе переменных
 - C) Требуется one-hot encoding
 - D) Использует только бинарные классы

Ответ: B

7. Как визуализация может быть использована для отладки качества модели?
- A) Показывает статистику ошибок

- В) Отображает графики предсказаний и сравнение с фактом
- С) Корректирует веса
- Д) Увеличивает скорость обучения

Ответ: В

8. Почему важно объединять визуализацию с фильтрацией данных?

Ответ: Это помогает сосредоточиться на подгруппах, сравнить сегменты, проверить гипотезы и провести качественный анализ.

9. В чем отличие Scatter Matrix от набора обычных Scatter Plot?

Ответ: Scatter Matrix отображает все возможные пары числовых признаков в единой сетке, позволяя быстро оценить взаимосвязи между ними.

10. Как визуализации могут помочь в объяснении результатов модели неэкспертной аудитории?

Ответ: Визуализации упрощают сложные зависимости, делают данные наглядными, выявляют закономерности и улучшают понимание выводов.