

# **Контрольные и тестовые вопросы по ПР13**

## **«Интерпретация моделей в RapidMiner» по вариантам с ответами**

### **Вариант 1**

**1. Какой критерий разделения узлов деревьев решений обычно обеспечивает наибольшую интерпретируемость модели?**

- A) Gini Impurity
- B) Information Gain
- C) Entropy Ratio
- D) Mean Squared Error

**Ответ: B**

**2. Что произойдет при слишком глубоком дереве решений?**

- A) Увеличится скорость обучения
- B) Модель будет более устойчива
- C) Повысится вероятность переобучения
- D) Снизится сложность модели

**Ответ: C**

**3. Какой инструмент RapidMiner позволяет вручную исследовать влияние признаков на предсказание?**

- A) Apply Model
- B) Split Data
- C) Model Simulator
- D) Replace Missing Values

**Ответ: C**

**4. Какие признаки из набора данных Titanic рекомендуется исключать перед моделированием?**

- A) Age, Fare
- B) Name, Ticket, Cabin
- C) Pclass, Sex
- D) SibSp, Parch

**Ответ: B**

**5. Какой параметр влияет на минимальное количество наблюдений в листе дерева решений?**

- A) Bootstrap size
- B) Minimal leaf size
- C) Maximal depth
- D) Number of trees

**Ответ: B**

**6. Почему важно проверять баланс классов перед моделированием?**

- A) Для исключения признаков
- B) Для ускорения вычислений
- C) Для предотвращения перекоса модели
- D) Для нормализации данных

**Ответ: C**

**7. Какой тип данных должен иметь признак Embarked в наборе Titanic после предобработки?**

- A) Real
- B) Integer
- C) Nominal
- D) Binominal

**Ответ: C**

**8. Почему важно интерпретировать дерево решений, а не только оценивать точность?**

**Ответ:** Интерпретация помогает понять причины решений модели и повысить доверие к ней.

**9. Какие признаки чаще всего имеют наибольшее влияние в задаче предсказания выживаемости на Титанике?**

**Ответ:** Пол пассажира (Sex) и класс билета (Pclass).

**10. Что дает использование симуляции «Что если?» в RapidMiner?**

**Ответ:** Возможность оценить, как изменения входных параметров влияют на предсказания модели.

---

## **Вариант 2**

**1. Что отражает параметр "maximal depth" в Decision Tree?**

- A) Максимальное количество ветвей
- B) Максимальный размер набора данных
- C) Максимальная глубина дерева
- D) Число классов для классификации

**Ответ: C**

**2. Для чего применяется оператор Numeric to Binominal?**

- A) Для преобразования текстовых признаков
- B) Для замены пропущенных значений
- C) Для преобразования числовых признаков в бинарные
- D) Для исключения признаков

**Ответ: C**

**3. Какой оператор в RapidMiner используется для оценки качества модели?**

- A) Split Data
- B) Set Role
- C) Performance (Classification)
- D) Replace Missing Values

**Ответ: C**

**4. Какой показатель информирует о важности признаков в RapidMiner?**

- A) ROC Curve
- B) Attribute Weights
- C) Accuracy
- D) Precision

**Ответ: B**

**5. Какой метод используется для замены пропущенных значений в числовом признаке Age?**

- A) Удаление строк
- B) Замена средним значением
- C) Замена медианой
- D) Замена случайным числом

**Ответ: B**

**6. Почему нежелательно делать слишком маленький `minimal leaf size` в дереве решений?**

- A) Увеличивается переобучение модели
- B) Снижается точность модели
- C) Замедляется обучение модели
- D) Уменьшается сложность модели

**Ответ: A**

**7. В чем преимущество визуализации дерева решений?**

- A) Ускорение обработки данных
- B) Улучшение интерпретации модели
- C) Уменьшение размера данных
- D) Исключение ненужных признаков

**Ответ: B**

**8. Почему пол пассажира является значимым фактором при моделировании выживаемости на Титанике?**

**Ответ:** Женщины имели преимущество при эвакуации, что исторически подтверждено и отражено в данных.

**9. Что произойдет с прогнозом выживаемости, если изменить класс билета с первого на третий для женщины?**

**Ответ:** Вероятность выживания значительно снизится.

**10. Зачем используют оператор `Split Data` перед обучением модели?**

**Ответ:** Для разделения данных на обучающую и тестовую выборки, чтобы объективно оценить качество модели.

---

### **Вариант 3**

**1. Какая метрика классификации наиболее чувствительна к дисбалансу классов?**

- A) Precision
- B) Accuracy
- C) Recall
- D) F1-score

**Ответ: B**

**2. Какую проблему решает замена пропущенных значений перед моделированием?**

- A) Увеличение сложности модели
- B) Снижение качества модели
- C) Предотвращение искажений в результатах модели
- D) Ускорение визуализации данных

**Ответ: C**

**3. Что оценивает оператор Statistics в RapidMiner?**

- A) Качество классификации
- B) Распределение признаков и классов
- C) Важность признаков
- D) Глубину дерева решений

**Ответ: B**

**4. Какой столбец данных должен быть обязательно установлен как целевой (label) при моделировании выживаемости на Титанике?**

- A) Age
- B) Fare
- C) Survived
- D) Sex

**Ответ: C**

**5. Какие признаки являются малозначимыми для предсказания выживаемости на Титанике?**

- A) Age и Sex
- B) Name, Ticket и Cabin
- C) Fare и Pclass
- D) SibSp и Parch

**Ответ: B**

**6. Как изменится модель, если сильно увеличить параметр `maximal depth`?**

- A) Снизится вероятность переобучения
- B) Упростится интерпретация модели
- C) Возрастет вероятность переобучения
- D) Уменьшится точность

**Ответ: C**

**7. Что определяет параметр `criterion` в Decision Tree?**

- A) Количество ветвей

- B) Критерий разделения узлов
- C) Максимальную глубину
- D) Количество листьев

**Ответ: B**

**8. Почему важно экспериментировать с гиперпараметрами модели?**

**Ответ:** Чтобы найти оптимальные значения, которые обеспечивают наилучшее качество и устойчивость модели.

**9. Как возраст пассажира влияет на прогноз выживаемости в модели?**

**Ответ:** Более молодые пассажиры имели более высокую вероятность выживания.

**10. Зачем нужна визуализация важности признаков (AttributeWeights)?**

**Ответ:** Для понимания, какие признаки оказывают наибольшее влияние на решения модели.

---

**Вариант 4**

**1. Каким способом лучше всего проверить влияние отдельного признака на итоговое предсказание модели?**

- A) Удалить признак
- B) Изменить значения признака и наблюдать изменения прогноза
- C) Увеличить глубину дерева
- D) Уменьшить размер выборки

**Ответ: B**

**2. Какой признак необходимо преобразовать из числового в номинальный для анализа данных Титаника?**

- A) Fare
- B) Age
- C) Pclass
- D) SibSp

**Ответ: C**

**3. Что такое переобучение в контексте деревьев решений?**

- A) Модель слишком проста
- B) Модель плохо интерпретируема
- C) Модель хорошо предсказывает обучающую выборку, но плохо – тестовую
- D) Модель слишком медленно обучается

**Ответ: C**

**4. Какой оператор используется для установки целевой переменной (label)?**

- A) Replace Missing Values
- B) Set Role
- C) Split Data
- D) Apply Model

**Ответ: B**

**5. Почему стоимость билета (Fare) может быть важной для прогноза выживания?**

**Ответ:** Более дорогие билеты обычно указывали на высший класс, пассажиры которого имели больше шансов выжить.

**6. Что означает параметр contamination threshold в задаче анализа аномалий?**

**Ответ:** Процент данных, который считается аномальным в выборке.

**7. Зачем нужно исключать признак Cabin при предсказании выживаемости?**

**Ответ:** Признак Cabin содержит много пропусков и незначим для точного моделирования.

**8. В чем польза анализа «Что если?» сценариев?**

**Ответ:** Позволяет понять, как изменения в данных влияют на выводы модели.

**9. Почему важно использовать разные критерии оценки качества модели?**

**Ответ:** Для всесторонней оценки производительности модели.

**10. Что показывает вкладка Plot View в Model Simulator?**

**Ответ:** Важность признаков в модели.