ТЕСТ ПО ДЕРЕВЬЯМ РЕШЕНИЙ

1. Какое главное преимущество деревьев решений перед другими алгоритмами машинного обучения?

а) Они всегда точнее нейронных сетей

**б) Они легко интерпретируются и объясняются**

в) Они работают только с числовыми данными

г) Они требуют меньше памяти, чем другие алгоритмы

2. Какая метрика обычно используется для измерения неоднородности в деревьях решений для задач классификации?

а) Среднеквадратическая ошибка

б) Стандартное отклонение

**в) Индекс Джини**

г) Евклидово расстояние

3. Что такое pruning (обрезка) в деревьях решений?

а) Добавление новых ветвей для улучшения точности

**б) Удаление лишних ветвей для предотвращения переобучения**

в) Объединение нескольких деревьев в одно

г) Случайное разделение узлов

4. Когда следует использовать критерий 'entropy' вместо 'gini' в дереве решений?

а) При работе только с непрерывными переменными

б) Когда нужны более быстрые вычисления

**в) Когда нужно сильнее штрафовать смешанные разделения классов**

г) Только при бинарной классификации

5. Что контролирует параметр max\_depth в дереве решений?

а) Максимальное количество признаков для рассмотрения

б) Максимальное количество листовых узлов

**в) Максимальную высоту/уровни дерева**

г) Максимальное количество обучающих примеров

6. Что такое листовой узел в дереве решений?

а) Первый узел в верхней части дерева

б) Узел, который разделяется на две ветви

**в) Узел без потомков, который даёт предсказание**

г) Узел, содержащий только числовые значения

7. Важность признаков в деревьях решений определяется:

а) Алфавитным порядком названий признаков

**б) Уменьшением неоднородности, взвешенным по количеству образцов**

в) Позицией признаков в наборе данных

г) Случайным присвоением значений важности

8. Какое минимальное количество образцов требуется в узле для разделения (значение по умолчанию в sklearn)?

а) 1

**б) 2**

в) 5

г) 10

9. Какой из следующих признаков указывает на переобучение в дереве решений?

а) Дерево имеет очень мало узлов

**б) Точность на обучающей выборке намного выше, чем на тестовой**

в) Все листовые узлы чистые

г) Дерево идеально сбалансировано

10. Какой метод можно использовать для улучшения производительности одиночного дерева решений?

а) Использование только категориальных признаков

б) Удаление всех шагов предобработки

**в) Применение кросс-валидации и настройка гиперпараметров**

г) Ограничение дерева только двумя уровнями