

1. Какой инструмент чаще всего используется для работы с табличными данными?
2. Какой язык программирования чаще всего применяется для анализа данных?
3. Дан набор данных о возрасте людей: [25, 30, 35, 40, 45]. Чему равна медиана?
4. Что такое Data Mining?
5. Как называются данные, собранные пользователями интернета (социальные сети, форумы)?
6. Что такое Big Data?
7. Какой метод анализа данных помогает выявить связь между двумя переменными?
8. Что такое парсинг данных?
9. Что такое релевантность данных?
10. Какой критерий важнее всего при выборе открытого источника данных?
11. Какой тип графика используется для отображения соотношения частей целого?
12. Какой показатель используется для оценки качества кластеризации методом измерения средней близости точки к центрам других кластеров и собственной группы?
13. Что обозначает коэффициент Кендалла ( $\tau$ )?
14. Что значит меньшее значение Индекса Дэвиса-Болдина (DB-index)?
15. При расчете среднеквадратичной ошибки (MSE) в каком случае достигается лучшее качество модели?
16. Рассчитайте среднеквадратичную ошибку (MSE) для следующей таблицы истинных и предсказанных значений:  
Истинное значение [5, 10, 15, 20]  
Предсказанное значение [6, 9, 16, 21]
17. Дано две пары рейтингов:  $(x_1, x_2)$  и  $(y_1, y_2)$ , рассчитайте коэффициент Кендалла ( $\tau$ ) для пар: (1,3) и (2,4).
18. Используя формулу коэффициента детерминации ( $R^2$ ), найдите  $R^2$  для случая, когда общая сумма квадратов равна 100, а сумма квадратов ошибок равна 20.



Магазин готовых ответов на тесты, практики, купить в магазине! ➔ [ОТВЕТЫ](#)

Нужна помощь с тестами, практикой, дипломной вкр? ➔ [КОНСУЛЬТАЦИЯ](#)

19. Рассчитайте индекс Дэвиса-Болдина (DB-index) для следующего примера: Имеются два кластера с радиусами окружностей  $r_1=2$  и  $r_2=3$ . Расстоянием между центрами кластеров равно 10.

20. Рассчитайте индекс силуэта (Silhouette Index) для точки A, принадлежащей кластеру 1, используя следующую таблицу расстояний:

Точка A

Расстояние до других точек кластера 1 = [2]

Минимальное расстояние до другого кластера = [5]

