1. **Для чего перед построением моделей машинного обучения предварительно используется корреляционный анализ?**

Корреляционный анализ устанавливает связь между исследуемыми случайными переменными машинной модели и оценивает тесноту этой связи. Однако в дополнение к этому желательно располагать моделью зависимости, полученной после обработки результатов моделирования.

1. **Как построить диаграмму типа «тепловая карта» на основе корреляционной матрицы?**
2. **О чем говорит сильная корреляция между исходным признаком и целевым признаком?**
3. **О чем говорит сильная корреляция между двумя исходным признаками? Как это влияет на построение модели?**
4. **В наборе данных содержатся признаки, в которых пропущено 20%, 50%, 80% данных. Какие из этих признаков Вы будете или не будете использовать при построении модели и почему?**
5. **В чем преимущества и недостатки методов «label encoding» и «one hot encoding»?**
6. **Какие способы заполнения пропусков в количественных признаках Вы знаете?**
7. **Как осуществляется масштабирование данных на основе стандартизованной оценки?**
8. **Опишите принцип работы метода ближайших соседей.**
9. **Опишите принцип работы машины опорных векторов.**
10. **Опишите принцип работы дерева решений.**
11. **Почему ансамблевые модели могут иметь преимущество перед одиночными моделями машинного обучения?**
12. **Опишите принцип работы бэггинга.**
13. **Опишите принцип работы случайного леса.**
14. **Опишите принцип работы метода AdaBoost.**
15. **Опишите принцип работы градиентного бустинга.**
16. **Что такое смещение и дисперсия при оценке модели машинного обучения?**
17. **Почему для задач классификации и регрессии используются различные метрики?**
18. **В каких соотношениях делят данные в тестовой и обучающей выборках?**
19. **Если решается задача классификации и наблюдается сильный дисбаланс классов, как это может повлиять на формирование обучающей и тестовой выборок?**
20. **Что такое кросс-валидация и для чего она используется?**
21. **Если решается задача классификации и наблюдается сильный дисбаланс классов, как это может повлиять на кросс-валидацию?**
22. **Какие схемы кросс-валидации Вы знаете?**
23. **Если мы закодируем категориальные признаки с использованием подходов «label encoding» или «one hot encoding», как это может повлиять на качество модели?**
24. **Мы используем для построения модели метод ближайших соседей. В наборе данных есть только количественные признаки. Улучшит ли предварительная нормализация данных качество построенной модели?**