1) Кластеризация это–

А) Разделение большой группы объектов на несколько поменьше (100%)

Б) Объединение нескольких маленьких групп объектов в одну большую

В) Система рекомендаций

Г) Арбуз

Д) Прогнозирование величину величины

2)В традиционном компьютерном программировании вы вводите команды. В машинное обучение мы загружаем…

А) Данные (100%)

Б) Узоры (0%)

В) Программы (0%)

Г) Правила (0%)

3) Какие метрики эффективности можно применять для оценки регрессионных моделей?

А) Средний квадрат ошибки (33%)

Б) Среднеквадратичная ошибка (33%)

В) Куб ошибки (0%)

Г) Среднеквадратичная логарифмическая ошибка (33%)

Д) Абсолютное процентное отклонение (0%)

4) Вы являетесь частью команды специалистов по обработке и анализу данных, работающей в национальной сети ресторанов быстрого питания. Вы создаете простой отчет, показывающий тенденцию: клиенты, которые посещают магазин чаще и покупают меньше еды, тратят больше, чем клиенты, которые посещают магазин реже и покупают больше еды. Какова наиболее вероятная диаграмма, созданная вашей командой?

А) Диаграмма многоклассовой классификации (0%)

Б) Линейная регрессия и диаграммы рассеяния (100%)

В) Сводная таблица (0%)

Г) Диаграмма кластера К-средних (0%)

Д) Параллельная регрессия (0%)

5) Метод градиентного спуска –

А) Это заливка, когда один цвет плавно переходит в другой (0%)

Б) Это алгоритм оптимизации, используемый для минимизации ошибок в модели машинного обучения (50%)

В) Это способ обучения и совершенствования модели машинного обучения (50%)

Г) Это метод обучения модели при её переобучении

6) Основным методом сглаживания в sklearn является

А) linear\_regression (100%)

Б) poly\_regression (0%)

В) ridge\_classifier (0%)

Г) lazy\_classifier (0%)

7) Как называется метод разделения данных на обучающую и тестовую выборки в sklearn?

А) train\_test\_split (100%)

Б) create\_test\_train\_split (0%)

В) split\_data (0%)

Г) create\_train\_test\_split (0%)

8) В библиотеке sklearn существуют такие регрессоры

А) Линейная регрессия (33%)

Б) Кубическая регрессия (0%)

В) Ридж-регрессия (33%)

Г) Градиентный бустинг (33%)

Д) Метод лучей (0%)

9) Коэффициент детерминации для модели регрессии

А) Показывает сколько в модели различных значений (0%)

Б) Показывает насколько хорошо модель соответствует данным, на которых она построена (50%)

В) Оценивает долю дисперсии зависимой переменной (50%)

Г) Ставит единицы и двойки (0%)

10) Random forest — это модифицированная и улучшенная версия

А) aggregated trees (0%)

Б) boosted trees (100%)

В) bagged trees (0%)

Г) stacked trees (0%)

11) Функция ошибки в моделях нужна, поскольку она

А) выводит неудобные для модели значения (0%)

Б) используется для настройки параметров модели таким образом, чтобы минимизировать ошибку между прогнозируемыми и фактическими значениями (50%)

В) создаёт массив данных, на которых модель ошиблась (0%)

Г) позволяет оценить, насколько хорошо модель соответствует данным (50%)

12) Для генерации массивов Numpy использует

А) numpy.zeros (50%)

Б) numpy.ranpom.randint (0%)

В) numpy.random.rand (50%)

Г) numpy.randint.random (0%)

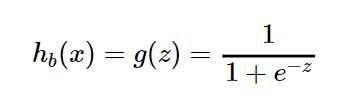
13) Метрика точности регрессии показывает

А) насколько не точны прогнозы модели (0%)

Б) насколько могут быть точны прогнозы модели, если убрать неудобные значения (0%)

В) насколько много в прогнозах модели ошибок (0%)

Г) насколько точны прогнозы модели (100%)



14) Данная функция является

А) Кубической регрессией (0%)

Б) Границей принятия решений (0%)

В) Логической регрессией (100%)

Г) Функцией ошибки градиентного спуска (0%)