

**федеральное государственное автономное образовательное учреждение высшего
образования
Первый Московский государственный медицинский университет им. И.М.
Сеченова Министерства здравоохранения Российской Федерации
(Сеченовский Университет)**

Институт цифрового биодизайна и искусственного интеллекта в медицине
Кафедра информационных технологий и обработки медицинских данных

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ
Фундаментальные основы программирования
основная профессиональная образовательная программа высшего
профессионального образования - программа бакалавриата

45.03.04 Интеллектуальные системы в гуманитарной сфере

Перечень вопросов по курсу для проведения централизованного тестирования.

Вопросы и варианты ответов. Правильные варианты отмечены знаком ✓.

Анализ медицинских данных Вопрос_1

Восстановление зависимости между парами (признаковое описание, целевая переменная) по данным, доступным для анализа выполняется при помощи алгоритмов машинного обучения

1. ✓ с учителем
2. без учителя
3. с подкреплением
4. без подкрепления

Анализ медицинских данных Вопрос_10

Что называют обучением с подкреплением?

1. ✓ моделирование обучения методом проб и ошибок
2. моделирование обучения с контролирующими параметрами
3. обучение с поддержкой экспертных знаний
- 4.

Анализ медицинских данных Вопрос_100

Для разметки нового (тестового) набора данных используется метод:

1. ✓ predict
2. fit
3. fit_transform
4. predict_proba

Анализ медицинских данных Вопрос_101

Какой метод используется для вывода количества строк и столбцов набора данных?

1. ✓ df.shape
2. df.dtypes
3. df.info()
4. df.size

Анализ медицинских данных Вопрос_104

Для вычисления частоты появления всех уникальных значений в датасете используется оператор

1. ✓ df.value_counts()
2. df.mean()
3. df.isnull().count()
4. df.duplicated().sum()

Анализ медицинских данных Вопрос_108

Для масштабирования числовых признаков используется метод

1. ✓ fit_transform()
2. reshape()
3. fit()
4. transform()

Анализ медицинских данных Вопрос_11

Корреляция может быть определена как

1. ✓ Связь, зависимость
2. Отношение, соотношение
3. Функция, уравнение
4. Коэффициент, вес

Анализ медицинских данных Вопрос_110

Программным кодом `X_train, X_test, y_train, y_test = train_test_split( data, digits.target,`

test_size=0.5, shuffle=False) clf.fit(X_train, y_train) тестовое и обучающее множество разделены в соотношении

1. ✓ 50:50
2. 80:20
3. 95:5
4. 95,5:0,5

Анализ медицинских данных Вопрос_111

В программном коде `cfr = svm.SVC(gamma=0.001)` классификации цифр используется классификатор

1. ✓ метод опорных векторов
2. метод наименьших квадратов
3. метод главных компонент
4. метод К-средних

Анализ медицинских данных Вопрос_112

Что такое Система управления базами данных "

1. ✓ комплекс языковых и программных средств, предназначенный для создания, ведения и совместного использования БД многими пользователями. "
2. это программа для создания и хранения информации
3. это устройство для отслеживания работы компьютеров, привязанных к сети "
4. комплекс языковых и программных средств, предназначенный для удаления, ведения и совместного использования БД многими пользователями. "

Анализ медицинских данных Вопрос_113

Что такое Хранилище данных? "

1. ✓ специализированная БД, в которой хранятся данные в формате, оптимизированном для поддержки принятия решений. ХД содержит исторические данные, полученные из оперативных БД, а также данные из других внешних источников "
2. это место, где хранятся изображения, которые передаются на компьютер
3. это специальное устройство для загрузки данных на компьютер или сервер
4. это хранилище любой информации одного типа

Анализ медицинских данных Вопрос_114

Что такое Аналитическая обработка в реальном времени? "

1. ✓ набор инструментов, которые работают вместе, чтобы обеспечить расширенную среду анализа данных для извлечения, обработки и моделирования данных из ХД. "
2. набор инструментов для извлечения данных
3. набор инструментов , где хранятся результаты обработки и анализа данных
4. набор инструментов для моделирования данных

Анализ медицинских данных Вопрос_115

Какие этапы относятся к нормализации данных? "

1. ✓ Устранить противоречивость + Упростить таблицы + Устранить избыточность "
2. Устранить таблицы + Сократить избыточность + Устранить противоречивость
3. Сократить таблицы + Объединить все таблицы в одну
4. Нет правильного ответа

Анализ медицинских данных Вопрос_116

MetoSheet это:

1. ✓ Это библиотека Python с открытым исходным кодом, которая используется для создания интерактивных информационных панелей.
2. Это библиотека Java с открытым исходным кодом, которая используется для создания интерактивных информационных панелей.

3. Это библиотека C# с открытым исходным кодом, которая используется для создания интерактивных информационных панелей.

4. Это библиотека C++ с открытым исходным кодом, которая используется для создания интерактивных информационных панелей.

Анализ медицинских данных Вопрос_117

Библиотека bamboolib это... "

1. ✓ это графический интерфейс для pandas Dataframes , который позволяет любому работать с Python в Jupyter Notebook или JupyterLab " "

2. это графический интерфейс для numpy Dataframes , который позволяет любому работать с Python в Jupyter Notebook или JupyterLab "

3. это графический интерфейс для numpy Dataframes , который позволяет любому работать с C#Python в Jupyter Notebook или JupyterLab "

4. это графический интерфейс для pandas Dataframes , который позволяет любому работать с C#Python в Jupyter Notebook или JupyterLab "

Анализ медицинских данных Вопрос_118

Что такое Разведочный анализ данных? "

1. ✓ анализ основных свойств данных, нахождение в них общих закономерностей, распределений и аномалий, построение начальных моделей"

2. анализ второстепенных свойств данных, нахождение в них общих закономерностей, распределений и аномалий, построение начальных моделей

3. анализ второстепенных свойств данных, нахождение в них различных значений, распределений и аномалий, построение начальных моделей

4. анализ основных свойств данных, нахождение в них различий, распределений и аномалий, построение конечных моделей

Анализ медицинских данных Вопрос_119

Функция дедупликации Mito нужна для

1. ✓ удаления нежелательных данных.

2. сохранения важных данных

3. удаления всех данных

4. сравнения разных групп данных

Анализ медицинских данных Вопрос_12

Зависимыми выборками являются:

1. ✓ одни и те же объекты в разные моменты времени

2. совокупность мужчин и совокупность женщин

3. больные сахарным диабетом и больные гриппом

4. выборки, полученные при рандомизации

Анализ медицинских данных Вопрос_120

Что такое Наука о данных? "

1. ✓ раздел информатики, изучающий проблемы анализа, обработки и представления данных в цифровой форме."

2. отдельная наука, изучающая способы анализа данных

3. отдельная наука, изучающая способы представления данных не в цифровой форме

4. раздел информатики, изучающий проблемы ввода информации и предоставления ее в цифровой форме

Анализ медицинских данных Вопрос_121

Что включает в себя нормализация данных в базе данных?

1. ✓ Устранение противоречивостей и избыточности, упрощение таблицы

2. Увеличение объема хранимых данных для ускорения доступа к информации

3. Слияние нескольких таблиц в одну для упрощения структуры базы данных

4. Добавление данных для создания дополнительных резервных копий информации

Анализ медицинских данных Вопрос_122

Из-за чего возникает парадокс высокой точности в машинном обучении?

1. ✓ Из-за несбалансированных наборов данных
2. Из-за использования недостаточно мощных алгоритмов обучения
3. Из-за недостатка данных для обучения машины
4. Из-за использования слишком сложных моделей

Анализ медицинских данных Вопрос_123

Что такое переобучение (overfitting) модели?

1. ✓ Модель запоминает или имитирует учебные данные и не может хорошо обобщить новые данные (тестовые данные)
2. Модель слишком проста и не может уловить сложные закономерности в данных
3. Модель использует недостаточно признаков для обучения
4. Модель использует неподходящий алгоритм машинного обучения

Анализ медицинских данных Вопрос_124

Что НЕ является признаком хорошо обученной (appropriate fitting) модели?

1. ✓ Модель не может определить важные закономерности в обучающих данных
2. Модель не слишком простая и не слишком сложная
3. Модель находит основные закономерности в данных, а не шум
4. Модель будет достаточно хорошо работать и на обучающих, и на тестовых данных

Анализ медицинских данных Вопрос_125

В чём разница между обучением с учителем и обучением без учителя в машинном обучении?

1. ✓ В обучении с учителем модель учится на примерах, а в обучении без учителя модель сама находит закономерности в данных
2. Обучение с учителем используется для анализа изображений, а обучение без учителя - для анализа текста
3. Обучение с учителем требует меньше данных, чем обучение без учителя
4. Обучение с учителем всегда более точное, чем обучение без учителя

Анализ медицинских данных Вопрос_126

Какого расстояния НЕТ в метриках кластеризации?

1. ✓ Расстояние Хэмминга
2. Евклидово расстояние
3. Расстояние Манхэттена
4. Расстояние Канберра

Анализ медицинских данных Вопрос_127

Что является характеристикой временных рядов?

1. ✓ Тренд
2. Среднее значение
3. Количество наблюдений
4. Стандартное отклонение

Анализ медицинских данных Вопрос_128

Что не относится к параметрам алгоритма Apriori

1. ✓ min_features
2. min_support
3. min_length
4. min_lift

Анализ медицинских данных Вопрос_129

Выберите неверное утверждение

1. ✓ В методе бэггинга модели обучаются последовательно, каждая следующая модель учится исправлять ошибки предыдущих.
2. В бустинге каждая новая модель фокусируется на примерах, на которых предыдущие модели работали менее эффективно.
3. В результате бустинга получается ансамбль моделей, где каждая следующая модель стремится улучшить предыдущие.
4. В бэггинге каждая модель может использовать один и тот же алгоритм обучения, но обучаться на разных подвыборках данных.

Анализ медицинских данных Вопрос_13

Зависимость веса от роста человека (росто-весовой индекс) описывается при помощи:

1. ✓ линейной регрессии
2. логистической регрессии
3. множественной регрессии
4. экспоненциальной регрессии

Анализ медицинских данных Вопрос_130

Что не относится к параметрам библиотеки LightGBM?

1. ✓ max_length
2. min_data_in_leaf
3. max_depth
4. max_depth

Анализ медицинских данных Вопрос_131

Какое утверждение неверно для модели AdaBoostClassifier

1. ✓ невосприимчив к шуму и выбросам в данных
2. использует простые деревья решений с одним разбиением
3. во время каждой итерации веса неправильно классифицированных образов увеличиваются
4. использует широкий спектр базовых базовых обучающихся, таких как деревья решений и линейные модели

Анализ медицинских данных Вопрос_132

RandomForest это

1. ✓ Алгоритм машинного обучения, используемый для классификации и регрессии.
2. Инструмент для автоматизации процесса анализа текста и выявления тематических трендов.
3. Система распределенных вычислений для обучения моделей машинного обучения на больших наборах данных.
4. Метод случайного выбора признаков и наблюдений для создания уникальных наборов обучающих данных в процессе обучения моделей.

Анализ медицинских данных Вопрос_133

Что не относится к ML-задачам?

1. ✓ Поиск точных математических решений
2. Кластеризация
3. Ранжирование
4. Обнаружение аномалий

Анализ медицинских данных Вопрос_134

Для чего предназначена функция дедупликации Mito?

1. ✓ для удаления нежелательных данных
2. для создания резервных копий
3. для группировки повторных строк в новый столбец
4. для замены повторных значений нулями

Анализ медицинских данных Вопрос_135

С помощью какого инструмента можно создать сводную таблицу в Mito?

1. ✓ PIVOT
2. GRAPH
3. MERGE
4. TABLE

Анализ медицинских данных Вопрос_136

С помощью какой агрегатной функции можно найти сумму значений в SQL?

1. ✓ SUM
2. COUNT
3. GETSUM
4. SUMVALUES

Анализ медицинских данных Вопрос_137

Что делает операнд ORDER BY в SQL?

1. ✓ задает порядок сортировки результирующего множества
2. позволяет выделять в результирующем множестве записей группы
3. задает условия, которым должны удовлетворять записи в результирующем векторе
4. используется для дополнительной селекции записей во время определения групп

Анализ медицинских данных Вопрос_139

К-кратная перекрестная валидация - это

1. ✓ метод проверки, чтобы увидеть, насколько хорошо обученная модель обобщается для независимого набора проверки.
2. способ увеличения точности модели путем добавления K новых признаков и повторного обучения на них.
3. разбиение данных на K частей, но каждая часть используется не только для оценки, но и для обучения модели.
4. увеличение скорости обучения модели за счет параллельного обучения на K рабочих станциях

Анализ медицинских данных Вопрос_14

Какой из методов является нелинейным?

1. ✓ Метод k-ближайших соседей (KNN)
2. Дискриминантный анализ (LDA)
3. Множественная регрессия
4. Линейная регрессия (LR)

Анализ медицинских данных Вопрос_140

Метод случайного леса - это

1. ✓ алгоритм машинного обучения, заключающийся в использовании ансамбля решающих деревьев. Алгоритм сочетает в себе две основные идеи: метод бэггинга и метод случайных подпространств
2. алгоритм для обучения искусственного интеллекта самообучению путем имитации поведения животных в лесу.
3. метод оценки вероятности случайного появления нескольких событий, основанный на анализе случайно выбранных выборок.
4. создание псевдослучайных чисел при генерации шифров для криптографических систем безопасности.

Анализ медицинских данных Вопрос_141

Гиперпараметры - это

1. ✓ параметры алгоритмов машинного обучения, которые влияют на структуру алгоритмов и производительность (качество) моделей.

2. специальные алгоритмы, позволяющие искусственным нейронным сетям с нулевым обучением самостоятельно определять, какие данные считать выбросами.
3. специальные параметры, которые можно исправить в процессе обучения нейронной сети, чтобы улучшить ее эмоциональный интеллект.
4. метод логистической регрессии, который позволяет оценить вероятность принадлежности объекта к одному из классов на основе его признаков.

Анализ медицинских данных Вопрос_142

Ассоциативным правилом называется

1. ✓ импликация $X \Rightarrow Y$, где $X, Y \subseteq I, X \cap Y \neq \emptyset$.
2. зависимость между элементами двух разных множеств.
3. конъюнкция $X \& Y$, где $X, Y \subseteq I, X \cap Y \neq \emptyset$.
4. независимость между элементами двух разных множеств.

Анализ медицинских данных Вопрос_143

Что такое "шум" в контексте данных?

1. ✓ Случайные ошибки или отклонения в данных
2. Данные, собранные с помощью некорректных методов
3. Данные, не относящиеся к решаемой задаче
4. Дубликаты данных

Анализ медицинских данных Вопрос_144

Какой метод масштабирования данных превращает среднее значение в 0, стандартное отклонение в 1?

1. ✓ StandardScaler
2. MinMaxScaler
3. RobustScaler
4. Normalizer

Анализ медицинских данных Вопрос_145

Какие существуют типы регуляризации?

1. ✓ L1 и L2
2. Ridge и Lasso
3. MinMax и Standard
4. Linear и Logistic

Анализ медицинских данных Вопрос_146

Какой метод pandas используется для кодирования категориальных переменных?

1. ✓ get_dummies()
2. label_encoder()
3. one_hot_encoder()
4. scale()

Анализ медицинских данных Вопрос_147

Что такое мультиколлинеарность?

1. ✓ Наличие сильной корреляции между предикторными переменными
2. Наличие выбросов в наборе данных
3. Наличие пропущенных значений в наборе данных
4. Наличие категориальных переменных в наборе данных

Анализ медицинских данных Вопрос_148

Что такое "feature engineering"?

1. ✓ Создание новых признаков на основе существующих
2. Удаление лишних признаков
3. Масштабирование признаков
4. Заполнение пропущенных значений

Анализ медицинских данных Вопрос_149

Что такое "pipeline" в scikit-learn?

1. ✓ Метод оценки качества модели
2. Последовательность шагов предобработки и моделирования
3. Метод визуализации данных
4. Метод оптимизации гиперпараметров

Анализ медицинских данных Вопрос_15

Какой метод используется для выявления выбросов в данных?

1. ✓ Метод главных компонент
2. Анализ временных рядов
3. Метод наименьших квадратов
4. Метод k-ближайших соседей

Анализ медицинских данных Вопрос_150

Что такое "аугментация данных"?

1. ✓ Искусственное увеличение объема данных
2. Удаление шума из данных
3. Масштабирование данных
4. Заполнение пропущенных значений

Анализ медицинских данных Вопрос_151

Какой метод pandas используется для обнаружения мультиколлинеарности?

1. ✓ corr()
2. dropna()
3. get_dummies()
4. clip()

Анализ медицинских данных Вопрос_152

Как получить среднее значение столбца "temperature" в pandas DataFrame?

1. ✓ data["temperature"].average()
2. data.average("temperature")
3. data["temperature"].mean()
4. data.mean("temperature")

Анализ медицинских данных Вопрос_153

Какой метод используется для разделения набора данных на обучающий и тестовый наборы?

1. ✓ train_test_split()
2. cross_val_score()
3. GridSearchCV()
4. StratifiedKFold()

Анализ медицинских данных Вопрос_155

Какой метод используется для преобразования типа данных столбца в Pandas DataFrame?

1. ✓ astype()
2. convert()
3. transform()
4. change()

Анализ медицинских данных Вопрос_156

Что такое OLAP? "

1. ✓ набор инструментов, которые работают вместе, чтобы обеспечить расширенную среду анализа данных для извлечения, обработки и моделирования данных из ХД." "
2. специализированная БД, в которой хранятся данные в формате, оптимизированном

для поддержки принятия решений. ХД содержит исторические данные, полученные из оперативных БД, а также данные из других внешних источников " "

3. комплекс языковых и программных средств, предназначенный для создания, ведения и совместного использования БД многими пользователями. " "

4. представляет собой совокупность специальным образом организованных данных, хранимых в памяти ВС и отображающих состояние объектов и их взаимосвязей в рассматриваемой предметной области."

Анализ медицинских данных Вопрос_157

Какого типа слияние включает только те строки, которые совпадают на обоих листах?

1. ✓ Inner Merge
2. Left Merge
3. Right Merge
4. Outer Merge

Анализ медицинских данных Вопрос_159

Какой метод позволяет увеличить размер редкого или малого класса?

1. ✓ Up-sampling
2. Down-sampling
3. Data generation
4. Sample weights

Анализ медицинских данных Вопрос_16

Идентификация точек данных со значительным отклонением от ожидаемого или нормального диапазона значений называется

1. ✓ Обнаружением и обработкой выбросов
2. Сокращением размерности данных
3. Обнаружением аномалий в формате точек данных
4. Обнаружением категориальных признаков

Анализ медицинских данных Вопрос_160

Какая особенность CatBoost позволяет эффективно работать с нечисловыми данными?

1. ✓ Поддержка категориальных функций
2. Быстрая и масштабируемая версия графического процессора
3. Применение обученной модели быстро и эффективно
4. Повышенная точность за счет уменьшения переобучения

Анализ медицинских данных Вопрос_161

Какой метод анализа временных рядов используется для разложения данных на составляющие части, такие как тренд, сезонность и остаток?

1. ✓ Методы декомпозиции
2. Методы сглаживания
3. Статистический анализ
4. Визуальный анализ

Анализ медицинских данных Вопрос_162

Какое из следующих определений соответствует методам сглаживания в анализе временных рядов?

1. ✓ Использование различных техник для удаления шума и выявления общих трендов в данных.
2. Использование графиков для визуального изучения трендов, сезонности и других характеристик.
3. Использование статистических методов для описания и моделирования временных рядов.
4. Использование моделей и алгоритмов для прогнозирования будущих значений.

Анализ медицинских данных Вопрос_163

Какое из следующих понятий реляционной модели определяет количество строк в таблице?

1. ✓ Кардинальность
2. Кортеж
3. Домен
4. Сущность

Анализ медицинских данных Вопрос_164

Что такое атрибуты в реляционной модели?

1. ✓ Свойства, характеризующие сущность, и соответствующие столбцам таблицы.
2. Строки таблицы.
3. Количество строк в таблице.
4. Значения, принимаемые свойством сущности.

Анализ медицинских данных Вопрос_165

Какое из следующих понятий реляционной модели используется для связывания двух отношений и обеспечения целостности данных?

1. ✓ Внешний ключ (FK)
2. Первичный ключ (PK)
3. Кортеж
4. Домен

Анализ медицинских данных Вопрос_166

Какой из следующих операторов языка манипулирования данными (DML) используется для обновления существующих данных в таблице?

1. ✓ UPDATE
2. SELECT
3. INSERT
4. DELETE

Анализ медицинских данных Вопрос_167

Какой из следующих операторов языка манипулирования данными (DML) используется для выбора всех строк из таблицы, удовлетворяющих определенному условию?

1. ✓ SELECT WHERE
2. SELECT ALL
3. SELECT DISTINCT
4. SELECT GROUP BY

Анализ медицинских данных Вопрос_168

Какой из следующих алгоритмов бустинга использует простые деревья решений с одним разбиением, известные как «пни решений слабых учеников»?

1. ✓ AdaBoostClassifier
2. GradientBoosting Classifier
3. XGBoost
4. LightGBM

Анализ медицинских данных Вопрос_169

Какая функция pandas используется для загрузки данных из CSV файла?

1. ✓ read_csv()
2. load_csv()
3. import_csv()
4. get_csv()

Анализ медицинских данных Вопрос_17

Какой метод анализа данных используется для выявления сходств и различий между группами симптомов заболевания?

1. ✓ Кластерный анализ
2. Факторный анализ
3. Дискриминантный анализ
4. Регрессионный анализ

Анализ медицинских данных Вопрос_170

Какой метод pandas используется для выбора уникальных значений из столбца?

1. ✓ unique()
2. distinct()
3. uniques()
4. get_unique()

Анализ медицинских данных Вопрос_171

Как сохранить график matplotlib в файл?

1. ✓ savefig()
2. export()
3. write()
4. dump()

Анализ медицинских данных Вопрос_172

Какой метод pandas используется для удаления строк или столбцов?

1. ✓ drop()
2. remove()
3. delete()
4. filter()

Анализ медицинских данных Вопрос_173

Что такое "масштабирование" данных и когда оно может быть полезным?

1. ✓ Изменение диапазона значений для улучшения визуализации
2. Удаление выбросов
3. Преобразование категориальных данных в числовые
4. Заполнение пропущенных значений

Анализ медицинских данных Вопрос_174

Какой тип диаграммы лучше всего подходит для визуализации распределения данных?

1. ✓ Гистограмма
2. Диаграмма рассеяния
3. Линейный график
4. Круговая диаграмма

Анализ медицинских данных Вопрос_175

Что такое кодирование категориальных признаков?

1. ✓ Преобразование текстовых значений в числовые
2. Удаление категориальных признаков из данных
3. Замена категориальных признаков числовыми значениями
4. Сортировка категориальных признаков по алфавиту

Анализ медицинских данных Вопрос_176

В чем заключается разница между разведочным и объяснительным анализом данных?

1. ✓ Разведочный анализ направлен на поиск закономерностей, а объяснительный - на подтверждение гипотез
2. Разведочный анализ использует визуализацию, а объяснительный - нет
3. Разведочный анализ сложнее, чем объяснительный
4. Разведочный анализ не использует статистические методы

Анализ медицинских данных Вопрос_177

Какой метод pandas используется для применения функции к каждой строке/столбцу?

1. ✓ `apply()`
2. `map()`
3. `transform()`
4. `iterate()`

Анализ медицинских данных Вопрос_178

Какая функция matplotlib используется для создания графика с несколькими осями?

1. ✓ `subplot()`
2. `multi_plot()`
3. `subplots()`
4. `axes()`

Анализ медицинских данных Вопрос_179

Как добавить подписи к осям графика seaborn?

1. ✓ Используя функции `set_xlabel()` и `set_ylabel()`
2. Используя аргументы "xlabel" и "ylabel"
3. Используя функции `label_x()` и `label_y()`
4. Используя аргументы "xlab" и "ylab"

Анализ медицинских данных Вопрос_18

Регрессионный анализ используется для

1. ✓ Прогнозирования будущих значений признаков
2. Выявления зависимостей между признаками
3. Выделения групп схожих между собой признаков
4. Определения причинно-следственных связей между признаками

Анализ медицинских данных Вопрос_180

К операторам управления транзакциями в SQL относятся:

1. ✓ COMMIT, ROLLBACK, SAVEPOINT, SET TRANSACTION
2. COMMIT, ROLLBACK, INSERT, SET TRANSACTION
3. SELECT, ROLLBACK, INSERT, DROP
4. ROLLBACK, SAVEPOINT, SET TRANSACTION

Анализ медицинских данных Вопрос_181

Агрегатная функция AVG в SQL:

1. ✓ Возвращает среднее значение
2. Возвращает среднее значение среди неповторяющихся элементов
3. Возвращает отношение минимального значения к максимальному
4. Возвращает сумму значений

Анализ медицинских данных Вопрос_182

`DataFrame.drop_duplicates()`

1. ✓ Удалит все повторяющиеся строки
2. Заполнит нулевые значения
3. Удалит все строки содержащие одинаковое значение во всех колонках
4. Удалить не повторяющиеся строки

Анализ медицинских данных Вопрос_185

Каков правильный синтаксис для возврата первой строки в Pandas DataFrame?

1. ✓ `df.loc[0]`
2. `df.get(0)`
3. `DF[0]`
4. `DF["0"]`

Анализ медицинских данных Вопрос_186

Что делает метод head() Pandas?

1. ✓ Возвращает заголовки и указанное количество строк, начиная сверху
2. Возвращает заголовки
3. Возвращает указанное количество строк, начиная сверху
4. Возвращает первую строку

Анализ медицинских данных Вопрос_187

Каков правильный метод Pandas для возврата последних строк?

1. ✓ tail()
2. foot()
3. last()
4. end()

Анализ медицинских данных Вопрос_188

Каков правильный метод удаления дубликатов из Pandas DataFrame?

1. ✓ df.drop_duplicates()
2. df.remove_duplicates()
3. df.delete_duplicates()
4. df.duplicates()

Анализ медицинских данных Вопрос_189

Какой синтаксис правильный, чтобы вернуть весь DataFrame

1. ✓ df.print_full()
2. df.print_full()
3. df.dump()
4. df.Full()

Анализ медицинских данных Вопрос_19

Какой из перечисленных методов является методом "без учителя"?

1. ✓ Кластерный анализ
2. Нейронные сети
3. Линейная регрессия
4. Решающие деревья

Анализ медицинских данных Вопрос_190

Какая функция Pandas подходит для загрузки файлов JSON в DataFrame?

1. ✓ read_json()
2. ReadFile()
3. read_file()
4. ReadJSON()

Анализ медицинских данных Вопрос_191

Какая функция не относится к библиотеке matplotlib.pyplot(plt)?

1. ✓ .heatmap()
2. .figure()
3. .subplots()
4. .ylabel()

Анализ медицинских данных Вопрос_192

Какой оператор создает представление БД?

1. ✓ CREATE VIEW
2. ALTER TABLE
3. DROP VIEW
4. CREATE TABLE

Анализ медицинских данных Вопрос_193

Выберите неверное утверждение о MitoSheet?

1. ✓ Не позволяет получить доступ у сводной статистике
2. Позволяет вам переключаться между анализом данных в Python и манипулированием данными

3. Каждая вкладка представляет собой dataframe Pandas

4. Использует графическую библиотеку Plotly Express

Анализ медицинских данных Вопрос_194

Какая метрика относится к регрессионной?

1. ✓ R Squared
2. Recall
3. Accuracy
4. Precision

Анализ медицинских данных Вопрос_195

Какая функция не вычисляет одномерные статистики (univariate)?

1. ✓ df.info()
2. df.describe()
3. df[feature].value_counts()
4. np.bincount(y)

Анализ медицинских данных Вопрос_196

Из какого ансамблевого метода получится случайный лес (RandomForest) если использовать в качестве слабого ученика деревья решений?

1. ✓ Bagging
2. Boosting
3. Voting
4. Stacking

Анализ медицинских данных Вопрос_197

Параметр RandomForest, определяющий функцию, измеряющую качество разбиения ветки дерева?

1. ✓ criterion
2. bootstrap
3. oob_score
4. random_state

Анализ медицинских данных Вопрос_198

В алгоритме CatBoosting пропущенные числовые значения обрабатываются:

1. ✓ Устанавливается минимальное значение для этой функции (default)
2. Направление ветвления изучено для пропущенных значений
3. Пропускает точку данных во время разделения, распределяет позже листьям
4. Не предусмотрено

Анализ медицинских данных Вопрос_199

Величина, показывающая как часто в записях базы данных, содержащих X, одновременно присутствует Y, т.е. оценивает вероятность того, что из наличия записи набора X следует наличие в ней набора Y

1. ✓ Достоверность(confidence)
2. Поддержка(support)
3. Улучшение(lift)
4. Алгоритм Apriori

Анализ медицинских данных Вопрос_2

Метрика, характеризующая качество модели, агрегированное по всем классам

1. ✓ Accuracy
2. Precision

3. Recall

4. F1-мера

Анализ медицинских данных Вопрос_20

Анализ временных рядов - это

1. ✓ Метод анализа, используемый для прогнозирования будущих значений переменной на основе ее прошлых значений
2. Метод анализа, используемый для определения различий между группами данных
3. Метод анализа, используемый для определения наиболее значимых переменных в наборе данных
4. Метод анализа, используемый для оценки взаимосвязи между несколькими переменными

Анализ медицинских данных Вопрос_200

Агломеративные методы кластеризации выделяются по признаку:

1. ✓ По способу обработки данных
2. По способу анализа данных
3. По количеству применений алгоритмов кластеризации
4. По возможности расширения объема обрабатываемых данных

Анализ медицинских данных Вопрос_201

Какой этап алгоритма обучения нейросети выполняется раньше других?

1. ✓ Задаются шаг обучения α ($0 < \alpha < 1$) и желаемая среднеквадратичная ошибка сети
2. Подаются последовательно образы из обучающей выборки на вход нейронной сети.
3. Инициализируются случайным образом весовые коэффициенты w и пороговые b , значения НС.
4. Вычисляется суммарная ошибка нейронной сети

Анализ медицинских данных Вопрос_202

Какой фреймворк о глубинном обучении не общего назначения, делает основной опрос на машинное зрение?

1. ✓ Caffe
2. Keras
3. TensorFlow
4. Theano

Анализ медицинских данных Вопрос_203

Код `a = model.fit(x_train, y_train, batch_size=200, epochs=100, verbose=1)`:

1. ✓ Обучает нейронную сеть
2. Описывает архитектуру нейронной сети
3. Компилирует сеть
4. Не относится к нейросетям

Анализ медицинских данных Вопрос_204

Какой оператор используется для вставки новых записей в SQL?

1. ✓ INSERT INTO
2. CREATE VIEW
3. ALTER TABLE
4. UPDATE

Анализ медицинских данных Вопрос_205

Что означает конструкция NOT NULL в sql-запросах при создании таблицы БД

1. ✓ означает, что в этом столбце должно быть определено значение
2. означает, что в этом столбце не должно быть определено значение
3. означает, что в этом столбце должно быть пустое значение
4. означает, что в этом столбце должен быть счетчик

Анализ медицинских данных Вопрос_206

Рекомендации - это..?

1. ✓ Предоставление пользователям того, в чём они могут быть наиболее заинтересованы.
2. Выяснение того, что это такое.
3. Помощь пользователям в поиске наиболее релевантной вещи.
4. Поиск необычных вещей.

Анализ медицинских данных Вопрос_207

Какие есть группы методов по способу анализа данных?

1. ✓ Четкие и нечеткие.
2. Быстрые и медленные.
3. Высокие и низкие.
4. Нет верного ответа.

Анализ медицинских данных Вопрос_208

Какие есть группы методов по времени выполнения кластеризации?

1. ✓ Поточковые (on-line) и не поточковые (off-line).
2. Четкие и нечеткие.
3. Иерархические и неиерархические.
4. Нет верного ответа.

Анализ медицинских данных Вопрос_209

Что такое Heat Maps?

1. ✓ Это тип визуализации, применяемый, когда нам нужно найти зависимые переменные
2. Это тип диаграммы, который используется для визуализации географических данных
3. Heat Maps всегда показывают данные в виде цветов от зеленого до желтого
4. Это тип диаграммы, который используется для визуализации временных рядов

Анализ медицинских данных Вопрос_21

Какой метод используется для оценки параметров временных рядов?

1. ✓ Экспоненциального сглаживания
2. Метод наименьших квадратов
3. Метод опорных векторов
4. Метод главных компонент

Анализ медицинских данных Вопрос_210

Выберите верное утверждение относительно гиперпараметров

1. ✓ Гиперпараметры определяют поведение и структуру модели машинного обучения и влияют на ее производительность
2. Гиперпараметры обучаются из данных в процессе обучения модели
3. Настройка гиперпараметров всегда приводит к улучшению производительности модели
4. Гиперпараметры не влияют на сложность модели

Анализ медицинских данных Вопрос_211

Что такое метод декомпозиции?

1. ✓ Разложение на составляющие части, такие как тренд, сезонность и остаток
2. Использование различных техник для удаления шума и выявления общих трендов в данных
3. Построение графиков для визуального изучения трендов, сезонности и других характеристик
4. Использование моделей и алгоритмов для прогнозирования будущих значений

Анализ медицинских данных Вопрос_212

Кто ввел понятие EDA?

1. ✓ Джон Тьюки
2. Соломон Лефшец
3. Джон Нейман
4. Эрик Белл

Анализ медицинских данных Вопрос_213

Какие цели формулирует Джон Тьюки для разведочного анализа данных?

1. ✓ Максимальное "проникновение" в данные
2. Минимизация использования визуализаций
3. Исключение всех аномалий из данных
4. Ограничение доступа к данным

Анализ медицинских данных Вопрос_214

Какие параметры может принимать функция потерь в методах бустинга?

1. ✓ Вес, применяемый к каждому классификатору на каждой итерации повышения
2. Максимальное количество оценок
3. Количество этапов повышения
4. Скорость обучения

Анализ медицинских данных Вопрос_215

Какие языки программирования поддерживает библиотека XGBoost?

1. ✓ C++, Python, R
2. C++, Python, Java
3. Python, R, Julia
4. Java, R, Julia

Анализ медицинских данных Вопрос_216

Какие методы прогнозирования временных рядов используются для прогнозирования будущих значений?

1. ✓ Все нижеперечисленное
2. EWMA (Exponentially Weighted Moving Average)
3. LSTM (Long Short-Term Memory)
4. ARIMA (Autoregressive Integrated Moving Average)

Анализ медицинских данных Вопрос_217

Какие методы классификации данных относятся к алгоритмам для прогнозирования временных рядов?

1. ✓ LSTM (Long Short-Term Memory)
2. Обучение HC
3. Automated Machine Learning
4. K-means

Анализ медицинских данных Вопрос_218

Какая функция Vamboolib используется для разделения строки на несколько столбцов?

1. ✓ Split
2. Divide
3. Separate
4. Substring

Анализ медицинских данных Вопрос_219

Какая функция Vamboolib используется для объединения двух наборов данных?

1. ✓ Join
2. Join
3. Combine

4. Concat

Анализ медицинских данных Вопрос_22

Какой метод может быть выбран для анализ динамики заболевания с целью планирования тактики лечения?

1. ✓ анализ временных рядов
2. регрессионный анализ
3. корреляционный анализ
4. кластерный анализ

Анализ медицинских данных Вопрос_220

Какой подход можно использовать для решения проблемы несбалансированных наборов данных?

1. ✓ Up-sampling (увеличение размера редкого класса) или Down-sampling (уменьшение размера доминирующего класса)
2. Удаление признаков с отсутствующими значениями
3. Нормализация признаков к одному масштабу
4. Использование метода опорных векторов (SVM)

Анализ медицинских данных Вопрос_221

Что такое LSTM в контексте анализа временных рядов?

1. ✓ Тип рекуррентной нейронной сети, часто используемой для прогнозирования временных рядов
2. Алгоритм поиска ассоциативных правил
3. Алгоритм поиска ассоциативных правил
4. Функция активации в нейронных сетях

Анализ медицинских данных Вопрос_222

Что показывает матрица корреляции?

1. ✓ линейную зависимость между признаками
2. пропорциональную зависимость между признаками
3. корректность взаимосвязи признаков
4. время обучения хорошей модели

Анализ медицинских данных Вопрос_223

Параметры алгоритмов машинного обучения, которые влияют на структуру алгоритмов и производительность (качество) моделей

1. ✓ Hyperparameters
2. ML Parameters
3. Optimization Parameters
4. General Parameters

Анализ медицинских данных Вопрос_224

На чём основан структурированный язык запросов SQL?

1. ✓ На реляционном исчислении
2. На операциях над таблицами
3. На операторах
4. На система управления базами данных (СУБД)

Анализ медицинских данных Вопрос_225

Что такое центроид в алгоритме кластеризации K-means?

1. ✓ Центр тяжести кластеров
2. Центр каждого кластера
3. Медиана распределения кластеров
4. Среднее значение объектов исходной выборки

Анализ медицинских данных Вопрос_226

В какой системе AutoML отсутствует предобработка данных?

1. ✓ TuPAQ
2. ML-Plan
3. Auto-WEKA
4. AlphaD3M

Анализ медицинских данных Вопрос_227

Что не является open-source фреймворком о глубинном обучении?

1. ✓ Apache Spark
2. Torch
3. Theano
4. Caffe

Анализ медицинских данных Вопрос_228

Выберите метод настройки нейросетевой модели для обучения

1. ✓ compile
2. fit
3. evaluate
4. predict

Анализ медицинских данных Вопрос_229

Что в ML обозначают как 1d, 2d, ..., nd?

1. ✓ Количество функций
2. Количество слоёв
3. Количество кластеров
4. Количество входных данных

Анализ медицинских данных Вопрос_23

ARIMA-модель используется

1. ✓ для прогнозирования временных рядов
2. для анализа связей между несколькими переменными
3. для определения наиболее значимых переменных в наборе данных
4. для определения различий между группами данных

Анализ медицинских данных Вопрос_230

Какой операнд позволяет выделять в результирующем множестве записей группы?

1. ✓ WHERE
2. HAVING
3. ORDER BY
4. ORDER BY

Анализ медицинских данных Вопрос_231

Кто ввел понятие "машинное обучение"?

1. ✓ Артур Сэмюэл
2. Алан Тьюринг
3. Джон Маккарти
4. Гарольд Саксон Паперт

Анализ медицинских данных Вопрос_232

Что не входит в основные параметры случайного дерева?

1. ✓ размер корня
2. глубина дерева
3. число признаков, по которым ищется лучшее разбиение в дереве
4. число объектов в листе

Анализ медицинских данных Вопрос_233

Какой язык не поддерживает библиотека LightGBM?

1. ✓ PascalABC
2. python
3. c#
4. c++

Анализ медицинских данных Вопрос_234

Операнд HAVING действует совместно с операндом ...

1. ✓ GROUP BY
2. ORDER BY
3. WHERE
4. DISTINCT

Анализ медицинских данных Вопрос_235

К алгоритмам обучения с учителем НЕ относится

1. ✓ Алгоритм k-средних
2. Линейные алгоритмы
3. Нейронные сети
4. Деревья решений

Анализ медицинских данных Вопрос_236

К фазам цикла исследования данных НЕ относится

1. ✓ Загрузка данных
2. Понимание бизнес-целей
3. Оценка
4. Подготовка данных

Анализ медицинских данных Вопрос_237

К признакам деления методов кластеризации на группы НЕ относится

1. ✓ По способу группировки данных
2. По способу обработки данных
3. По времени выполнения кластеризации
4. По количеству применений алгоритмов кластеризации

Анализ медицинских данных Вопрос_24

Что такое стационарность временного ряда?

1. ✓ статистические свойства временного ряда не меняются со временем
2. временной ряд содержит тренд и/или сезонность
3. временной ряд содержит случайные колебания
4. значения временного ряда имеют нормальное распределение

Анализ медицинских данных Вопрос_240

Что происходит во 4 шаге обучения при бустинге?

1. ✓ Алгоритм повторяет 2 и 3 шага до тех пор, пока ошибки обучения не опустятся ниже определённого порога
2. Алгоритм повторяет 1 и 2 шага до тех пор, пока ошибки обучения не опустятся ниже определённого порога
3. Алгоритм повторяет все прошлые шаги до тех пор, пока ошибки обучения не опустятся ниже определённого порога
4. Алгоритм повторяет 3 шаг до тех пор, пока ошибки обучения не опустятся ниже определённого порога

Анализ медицинских данных Вопрос_241

О каком явлении в оценке моделей идёт речь? Высокая точность даже тогда, когда мало истинных срабатываний.

1. ✓ Парадокс высокой точности
2. Парадокс аномальной точности

3. Парадокс полноты

4. Парадокс завышенной точности

Анализ медицинских данных Вопрос_242

Как называется метод присвоения более высокого веса редким классам и более низкого веса доминирующим классам?

1. ✓ Sample weights
2. Up-sampling
3. Down-sampling
4. Weight reversing

Анализ медицинских данных Вопрос_243

Почему ансамблевые методы так популярны?

1. ✓ Они создают сильную модель, комбинируя прогнозы слабых моделей
2. Они позволяют ускорить разработку модель за счёт слабых моделей
3. Они улучшают работоспособность данной модели за счёт использования слабых моделей
4. Они перераспределяют нагрузку на методы слабых моделей, тем самым ускоряя обучение

Анализ медицинских данных Вопрос_244

Обучающую выборку делится на две части: на первой обучают базовые алгоритмы, затем получают их ответы на второй части и на тестовой выборке. Как называется подобная схема стекинга?

1. ✓ Blending
2. Blendering
3. Mixing
4. Integrating

Анализ медицинских данных Вопрос_245

Что такое бутстрэп в беггинге?

1. ✓ Случайный выбор данных из датасета и представление их в модель
2. Определённый выбор данных из датасета и представление их в модель
3. Случайный выбор данных из нескольких датасетов и представление их в модель
4. Выбор определённых данных из нескольких датасетов и представление их в модель

Анализ медицинских данных Вопрос_246

Какая операция реляционной алгебры удаляет дубликаты из результата?

1. ✓ Операция проекции (π) может удалять дубликаты, оставляя только уникальные строки в результате.
2. Операция объединения
3. Операция пересечения
4. Операция разности

Анализ медицинских данных Вопрос_247

Какая операция реляционной алгебры используется для фильтрации строк?

1. ✓ Операция выборки (σ) используется для фильтрации строк по определенному критерию.
2. Операция умножения
3. Операция деления
4. Операция вычитания

Анализ медицинских данных Вопрос_248

Какие методы используются для удаления дубликатов в базе данных?

1. ✓ Для удаления дубликатов можно использовать операцию проекции с уникальными атрибутами или специальные SQL-команды, такие как DISTINCT.

2. Использование случайного выбора.
3. Применение фильтрации по дате создания.
4. Удаление всех записей и создание новой таблицы.

Анализ медицинских данных Вопрос_249

Как в MitoSheet осуществляется импорт данных?

1. ✓ Данные можно импортировать в MitoSheet через интерфейс, который поддерживает чтение файлов в форматах CSV, Excel и других.
2. Данные в MitoSheet загружаются через Bluetooth.
3. Данные в MitoSheet передаются с помощью QR-кодов.
4. Данные в MitoSheet импортируются sql-запросами

Анализ медицинских данных Вопрос_25

Что такое регуляризация в машинном обучении?

1. ✓ уменьшение ошибки модели за счёт штрафа за большие значения коэффициентов
2. увеличение ошибки модели за счёт добавления шума к данным
3. сокращение размерности данных для ускорения обучения
4. увеличение количества признаков для улучшения точности предсказаний

Анализ медицинских данных Вопрос_250

Какие возможности предоставляет MitoSheet для очистки данных?

1. ✓ MitoSheet предлагает инструменты для удаления дубликатов, обработки пропущенных значений, фильтрации данных и преобразования типов данных.
2. MitoSheet очищает данные с помощью антивирусных сканирований.
3. MitoSheet использует физическую очистку данных с диска.
4. MitoSheet применяет sql-запросы для очистки данных.

Анализ медицинских данных Вопрос_251

Что представляет собой операция пересечения в реляционной алгебре?

1. ✓ Получение общих строк из двух таблиц
2. Выборка определенного столбца
3. Объединение таблиц
4. Выборка строк, удовлетворяющих условию

Анализ медицинских данных Вопрос_252

Зачем используются алгоритмы временных рядов в прогнозировании?

1. ✓ Для предсказания будущих значений на основе паттернов из прошлых данных
2. Для обработки статических данных
3. Для классификации изображений по времени
4. Для анализа текстовых данных

Анализ медицинских данных Вопрос_253

Выберите основные операторы языка манипулирования данными SQL

1. ✓ SELECT, UPDATE, INSERT и DELETE.
2. ADD, MODIFY, SEARCH и CLEAR.
3. CREATE, DROP, RENAME и LIST.
4. PRINT, COPY, PASTE и CUT.

Анализ медицинских данных Вопрос_254

Почему важно обнаруживать и удалять выбросы при проведении EDA?

1. ✓ выбросы могут привести к менее точным моделям данных. Удаление выбросов помогает улучшить качество анализа и моделирования данных.
2. выбросы обогащают данные
3. их обнаружение не имеет значения
4. Удаление выбросов может исказить данные, поэтому лучше оставить их без изменений

Анализ медицинских данных Вопрос_255

Какой метод можно использовать для замены пропущенных числовых значений средним значением столбца?

1. ✓ `df['col'].fillna(df['col'].mean())`
2. `df['col'].fillna(df['col'].mode())`
3. `df.dropna()`
4. `df.fillna(0)`

Анализ медицинских данных Вопрос_256

Как обучить модель `VotingClassifier` на обучающем наборе данных

1. ✓ `VC.fit(X_train, y_train)`
2. `VC.fit(X_test, y_train)`
3. `VC.predict(X_test)`
4. `VC.predict(X_train)`

Анализ медицинских данных Вопрос_257

Что определяет параметр `min_confidence` при поиске ассоциативных правил?

1. ✓ фильтрует те правила, достоверность которых превышает порог достоверности, указанный в параметре
2. определяет минимальное значение подъема для правил
3. используется для выбора элементов со значениями поддержки
4. указывает минимальное количество элементов, которые вы хотите использовать в своих правилах

Анализ медицинских данных Вопрос_258

Какой параметр в `RandomForestRegressor` отвечает за количество деревьев?

1. ✓ `n_estimators`
2. `criterion`
3. `max_features`
4. `max_depth`

Анализ медицинских данных Вопрос_259

Что определяет параметр `min_length` при формировании правил?

1. ✓ Минимальное количество элементов, которые нужно использовать в правиле
2. Максимальное количество элементов в правиле
3. Порог для поддержки элементов
4. Порог для достоверности правил

Анализ медицинских данных Вопрос_260

Модель переобучена, если

1. ✓ модель слишком сложная и точно подстраивается под обучающие данные, что приводит к плохой обобщающей способности на новых данных
2. модель слишком простая и не может точно предсказывать значения целевой переменной
3. модель не успела получить достаточно знаний из данных
4. модель получила некорректные данные и не может обучиться

Анализ медицинских данных Вопрос_261

Какой параметр используется для выбора элементов со значениями поддержки, превышающими указанное значение?

1. ✓ `min_support`
2. `min_confidence`
3. `min_lift`
4. `min_length`

Анализ медицинских данных Вопрос_262

Какие меры регуляризации применяются в классическом градиентном бустинге?

1. ✓ Ограничение на структуру дерева, контролирование темпа обучения и увеличение "непохожести" деревьев за счет рандомизации
2. В классическом градиентном бустинге не применяются меры регуляризации
3. Применяются только меры регуляризации, связанные с ограничением на структуру дерева
4. Применяются только меры регуляризации, связанные с контролированием темпа обучения

Анализ медицинских данных Вопрос_262

Какая метрика используется для измерения способности модели прогнозировать положительный результат в контексте классификации?

1. ✓ Полнота
2. Точность
3. F1-Score
4. ROC AUC

Анализ медицинских данных Вопрос_263

В результате выполнения оператора `df.tail(5)`?

1. ✓ Выводятся последние пять строк из DataFrame `df`.
2. Отфильтровываются строки DataFrame `df`, оставляя только те, которые содержат последние пять значений.
3. Изменяются последние пять строк в DataFrame `df` согласно заданным критериям.
4. Создается новый DataFrame, содержащий последние пять строк из DataFrame `df`.

Анализ медицинских данных Вопрос_264

Что не является видом ансамблевых методов?

1. ✓ Линейная регрессия
2. Стэкинг
3. Бустинг
4. Бэггинг

Анализ медицинских данных Вопрос_265

Для каких целей можно использовать Apache Hadoop?

1. ✓ Для анализа "озер данных", индексации веб-сайтов, финансового анализа, научных исследований
2. Для создания шины микросервисов и анализа в реальном времени
3. Для выполнения мгновенных аналитических запросов на структурированных данных
4. Для мониторинга транспорта

Анализ медицинских данных Вопрос_266

Какой метод может быть использован для прогнозирования недостающих значений из полных выборок с помощью методов машинного обучения?

1. ✓ Advanced imputation
2. Замена пропущенных значений на среднее
3. Удаление строк с пропущенными значениями
4. Placeholder

Анализ медицинских данных Вопрос_267

Какой класс необходимо импортировать для использования метода классификации с голосованием?

1. ✓ VotingClassifier из модуля `sklearn.ensemble`
2. VotingRegressor из модуля `sklearn.ensemble`
3. ClassifierMixin из модуля `sklearn.base`

4. VotingModel из модуля sklearn.model_selection

Анализ медицинских данных Вопрос_268

Какой класс необходимо импортировать для использования метода классификации с использованием бэггинга?

1. ✓ BaggingClassifier из модуля sklearn.ensemble
2. BaggingRegressor из модуля sklearn.ensemble
3. ClassifierMixin из модуля sklearn.base
4. RegressorMixin из модуля sklearn.base

Анализ медицинских данных Вопрос_27

Среднеквадратичная ошибка (в том числе в процентах), корень из среднеквадратичной ошибки, средняя абсолютная ошибка, средняя относительная ошибка, среднеквадратичная логарифмическая ошибка входят в аппарат оценки качества

1. ✓ регрессионных моделей
2. моделей классификации
3. моделей нейронных сетей
4. моделей временных рядов

Анализ медицинских данных Вопрос_271

Как задать критерий для разделения в узлах деревьев решений в RandomForestClassifier?

1. ✓ Указать параметр criterion при создании экземпляра класса и передать значение "gini" или "entropy".
2. Использовать метод set_criterion после создания экземпляра класса.
3. Указать параметр max_features при создании экземпляра класса и передать значение "gini" или "entropy".
4. Использовать метод set_params после создания экземпляра класса.

Анализ медицинских данных Вопрос_272

Какой параметр определяет функцию, которая измеряет качество разбиения ветки дерева в RandomForestRegressor?

1. ✓ max_depth
2. min_samples_split
3. min_samples_leaf
4. criterion

Анализ медицинских данных Вопрос_273

Что представляет собой первый шаг алгоритма K-means?

1. ✓ Выбор случайных центров кластеров из исходной выборки
2. Определение числа кластеров, которые будут сформированы
3. Определение ближайшего к каждой записи центра кластера
4. Вычисление центроидов кластеров

Анализ медицинских данных Вопрос_274

Каким образом вычисляются центроиды в алгоритме K-means?

1. ✓ Выбираются случайные точки из исходной выборки
2. Находится среднее значение всех точек в кластере
3. Находится точка, которая находится на максимальном расстоянии от всех остальных точек в кластере
4. Определяется такая точка, которая находится на минимальном расстоянии от всех остальных точек в кластере

Анализ медицинских данных Вопрос_275

Как задать параметр, отвечающий за использование out-of-bag объектов для оценки R^2 в RandomForestRegressor?

1. ✓ n_jobs
2. random_state
3. oob_score
4. bootstrap

Анализ медицинских данных Вопрос_277

Какой класс необходимо импортировать для использования метода классификации с использованием случайного леса (Random Forest)?

1. ✓ RandomForestClassifier из модуля sklearn.ensemble
2. RandomForestRegressor из модуля sklearn.ensemble
3. ClassifierMixin из модуля sklearn.base
4. RegressorMixin из модуля sklearn.base

Анализ медицинских данных Вопрос_279

Какая характеристика временных рядов описывает случайные колебания данных?

1. ✓ Шум
2. Тренд
3. Стационарность
4. Автокорреляция

Анализ медицинских данных Вопрос_28

Для его используется кросс-валидация?

1. ✓ для обучения модели на случайно выбранных подмножествах исходных данных, чтобы избежать переобучения
2. для поиска наилучших гиперпараметров модели путём последовательного перебора значений гиперпараметров и оценки качества модели на каждом наборе гиперпараметров
3. для использования нескольких моделей для улучшения качества предсказаний
4. для проверки качества модели на новых данных

Анализ медицинских данных Вопрос_280

Что происходит, если внешний ключ ссылается на несуществующее значение в связанной таблице(БД)?

1. ✓ Операция добавления или обновления будет отклонена, нарушение целостности ссылочной целостности
2. Значение во внешнем ключе заменяется на NULL
3. Будет создана новая запись в связанной таблице
4. Будет выдано предупреждение, но операция не будет прервана

Анализ медицинских данных Вопрос_281

Какое утверждение о внешнем ключе в базе данных верно?

1. ✓ Внешний ключ представляет собой атрибут, который ссылается на первичный ключ другой таблицы
2. Внешний ключ используется для создания уникальных значений в таблице
3. Внешний ключ используется для уникальной идентификации каждой записи в таблице
4. Внешний ключ не может быть использован в SQL запросах

Анализ медицинских данных Вопрос_282

Что такое параметры модели (weights)?

1. ✓ Набор чисел, встроенных в модель, который может предсказывать метки.
2. Данные, которые помогут делать прогнозы.
3. Количество функций.
4. То, что вы пытаетесь предсказать.

Анализ медицинских данных Вопрос_283

Что необходимо сделать для повышения качества стекинга?

1. ✓ Необходимо усреднить несколько блендингов с разными разбиениями обучения.
2. Необходимо исключить из ансамбля все модели, кроме одной, чтобы упростить процесс прогнозирования.
3. Необходимо использовать только один тип моделей без разнообразия.
4. Необходимо уменьшить количество обучающих данных, чтобы ускорить процесс обучения.

Анализ медицинских данных Вопрос_284

Какое из перечисленных ниже понятий представляет собой специализированную базу данных, в которой хранятся данные в формате, оптимизированном для поддержки принятия решений?

1. ✓ Хранилище данных
2. Система управления базами данных
3. База данных
4. Аналитическая обработка в реальном времени (OLAP)

Анализ медицинских данных Вопрос_285

Какая функция активации была использована в сверточной сети AlexNet, победившей на соревновании ImageNet в 2012 году?

1. ✓ ReLU
2. Sigmoid
3. Tanh
4. Leaky ReLU

Анализ медицинских данных Вопрос_286

Какая функция используется для расчета ошибки между реальными и полученными ответами в процессе обучения нейронной сети?

1. ✓ Кросс-энтропия
2. Среднеквадратичное отклонение
3. Экспоненциальная функция потерь
4. Расстояние Кульбака-Лейблера

Анализ медицинских данных Вопрос_287

Какое ключевое слово используется для изменения объектов БД?

1. ✓ ALTER
2. MODIFY
3. INTERCHANGE
4. CHANGE

Анализ медицинских данных Вопрос_288

Каким запросом можно добавить запись в таблицу T1?

1. ✓ INSERT INTO T1(p1,p4) VALUES ('text',now())
2. INSERT VALUES ('text',now()) INTO T1
3. INSERT T1 FROM VALUES ('text',now())
4. ADD T1(p1,p4) RECORD ('text',now())

Анализ медицинских данных Вопрос_289

Что включает в себя нормализация данных в базе данных?

1. ✓ Устранение противоречивостей и избыточности, упрощение таблицы
2. Увеличение объема хранимых данных для ускорения доступа к информации
3. Слияние нескольких таблиц в одну для упрощения структуры базы данных
4. Добавление данных для создания дополнительных резервных копий информации

Анализ медицинских данных Вопрос_29

Метод опорных векторов (SVM) использует

1. ✓ гиперплоскости для разделения классов
2. градиентный спуск для оптимизации весов модели
3. результаты анализа кластеров данных
4. случайный лес для классификации данных

Анализ медицинских данных Вопрос_290

Какой тип данных используется для хранения текстовой информации переменной длины в PostgreSQL?

1. ✓ TEXT
2. VARCHAR
3. CHAR
4. STRING

Анализ медицинских данных Вопрос_291

За что отвечает аргумент batch_size в градиентном бустинге?

1. ✓ Количество выборок на обновление градиента
2. Массив данных обучения
3. Массив целевых данных
4. Количество эпох для обучения модели

Анализ медицинских данных Вопрос_292

Что такое сущность в реляционной модели?

1. ✓ объект любой природы, данные о котором хранятся в БД
2. двумерная таблица, содержащую некоторые данные
3. количество строк в таблице
4. строка таблицы

Анализ медицинских данных Вопрос_293

SQL запрос для удаления таблицы:

1. ✓ DROP TABLE
2. INSERT TABLE
3. DELETE TABLE
4. UPDATE TABLE

Анализ медицинских данных Вопрос_294

Как вернуть кол-во записей в таблице patient?

1. ✓ SELECT COUNT(*) FROM patient
2. SELECT COLUMNS(*) FROM patient
3. SELECT COUNT() FROM patient
4. SELECT COLUMNS() FROM patient

Анализ медицинских данных Вопрос_295

Что вернёт запрос «select 1 from Users» в таблице Users, если она существует?

1. ✓ Запрос выведет 1 столько раз, сколько строк в таблице Users
2. Запрос выведет количество строк в таблице Users
3. Запрос выведет 1
4. Запрос выведет первую строку таблицы Users

Анализ медицинских данных Вопрос_296

Какое ключевое слово SQL используется для группировки результирующей выборки?

1. ✓ group by
2. ORDER BY
3. sort by
4. pivot

Анализ медицинских данных Вопрос_298

При каком соединении могут остаться «висящие» кортежи?

1. ✓ при полном внешнем соединении (full outer join)
2. при левом внешнем соединении (left outer join)
3. при перекрёстном соединении (cross join)
4. при тета-соединении (theta-join)

Анализ медицинских данных Вопрос_299

Какой тип оператора будет анализироваться в первую очередь (имеет больший приоритет) в выражении SQL из представленных:

1. ✓ унарные '+' и '-'
2. is null
3. умножение (*), деление (/)
4. and

Анализ медицинских данных Вопрос_3

Измеряет долю истинных положительных результатов (TP) среди всех прогнозируемых положительных результатов

1. ✓ Precision
2. Accuracy
3. Recall
4. F1-мера

Анализ медицинских данных Вопрос_30

Какой метод машинного обучения используется для обработки изображений?

1. ✓ Нейронные сети
2. Решающее дерево
3. Линейная регрессия
4. Алгоритм k-ближайших соседей

Анализ медицинских данных Вопрос_300

Каким запросом можно получить все поля и записи таблицы patient?

1. ✓ SELECT * FROM patient
2. SELECT * .patient
3. SELECT [all] FROM patient
4. SELECT patient

Анализ медицинских данных Вопрос_31

В методологию CRISP-DM НЕ входит этап:

1. ✓ обучения модели
2. понимания данных
3. оценки
4. моделирования

Анализ медицинских данных Вопрос_32

Какой набор данных недоступен при обучении?

1. ✓ Тестовый
2. Обучающий
3. Контрольный
4. Демонстрационный

Анализ медицинских данных Вопрос_33

Для получения объективной оценки производительности модели используется набор данных

1. ✓ Тестовый
2. Обучающий
3. Контрольный
4. Демонстрационный

Анализ медицинских данных Вопрос_34

Для оптимизации параметров модели используется набор данных

1. ✓ Проверочный
2. Обучающий
3. Тестовый
4. Контрольный

Анализ медицинских данных Вопрос_35

Ошибки, используемые для понимания того, какая модель дает наименьшую частоту ошибок

1. ✓ MAE, RMSE
2. MSE, MAPE
3. R Squared, Loss
4. WAPE

Анализ медицинских данных Вопрос_36

При одномерной статистике для визуализации распределения категориальных признаков используются

1. ✓ Гистограммы
2. Диаграммы рассеяния
3. Корреляционные матрицы
4. Линейные графики

Анализ медицинских данных Вопрос_37

Для визуализации взаимосвязи между признаками используются

1. ✓ Точечные диаграммы
2. Линейные графики
3. Гистограммы
4. Круговые диаграммы

Анализ медицинских данных Вопрос_38

В корреляционных матрицах идеальной положительной корреляции между признаками соответствует значение

1. ✓ 1
2. 0
3. -1
4. NaN

Анализ медицинских данных Вопрос_39

Модель машинного обучения имеет низкие значения метрик классификации, если

1. ✓ используется несбалансированный набор данных
2. количество образцов в классе распределено равномерно
3. используется неидентичный синтетический датасет
4. используется набор данных с большим числом признаков

Анализ медицинских данных Вопрос_4

Измеряет долю истинных положительных результатов (TP) среди всех истинных положительных результатов

1. ✓ Recall
2. Precision
3. Accuracy
4. F1-мера

Анализ медицинских данных Вопрос_40

Замена данных выбранного признака на среднее значение, либо наиболее частое выполняется

1. ✓ в столбцах с пропущенными значениями
2. в повторяющихся строках
3. в повторяющихся столбцах
4. в столбцах с категориальными признаками

Анализ медицинских данных Вопрос_41

Ансамблевый метод, при котором несколько экземпляров базовой модели обучаются параллельно (независимо друг от друга) на разных бутстрэп выборках, а затем агрегируются выбранным процессом «усреднения»

1. ✓ Бэггинг
2. Стекинг
3. Бустинг
4. Блендинг

Анализ медицинских данных Вопрос_42

Ансамблевый метод, при котором комбинирование слабых функций, которые строятся в ходе итеративного процесса, где на каждом шаге новая модель обучается с использованием данных об ошибках предыдущих

1. ✓ Бустинг
2. Бэггинг
3. Стекинг
4. Бутстрэп

Анализ медицинских данных Вопрос_43

Рекомендательные системы, обнаружение аномалий, группировка справочных данных являются примерами применений методов обучения

1. ✓ Без учителя
2. С частичным привлечением учителя
3. С подкреплением
4. С учителем

Анализ медицинских данных Вопрос_44

Выявление скрытых связей или правил между объектами в наборе данных является целью

1. ✓ Алгоритмов ассоциативного анализа
2. Алгоритмов регрессионного анализа
3. Алгоритмов анализа временных рядов
4. Алгоритмов кластеризации

Анализ медицинских данных Вопрос_45

Для предотвращения переобучения модели может быть применен

1. ✓ L1 и L2 регуляризация
2. метод главных компонент
3. перекрестная проверка
4. сеточный поиск параметров

Анализ медицинских данных Вопрос_46

Для чего необходима настройка гиперпараметров модели?

1. ✓ для достижения лучшей производительности модели и улучшений способности к обобщению
2. для того, чтобы модель слишком хорошо запомнила обучающие данные
3. для обеспечения сбалансированности классов модели
4. для оценки, насколько повышение сложности модели значительно увеличивает её точность

Анализ медицинских данных Вопрос_47

Процесс настройки гиперпараметров НЕ включает шаг

1. ✓ определения архитектуры модели
2. определения пространства поиска значений
3. выбора оптимальных значений
4. проведения экспериментов с комбинациями значений

Анализ медицинских данных Вопрос_48

В алгоритме случайного леса типичным гиперпараметром является

1. ✓ Максимальное количество признаков для разбиения (max_features)
2. Вес каждого класса (class_weight)
3. Порог для останова наращивания дерева (min_impurity_split)
4. Значение для генерации случайных чисел (random_state)

Анализ медицинских данных Вопрос_49

Поддержка, достоверность и улучшение являются характеристиками

1. ✓ ассоциативных правил
2. алгоритмов регрессии
3. алгоритмов кластеризации
4. алгоритмов диагностики

Анализ медицинских данных Вопрос_5

Среднее гармоническое между точностью и полнотой

1. ✓ F1-мера
2. Precision
3. Recall
4. Accuracy

Анализ медицинских данных Вопрос_50

Apriori - алгоритм, использующийся

1. ✓ для поиска ассоциативных правил
2. для анализа временных рядов
3. для анализа связей между несколькими переменными
4. для определения наиболее значимых переменных в наборе данных

Анализ медицинских данных Вопрос_51

Выберите метрики кластеризации

1. ✓ расстояние Манхэттена, Канберра, Евклидово расстояние
2. средняя абсолютная масштабированная ошибка, средняя относительная ошибка
3. коэффициент детерминации R-квадрат, скорректированный коэффициент детерминации
4. поддержка, достоверность и улучшение

Анализ медицинских данных Вопрос_52

Какие ML-алгоритмы не чувствительны к данным, находящимся в разных масштабах?

1. ✓ Decision Tree, Random Forest
2. KNN, K Means
3. XGBoost, CatBoost
4. MLPRegressor, Sequential

Анализ медицинских данных Вопрос_53

Визуальные методы для обнаружения выбросов в данных

1. ✓ Box plot, Scatter plot
2. Bubble plot, bar
3. Hist, lollipop chart
4. Dot plot, slopechart

Анализ медицинских данных Вопрос_54

OneHotEncoder, OrdinalEncoder, LabelEncoder, BinaryEncoder используются для

1. ✓ преобразования категориальных или текстовых данных в числовые
2. изменение сценария обучения модели
3. кодирования числовых признаков в категориальные
4. преобразования размерности признаков

Анализ медицинских данных Вопрос_55

StandardScaler, MinMaxScaler, MinAbsScaler, RobustScaler используются для

1. ✓ способы масштабирования признаков
2. кодирования бинарных признаков
3. сокращение размерности признаков
4. уменьшение количества признаков

Анализ медицинских данных Вопрос_56

Как называется комплекс языковых и программных средств, предназначенный для создания, ведения и совместного использования БД многими пользователями?

1. ✓ Система управления базами данных
2. База данных
3. Хранилище данных
4. Аналитическая обработка в реальном времени

Анализ медицинских данных Вопрос_57

Что выполняет данная команда DROP TABLE ?

1. ✓ Удаление таблицы
2. Изменение структуры таблицы
3. Выборка записей
4. Создание представления

Анализ медицинских данных Вопрос_58

Какая команда используется для проверки типов данных при исследовательском анализе данных?

1. ✓ df.dtypes
2. df.drop
3. df.head
4. df.rename

Анализ медицинских данных Вопрос_59

Выберите верное определение понятию "Выброс" в EDA?

1. ✓ точка или набор точек, которые отличаются от других точек
2. частота появления переменных в определенном интервале
3. тип визуализации, применяемый, когда нам нужно найти зависимые переменные
4. диаграмма разброса для нахождения корреляции (взаимосвязи) между двумя переменными

Анализ медицинских данных Вопрос_6

Для определения стабильных прогнозов в условиях, когда выборка сбалансирована, лучшей метрикой является:

1. ✓ Точность
2. Полнота
3. Средняя абсолютная ошибка
4. Средняя квадратическая ошибка

Анализ медицинских данных Вопрос_60

Выберите библиотеку Python с открытым исходным кодом, которая используется для создания интерактивных информационных панелей, где можно анализировать данные, визуализировать различные функции и соответствующим образом манипулировать

данными.

1. ✓ MitoSheet
2. NumPy
3. Pandas
4. TensorFlow

Анализ медицинских данных Вопрос_61

Выберите верную ML-задачу по ее описанию: помощь пользователям в поиске наиболее релевантной вещи.

1. ✓ Ранжирование
2. Рекомендации
3. Классификация
4. Регрессия

Анализ медицинских данных Вопрос_62

Что такое машинное обучение?

1. ✓ основная ветвь искусственного интеллекта, цель которой - дать компьютерам возможность учиться без явного программирования
2. раздел информатики, изучающий проблемы анализа, обработки и представления данных в цифровой форме
3. метод быстрого прототипирования, который позволяет с помощью множества визуальных представлений одних и тех же данных обнаружить скрытые взаимосвязи и зависимости
4. математическая модель, а также ее программное или аппаратное воплощение, построенная по принципу организации и функционирования биологических нейронных сетей

Анализ медицинских данных Вопрос_63

Как называется простейшая схема стекинга?

1. ✓ блендинг
2. фолд
3. бутстрэп
4. бэггинг

Анализ медицинских данных Вопрос_64

Какой вид бустинга не относится к данному методу?

1. ✓ категориальный
2. адаптивный
3. градиентный
4. экстремальный

Анализ медицинских данных Вопрос_66

Выберите верное утверждение:

1. ✓ Визуальный анализ - включает построение графиков для визуального изучения трендов, сезонности и других характеристик.
2. Статистический анализ - использование различных техник для удаления шума и выявления общих трендов в данных. Например, скользящее среднее или экспоненциальное сглаживание.
3. Методы сглаживания - использование статистических методов для описания и моделирования временных рядов. Например, расчет среднего значения, стандартного отклонения, анализ автокорреляции и других статистических показателей.
4. Методы прогнозирования - разложение на составляющие части, такие как тренд, сезонность и остаток. Это позволяет лучше понять структуру данных и выделить важные компоненты.

Анализ медицинских данных Вопрос_67

Что такой тренд в характеристике временных рядов?

1. ✓ долгосрочное изменение в данных
2. циклические изменения данных, которые повторяются с постоянным интервалом времени
3. случайные колебания данных, которые не подчиняются определенным закономерностям
4. долгосрочные колебания данных, которые не связаны с сезонностью.

Анализ медицинских данных Вопрос_68

Как правильно расшифровывается ARIMA?

1. ✓ Autoregressive Integrated Moving Average
2. Autoregressive Integrated Moving Average
3. Autoregressive Moving Average
4. Autoregressive Integrated Average

Анализ медицинских данных Вопрос_69

Что не относится к метрикам кластеризации?

1. ✓ Нет верного ответа
2. Евклидово расстояние
3. Расстояние Манхэттена
4. Расстояние Канберра

Анализ медицинских данных Вопрос_7

Задача кластеризации решается для

1. ✓ группировки похожих объектов
2. поиске наиболее релевантного объекта
3. обнаружения аномалий
4. прогнозирования численного значения

Анализ медицинских данных Вопрос_70

Что показывает Support ассоциативного правила?

1. ✓ показывает то, как часто в базе данных вместе встречаются X и Y
2. показывает, как часто в записях базы данных, содержащих X, одновременно присутствует Y
3. показывает то, как часто X и Y появляются вместе, одновременно учитывая, с какой частотой появляется каждый из них
4. Нет верного ответа

Анализ медицинских данных Вопрос_71

Для чего используется параметр min_support в Apriori?

1. ✓ для выбора элементов со значениями поддержки, превышающими значение, указанное в параметре
2. фильтрует те правила, достоверность которых превышает порог достоверности, указанный в параметре
3. определяет минимальное значение подъема для правил, включенных в список
4. указывает минимальное количество элементов, которые вы хотите использовать в своих правилах

Анализ медицинских данных Вопрос_72

Выберите верное определение кластеризации в анализе данных

1. ✓ выделение групп или классов в заданном множестве наблюдаемых объектов, при этом всякая группа должна содержать объекты, схожие между собой в большей степени, чем с объектами других выделяемых групп
2. выделение групп или классов в заданном множестве наблюдаемых объектов, при

этом всякая группа должна содержать объекты, различные между собой в большей степени, чем с объектами других выделяемых групп

3. выделение групп или классов в заданном множестве наблюдаемых объектов, при этом всякая группа должна содержать объекты, равные между собой в большей степени, чем с объектами других выделяемых групп

4. Нет верного ответа

Анализ медицинских данных Вопрос_73

Как можно заменить значения NaN на нуль в DataFrame?

1. ✓ `df.fillna(0)`

2. `df.replace_na(0)`

3. `df.impute_na(0)`

4. `df.null_to_zero()`

Анализ медицинских данных Вопрос_74

Какая функция используется для группировки данных в DataFrame по одному или нескольким столбцам?

1. ✓ `df.group_by()`

2. `df.aggregate()`

3. `df.pivot_table()`

4. `df.groupby()`

Анализ медицинских данных Вопрос_75

Какая функция используется для вычисления описательной статистики DataFrame?

1. ✓ `df.stats()`

2. `df.analysis()`

3. `df.describe()`

4. `df.summary()`

Анализ медицинских данных Вопрос_76

Какое ключевое слово используется для выбора конкретных столбцов из таблицы?

1. ✓ SELECT

2. PICK

3. CHOOSE

4. FETCH

Анализ медицинских данных Вопрос_77

С помощью какой функции библиотеки Pandas можно построить ящик с усами?

1. ✓ `boxplot()`

2. `heatmap()`

3. `scatter()`

4. `box()`

Анализ медицинских данных Вопрос_78

Что такое Kaggle?

1. ✓ Платформа для обмена данными и соревнований по решению задач с применением машинного обучения

2. Поисковая система для научных статей

3. Инструмент для визуализации данных

4. Библиотека для Python для работы с данными

Анализ медицинских данных Вопрос_79

Какие основные преимущества предоставляет Google Colab?

1. ✓ Бесплатное использование с доступом к GPU и TPU

2. Возможность совместной работы в реальном времени

3. Интеграция с Google Drive для хранения файлов

4. Интеграция с Google Drive для хранения файлов

Анализ медицинских данных Вопрос_8

Определение диагноза пациента является задачей

1. ✓ классификации
2. регрессии
3. кластеризации
4. прогнозирования

Анализ медицинских данных Вопрос_80

Какой из графиков НЕ является визуализацией данных?

1. ✓ ROC-кривая (Roc_curve)
2. Ящик с усами (Boxplot)
3. Гистограмма (Hisp)
4. Тепловая карта (Heatmap)

Анализ медицинских данных Вопрос_81

Что из перечисленного НЕ является этапом CRISP-DM?

1. ✓ Сбор данных (Collect Initial Data)
2. Подготовка данных (Data preparation)
3. Внедрение (Deployment)
4. Бизнес-анализ (Business understanding)

Анализ медицинских данных Вопрос_82

За что отвечает команда `model.fit(x, y)` из библиотеки `RandomForestClassifier`?

1. ✓ Обучение модели `model` на тренировочных данных `features` и `targets`
2. Показ точности модели `model` на тренировочных данных `features` и `targets`
3. Обучение модели `model` на тестовых данных `features` и `targets`
4. Показ точности модели `model` на тестовых данных `features` и `targets`

Анализ медицинских данных Вопрос_83

В чём основное различие команд `loc` и `iloc`?

1. ✓ `loc` возвращает строки/столбцы по индексам в индексных строках/столбцах, `iloc` - по индексам позиций
2. `loc` возвращает строки/столбцы по индексам позиций, `iloc` - по индексам в индексных строках/столбцах
3. Оба возвращают строки/столбцы по индексам в индексных строках/столбцах, но `loc` - весь диапазон, а `iloc` - кроме последнего индекса
4. Оба возвращают строки/столбцы по индексам позиций, но `loc` - весь диапазон, а `iloc` - кроме последнего индекса

Анализ медицинских данных Вопрос_84

Что выдаст следующая команда: `pd.read_csv('Dataframe.csv', decimal = '.', usecols = [2])`?

1. ✓ Третий столбец датафрейма, где точка - обозначение десятичной части числа
2. Второй столбец датафрейма, где точка - обозначение десятичной части числа
3. Датафрейм, где столбец с индексом 2 является индексируемым (главным) и где точка - обозначение десятичной части числа
4. Датафрейм, где второй столбец является индексируемым (главным) и где точка - обозначение десятичной части числа

Анализ медицинских данных Вопрос_85

Что выдаст следующая команда: `s.groupby('Work').agg(age=('Age','mean'))`?

1. ✓ Таблицу `S` со значениями столбцов `Work` и `Age`, где первый столбец отсортирован по алфавиту и где второй столбец - средний возраст рабочих по каждой профессии
2. Таблицу `S` со значениями столбцов `Work` и `Age`, где первый столбец отсортирован по алфавиту и где второй столбец - общий средний возраст рабочих

3. Таблицу S со значениями столбцов Work и Age, где второй столбец - средний возраст рабочих по каждой профессии
4. Таблицу S со значениями столбцов Work и Age, где второй столбец - общий средний возраст рабочих

Анализ медицинских данных Вопрос_86

Что произойдёт с датафреймом, в котором расписаны авиарейсы: `df = pd.get_dummies(df, columns=['Takeoff', 'Boarding'])`?

1. ✓ Аэропорты взлёта и посадки станут категориальным типом данных, значения у которых: True - самолёт был в аэропорте, False - не был
2. Аэропорты взлёта и посадки станут категориальным типом данных, значения у которых: 1 - самолёт был в аэропорте, 0 - не был
3. В датафрейме будут только аэропорты взлёта и посадки, но они станут категориальным типом данных, значения у которых: True - самолёт был в аэропорте, False - не был
4. В датафрейме будут только аэропорты взлёта и посадки, но они станут категориальным типом данных, значения у которых: 1 - самолёт был в аэропорте, 0 - не был

Анализ медицинских данных Вопрос_87

Какую метрику оценивает следующая формула: $TP/(TP+FP)$?

1. ✓ Precision
2. Accuracy
3. Recall
4. Veracity

Анализ медицинских данных Вопрос_89

Какой тип выравнивания данных включает в себя создание новых датасетов, похожих с имеющимися?

1. ✓ Data generation
2. Down-sampling
3. Up-sampling
4. Sample weights

Анализ медицинских данных Вопрос_9

Какую задачу из нижеперечисленных можно отнести к задаче классификации?

1. ✓ определение наиболее целесообразного способа лечения
2. определение длительности и исхода заболевания
3. оценивание тяжести заболевания
4. поиск ассоциативных признаков заболевания

Анализ медицинских данных Вопрос_90

Что делает следующая формула (используемая при анализе данных): $IQR = Q3 - Q1$?

1. ✓ Определяет разброс средней половины данных вычитанием третьего и первого квартилей для выявления "выбросов", выходящих за значения $Q1 - 1.5 * IQR$ или $Q3 + 1.5 * IQR$
2. Определяет второй квартиль вычитанием третьего и первого квартилей для определения центра всех значений
3. Определяет разброс средней половины данных вычитанием третьего и первого квартилей для выявления "выбросов", выходящих за значения $Q1$ или $Q3$
4. Определяет разброс средней половины данных вычитанием третьего и первого квартилей для выявления "выбросов", выходящих за значения $Q1 - Q2$ или $Q3 + Q2$

Анализ медицинских данных Вопрос_91

Какое обозначение у данной регрессионной метрики: $1/n * \sum |знач. - прогноз. знач.|$?

1. ✓ MAE
2. MSE
3. RMS
4. R²

Анализ медицинских данных Вопрос_92

Какая формула обозначает расстояние Канберра (перемещение по координатной сетке)?

1. ✓ $\sum |x - y| / (|x| + |y|)$
2. $\sum |x - y|$
3. $\sqrt{\sum (x - y)^2}$
4. $\sum |x - y| / (x - y)$

Анализ медицинских данных Вопрос_94

Для чего в SQL нужна виртуальная таблица VIEW?

1. ✓ Сохранение результата выборки данных с помощью команды SELECT
2. Сохранение результата отбора данных с помощью команды CASE
3. Сохранение всех данных из таблицы, указанной после названия VIEW
4. Сохранение указанных в коде столбцов из определённой таблицы

Анализ медицинских данных Вопрос_95

Что такое дата фрейм (DataFrame)?

1. ✓ Двумерная структура данных, представляющая собой таблицу с метками для строк и столбцов и предназначенная для хранения данных с возможностью легкого доступа и манипуляции с ними
2. Основная единица измерения данных в компьютерной графике, используемая для создания трехмерных моделей
3. Метод шифрования данных, применяемый для защиты информации в компьютерных сетях
4. Специализированный процессор, предназначенный для выполнения операций с данными в реальном времени

Анализ медицинских данных Вопрос_96

Какая библиотека позволяет работать с дата фреймами (data frames) в python?

1. ✓ pandas
2. dogas
3. catas
4. tigras

Анализ медицинских данных Вопрос_97

Altair, Lux, SweetViz, Seaborn - это

1. ✓ Библиотеки визуализации данных
2. Библиотеки анализа данных
3. Библиотеки для решения задач построения и тренировки нейронной сети
4. Библиотеки для математических вычислений

Анализ медицинских данных Вопрос_98

Самая популярная библиотека машинного обучения для Python

1. ✓ scikit-learn
2. pandas
3. tensorflow
4. numpy

Анализ медицинских данных Вопрос_99

Библиотека Python для обработки структурированных данных и их анализа

1. ✓ pandas

- 2. scikit-learn
- 3. tensorflow
- 4. numpy

Принята на заседании кафедры Информационных технологий и обработки медицинских данных ЦЦМиИИМ ИЦБиИИМ

от «23» ноября 2024 г., протокол № 5

Заведующий кафедрой

Информационных технологий
и обработки медицинских
данных ЦЦМиИИМ
ИЦБиИИМ



(подпись)

Лебедев Г.С.

(фамилия, инициалы)