

Документ подписан простой электронной подписью  
 Информация о владельце:  
 ФИО: Косенок Сергей Михайлович  
 Должность: ректор  
 Дата подписания: 19.06.2024 07:20:13  
 Уникальный программный идентификатор:  
 e3a68f3eaa1e62674b54f4998099d3d6bfdcf836

## Тестовое задание для диагностического тестирования по дисциплине

### Инструменты, подходы и методы обработки структурированных и неструктурированных данных, 5 семестр

Код, направление подготовки	09.03.01 ИНФОРМАТИКА И ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА
Направленность (профиль)	Искусственный интеллект и экспертные системы
Форма обучения	Очная
Кафедра разработчик	Автоматизированных систем обработки информации и управления
Выпускающая кафедра	Автоматизированных систем обработки информации и управления

№	Проверяемая компетенция	Задание	Варианты ответов	Тип сложности вопроса
1	ОПК-2.1 ОПК-2.2 ОПК-9.1 ОПК-9.2 ПК-7.1 ПК- 7.2 ПК-7.3 ПК-11.1 ПК -11.2	... - это процесс преобразования текстовых данных в числовой формат, который может быть обработан компьютером.	—	Низкий
2	ОПК-2.1 ОПК-2.2 ОПК-9.1 ОПК-9.2 ПК-7.1 ПК- 7.2 ПК-11.1 ПК-11.2	Какой тип базы данных предназначен для хранения неструктурированных данных?	1. Объектно-ориентированная база данных 2. NoSQL база данных 3. Реляционная база данных 4. Иерархическая база данных	Низкий

3	ОПК-2.1 ОПК-2.2 ОПК-9.1 ОПК-9.2 ПК-7.1 ПК- 7.2 ПК-11.1 ПК-11.2	Какой метод используется для обработки текстовых данных?	1. Векторизация текстов 2. Случайный лес 3. DBSCAN 4. К-средних	Низкий
4	ОПК-2.1 ОПК-2.2 ОПК-9.1 ОПК-9.2 ПК-7.1 ПК- 7.2 ПК-7.3 ПК-11.1 ПК -11.2	Какой инструмент используется для поиска ассоциативных правил в структурированных данных?	1. К-средних 2. Метод главных компонент 3. Алгоритм Apriori 4. SVM	Низкий
5	ОПК-2.1 ОПК-2.2 ОПК-9.1 ОПК-9.2 ПК-7.1 ПК- 7.2 ПК-11.1 ПК-11.2	... - это процесс превращения неструктурированных данных в структурированный формат, который может быть обработан компьютером.	—	Низкий
6	ОПК-2.2 ОПК-9.2 ПК-7.2 ПК- 7.3 ПК-11.2	Укажите процент неструктурированных данных в Интернете?	—	Средний
7	ОПК-2.2 ОПК-9.2 ПК-7.2 ПК- 7.3 ПК-11.2	Какие инструменты можно использовать для обработки структурированных данных?	1. базы данных 2. алгоритмы 3. машинное обучение 4. математические модели	Средний
8	ОПК-2.2 ОПК-9.2 ПК-7.2 ПК- 7.3 ПК-11.2	Метод опорных векторов (SVM) используется для задачи ... .	—	Средний
9	ОПК-2.1 ОПК-2.2 ОПК-9.1 ОПК-9.2 ПК-7.1 ПК- 7.2 ПК-7.3 ПК-11.1 ПК -11.2	Метод наименьших квадратов (OLS) используется в задачах ... .	—	Средний

10	ОПК-2.1 ОПК-2.2 ОПК-9.1 ОПК-9.2 ПК-7.1 ПК- 7.2 ПК-7.3 ПК-11.1 ПК -11.2	Какой инструмент используется для обработки структурированных данных?	1. Графические редакторы 2. Текстовые редакторы 3. Операционные системы 4. Базы данных	Средний
11	ОПК-2.1 ОПК-2.2 ОПК-9.1 ОПК-9.2 ПК-7.1 ПК- 7.2 ПК-7.3 ПК-11.1 ПК -11.2	Какой подход используется для анализа больших объемов данных?	1. Бизнес-аналитика 2. Статистический анализ 3. Машинное обучение 4. Нет верного ответа	Средний
12	ОПК-2.2 ОПК-9.2 ПК-7.2 ПК- 7.3 ПК-11.2	Каково количество возможных различных бинарных строк длины 8?	—	Средний
13	ОПК-2.1 ОПК-2.2 ОПК-9.1 ОПК-9.2 ПК-7.1 ПК- 7.2 ПК-7.3 ПК-11.1 ПК -11.2	Соотнесите термин и его определение:	1. Токенизация разбиение текста на отдельные слова или токены. 2. Лемматизация процесс приведения слова к его основе путём отбрасывания окончаний. 3. Стемминг процесс приведения слова к его базовой форме	Средний

14	ОПК-2.1 ОПК-2.2 ОПК-9.1 ОПК-9.2 ПК-7.1 ПК- 7.2 ПК-7.3 ПК-11.1 ПК -11.2	Соотнесите метод и его описание:	<p>1. К-средних метод кластеризации, который разбивает данные на К кластеров, минимизируя сумму квадратов расстояний между точками и центроидами кластеров.</p> <p>2. Иерархическая кластеризация метод кластеризации, который находит плотные регионы в пространстве объектов и считает их кластерами, а объекты, находящиеся вне плотных регионов, - выбросами.</p> <p>3. DBSCAN метод кластеризации, который строит иерархическое дерево кластеров, объединяя на каждом шаге два ближайших кластера.</p>	Средний
15	ОПК-2.1 ОПК-2.2 ОПК-9.1 ОПК-9.2 ПК-7.1 ПК- 7.2 ПК-7.3 ПК-11.1 ПК -11.2	Соотнесите тип данных и пример:	<p>1. Категориальные данные, которые можно измерить и представить числами (например, возраст, доход, количество товаров).</p> <p>2. Количественные данные, которые могут принимать только два значения (например, да/нет, 0/1, мужской/женский).</p> <p>3. Бинарные данные, которые могут принимать ограниченное число значений из заданного набора (например, цвет, пол, тип автомобиля).</p>	Средний

16	ОПК-2.1 ОПК-2.2 ОПК-9.1 ОПК-9.2 ПК-7.1 ПК- 7.2 ПК-7.3 ПК-11.1 ПК -11.2	Упорядочите следующие методы предобработки данных по их порядку применения:	1. Нормализация данных 2. Кодирование категориальных признаков 3. Обработка пропущенных значений 4. Удаление выбросов	Высокий
17	ОПК-2.1 ОПК-2.2 ОПК-9.1 ОПК-9.2 ПК-7.1 ПК- 7.2 ПК-7.3 ПК-11.1 ПК -11.2	Упорядочите следующие методы обработки текстовых данных по стадиям обработки:	1. Синтаксический анализ 2. Стемминг / Лемматизация 3. Токенизация	Высокий
18	ОПК-2.1 ОПК-2.2 ОПК-9.1 ОПК-9.2 ПК-7.1 ПК- 7.2 ПК-7.3 ПК-11.1 ПК -11.2	Какие из следующих языков программирования наиболее часто используются для обработки и анализа данных?	1. Python 2. R 3. Java 4. C++	Высокий
19	ОПК-2.1 ОПК-2.2 ОПК-9.1 ОПК-9.2 ПК-7.1 ПК- 7.2 ПК-7.3 ПК-11.1 ПК -11.2	Какие из следующих методов можно использовать для обработки пропущенных значений в структурированных данных?	1. Интерполяция 2. Удаление строк с пропущенными значениями 3. Замена пропущенных значений на среднее значение столбца 4. Замена пропущенных значений на медианное значение столбца	Высокий

20	ОПК-2.1 ОПК-2.2 ОПК-9.1 ОПК-9.2 ПК-7.1 ПК- 7.2 ПК-7.3 ПК-11.1 ПК -11.2	Какие из следующих методов обработки текстовых данных используются для извлечения ключевых слов из текста?	1. LDA 2. Лемматизация 3. TF-IDF 4. Стемминг	Высокий
----	--	---	---	---------

№	ПРАВИЛЬНЫЕ ОТВЕТЫ
1	Векторизация; Токенизация; Кластеризация; Обучение без учителя
2	NoSQL база данных
3	Векторизация текстов
4	Алгоритм Apriori
5	Токенизация; Векторизация; Кластеризация; Обнаружение аномалий
6	80 %
7	базы данных; машинное обучение; алгоритмы; математические модели
8	классификации; регрессии; кластеризации; обучения с подкреплением
9	регрессии; классификации; кластеризации; обучения с подкреплением
10	Базы данных
11	Машинное обучение
12	256
13	Токенизация разбиение текста на отдельные слова или токены.; Лемматизация процесс приведения слова к его базовой форме; Стемминг процесс приведения слова к его основе путём отбрасывания окончаний.

14	<p>K-средних метод кластеризации, который разбивает данные на K кластеров, минимизируя сумму квадратов расстояний между точками и центроидами кластеров.; Иерархическая кластеризация метод кластеризации, который строит иерархическое дерево кластеров, объединяя на каждом шаге два ближайших кластера.; DBSCAN метод кластеризации, который находит плотные регионы в пространстве объектов и считает их кластерами, а объекты, находящиеся вне плотных регионов, - выбросами.</p>
15	<p>Категориальные данные данные, которые могут принимать ограниченное число значений из заданного набора (например, цвет, пол, тип автомобиля).; Количественные данные данные, которые можно измерить и представить числами (например, возраст, доход, количество товаров).; Бинарные данные данные, которые могут принимать только два значения (например, да/нет, 0/1, мужской/женский).</p>
16	<p>Обработка пропущенных значений; Удаление выбросов; Нормализация данных; Кодирование категориальных признаков</p>
17	<p>Токенизация; Стемминг / Лемматизация; Синтаксический анализ</p>
18	<p>Python; R</p>
19	<p>Замена пропущенных значений на среднее значение столбца; Замена пропущенных значений на медианное значение столбца; Интерполяция</p>
20	<p>TF-IDF; LDA</p>