

Документ подписан простой электронной подписью
Информация о владельце:

ФИО: Косенок Сергей Михайлович

Должность: ректор

Дата подписания: 11.06.2026 09:32:23

Уникальный программный ключ:

e3a68f3eaa1e62674b544499809203d0b1dc183e

Тестовое задание для диагностического тестирования по дисциплине

Инструменты, подходы и методы обработки структурированных и неструктурированных данных, 7 семестр

Код, направление подготовки	09.03.01 ИФОРМАТИКА 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	ПК-6.1 ПК-6.2 ПК-6.3	... - это процесс преобразования текстовых данных в числовой формат, который может быть обработан компьютером.	—	Низкий
2	ПК-6.1 ПК-6.2 ПК-6.3	Какой тип базы данных предназначен для хранения неструктурированных данных?	<ol style="list-style-type: none"> 1. NoSQL база данных 2. Иерархическая база данных 3. Объектно-ориентированная база данных 4. Реляционная база данных 	Низкий
3	ПК-6.1 ПК-6.2 ПК-6.3	Какой метод используется для обработки текстовых данных?	<ol style="list-style-type: none"> 1. DBSCAN 2. Случайный лес 3. К-средних 4. Токенизация 	Низкий
4	ПК-6.1 ПК-6.2 ПК-6.3	В чем принципиальная необходимость превращения неструктурированных данных в структурированный формат?	<ol style="list-style-type: none"> 1. возможность компьютерной обработки 2. возможность обработки человеком 3. наглядность 4. уменьшение объема 	Низкий

5	ПК-6.1 ПК-6.2 ПК-6.3	Какой процент неструктурированных данных в Интернете?	0% 2. 20% 3. 80% 4. 100%	Низкий
6	ПК-6.1 ПК-6.2 ПК-6.3	Как в Python хранится вещественное число?	1. float, занимает 32 бита (4 байта) 2. float, но, по сути, double, занимает 64 бита (8 байта) 3. массив необходимой (переменной) длины из 32 битных целых чисел 4. два массива (мантисса и экспонента) необходимой длины из 32 битных целых чисел	Средний
7	ПК-6.1 ПК-6.2 ПК-6.3	Что используется для решения задачи линейной регрессии?	1. Метод бисекции 2. Метод опорных векторов 3. Метод наименьших квадратов 4. Логистическая функция потерь	Средний
8	ПК-6.1 ПК-6.2 ПК-6.3	Какой инструмент используется для обработки структурированных данных?	1. Графические редакторы 2. Текстовые редакторы 3. Операционные системы 4. Базы данных	Средний
9	ПК-6.1 ПК-6.2 ПК-6.3	Какая информация о словах не сохраняется при использовании представления мешок слов (bag of words)?	1. количество разных слов в документе 2. порядок вхождений слов в документ 3. число вхождений слов в документ 4. число всех слов в документе	Средний

10	ПК-6.1 ПК-6.2 ПК-6.3	Как будет выглядеть векторное представление мешка слов для набора текстов: «Это были лучшие времена» и «Это было худшее время»?	1. [1 1 1 0 1], [1 1 0 1 1] 2. [1 1 0 1 0 1], [1 0 0 1 1 1] 3. [1 1 1 0], [1 1 0 1] 4. [0 1 1 0 1], [1 1 0 0 1]	Средний									
11	ПК-6.1 ПК-6.2 ПК-6.3	Какова доля правильных ответов модели, если для нее $TP = 3$, $FP = 7$, $FN = 7$, $TN = 3$?	1. 30% 2. 35% 3. 50% 4. 70%	Средний									
		<table border="1"> <thead> <tr> <th></th> <th>$y = 1$</th> <th>$y = -1$</th> </tr> </thead> <tbody> <tr> <th>$a(x) = 1$</th> <td>True Positive (TP)</td> <td>False Positive (FP)</td> </tr> <tr> <th>$a(x) = -1$</th> <td>False negative (FN)</td> <td>True Negative (TN)</td> </tr> </tbody> </table> <p>Таблица 1. Матрица ошибок</p>		$y = 1$	$y = -1$	$a(x) = 1$	True Positive (TP)	False Positive (FP)	$a(x) = -1$	False negative (FN)	True Negative (TN)		
	$y = 1$	$y = -1$											
$a(x) = 1$	True Positive (TP)	False Positive (FP)											
$a(x) = -1$	False negative (FN)	True Negative (TN)											
12	ПК-6.1 ПК-6.2 ПК-6.3	Каково количество возможных различных бинарных строк длины 8?	—	Средний									
13	ПК-6.1 ПК-6.2 ПК-6.3	Соотнесите термин и его определение:	1. Токенизация \leftrightarrow разбиение текста на отдельные слова или токены. 2. Лемматизация \leftrightarrow процесс приведения слова к его базовой форме 3. Стемминг \leftrightarrow процесс приведения слова к его основе путём отбрасывания окончаний.	Средний									

14	ПК-6.1 ПК-6.2 ПК-6.3	Соотнесите метод и его описание:	<p>1. К-средних ↔ находит плотные регионы в пространстве объектов и считает их кластерами, а объекты, находящиеся вне плотных регионов, - выбросами.</p> <p>2. Иерархическая кластеризация ↔ разбивает данные на К кластеров, минимизируя сумму квадратов расстояний между точками и центроидами кластеров.</p> <p>3. DBSCAN ↔ строит иерархическое дерево кластеров, объединяя на каждом шаге два ближайших кластера.</p>	Средний
15	ПК-6.1 ПК-6.2 ПК-6.3	Соотнесите тип данных и пример:	<p>1. Категориальные данные ↔ могут принимать ограниченное число значений из заданного набора (например, цвет, пол, тип автомобиля).</p> <p>2. Количественные данные ↔ могут принимать только два значения (например, да/нет, 0/1, мужской/женский).</p> <p>3. Бинарные данные ↔ можно измерить и представить числами (например, возраст, доход, количество товаров).</p>	Средний

16	ПК-6.1 ПК-6.2 ПК-6.3	Выберите правильный порядок применения методов предобработки данных.	<p>1. Обработка пропущенных значений; Удаление выбросов; Нормализация данных; Кодирование категориальных признаков</p> <p>2. Удаление выбросов; Нормализация данных; Кодирование категориальных признаков; Обработка пропущенных значений;</p> <p>3. Нормализация данных; Обработка пропущенных значений; Удаление выбросов; Кодирование категориальных признаков</p> <p>4. Кодирование категориальных признаков; Обработка пропущенных значений; Удаление выбросов; Нормализация данных</p>	Высокий
17	ПК-6.1 ПК-6.2 ПК-6.3	Выберите правильный порядок применения методов обработки текстовых данных.	<p>1. Стемминг / Лемматизация; Токенизация; Синтаксический анализ;</p> <p>2. Токенизация; Стемминг / Лемматизация; Синтаксический анализ</p> <p>3. Токенизация; Синтаксический анализ; Стемминг / Лемматизация;</p> <p>4. Синтаксический анализ; Токенизация; Стемминг / Лемматизация</p>	Высокий

18	ПК-6.1 ПК-6.2 ПК-6.3	Какие из следующих языков программирования наиболее часто используются для обработки и анализа данных?	<ol style="list-style-type: none"> 1. Python 2. R 3. C++ 4. Java 	Высокий
19	ПК-6.1 ПК-6.2 ПК-6.3	Какие из следующих методов можно использовать для обработки пропущенных значений в структурированных данных?	<ol style="list-style-type: none"> 1. Замена пропущенных значений на медианное значение столбца 2. Замена пропущенных значений на среднее значение столбца 3. Интерполяция 4. Удаление строк с пропущенными значениями 	Высокий
20	ПК-6.1 ПК-6.2 ПК-6.3	Какие из следующих методов обработки текстовых данных используются для извлечения ключевых слов из текста?	<ol style="list-style-type: none"> 1. TF-IDF 2. Стемминг 3. Лемматизация 4. LDA 	Высокий