

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Косенок Сергей Михайлович

Должность: ректор

Дата подписания: 19.06.2024 06:18:08

Уникальный программный ключ:

e3a68f3eaa1e62674b541499809903d6bfdcf836

Тестовое задание для диагностического тестирования по дисциплине:

Основы научных исследований в области анализа данных, 1 семестр

Код направления подготовки	09.04.02 ИНФОРМАЦИОННЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ
Направленность (профиль)	УПРАВЛЕНИЕ ДАННЫМИ
Форма обучения	Очная
Кафедра-разработчик	Информатики и вычислительной техники
Выпускающая кафедра	Информатики и вычислительной техники

Проверяемая компетенция	№	Задание	Варианты ответов	Тип сложности вопроса
УК-1 УК-6 ОПК-4	1	Задача классификации сводится к ...	(1) нахождению частых зависимостей между объектами или событиями; (2) определению класса объекта по его характеристиками; (3) определению, по известным характеристикам объекта, значения некоторого его параметра; (4) поиска независимых групп и их характеристик во всем множестве анализируемых данных.	низкий
УК-1 УК-6 ОПК-4	2	Задача регрессии сводится к ...	(1) нахождения частых зависимостей между объектами или событиями; (2) определению класса объекта по его характеристиками; (3) определению, по известным характеристикам объекта, значения некоторого его параметра; (4) поиска независимых групп и их характеристик во всем	низкий

			множестве анализируемых данных.	
УК-1 УК-6 ОПК-4	3	Целью поиска ассоциативных правил является ...	(1) нахождения частых зависимостей между объектами или событиями; (2) определения класса объекта по его характеристикам; (3) определение, по известным характеристикам объекта, значений некоторого его параметра; (4) поиск независимых групп и их характеристик во всем множестве анализируемых данных.	низкий
УК-1 УК-6 ОПК-4	4	Очистка данных — ...	(1) комплекс методов и процедур, направленных на устранение причин, мешающих корректной обработке: аномалий, пропусков, дубликатов, противоречий, шумов и т.д. (2) процесс дополнения данных некоторой информацией, позволяющей повысить эффективность аналитических задач (3) объект, содержащий структурированные данные, которые могут оказаться полезными (4) комплекс методов и процедур, направленных на извлечение данных из различных источников, обеспечение необходимого уровня их информативности и качества, преобразования в единый формат, в котором они могут быть загружены в хранилище	низкий

			данных или аналитическую систему	
УК-1 УК-6 ОПК-4	5	При использовании какого метода необходимо задавать количество кластеров?	(1) метод k-средних (2) метод ближнего соседа (3) вся группа иерархических методов (4) все ответы неверны	низкий
УК-1 УК-6 ОПК-4	6	К классу описательных задач Data Mining относятся такие задачи:	(1) прогнозирование (2) классификация (3) кластеризация (4) визуализация	средний
УК-1 УК-6 ОПК-4	7	К классу прогнозирующих задач Data Mining относятся такие задачи:	(1) прогнозирование (2) классификация (3) кластеризация (4) визуализация	средний
УК-1 УК-6 ОПК-4	8	Основные меры расстояния между объектами при использовании иерархического метода кластерного анализа	(1) евклидово расстояние (2) квадрат евклидова расстояния (3) манхэттенское расстояние (4) расстояние Чебышева	средний
УК-1 УК-6 ОПК-4	9	Назовите достоинства алгоритма кластеризации k-средних	(1) простота использования (2) быстрота использования (3) понятность и прозрачность алгоритма (4) нечувствительность к выбросам	средний
УК-1 УК-6 ОПК-4	10	Алгоритм конструирования дерева решений ...	(1) не требует от пользователя выбора из набора входных атрибутов (независимых переменных), наиболее значимых (2) требует от пользователя выбора из набора входных атрибутов (независимых переменных), наиболее значимых (3) на вход алгоритма можно подавать все существующие атрибуты, алгоритм сам выберет наиболее значимые среди них, и только они будут	средний

			использованы для построения дерева	
УК-1 УК-6 ОПК-4	11	Задачи классификации решаются следующими алгоритмами:	(1) методом дерева решений (2) методы линейной регрессии (3) алгоритмом Apriori	средний
УК-1 УК-6 ОПК-4	12	Какая из технологий анализирует массивы неструктурированной информации и одним из ее методов является поиск подстроки в строке?	(1) Data Mining (2) Text Mining (3) Web Mining	средний
УК-1 УК-6 ОПК-4	13	Что такое токенизация в обработке текстов?	(1) Один из этапов обработки текста (2) Разбиение последовательности символов на последовательность слов (3) Удаление знаков препинания из строки (4) Добавление зашифрованных элементов к последовательности символов (5) Расшифровка регулярных выражений	средний
УК-1 УК-6 ОПК-4	14	Клиент покупает билет на самолет через интернет. Как бы в данном случае формулировалась задача прогнозирования?	(1) в момент покупки предсказать, насколько и когда может упасть стоимость этого билета в ближайшем будущем (2) в момент покупки определить, к какому сегменту относится пользователь и предложить выгодные для него условия сделки (3) выдать пользователю прогноз погоды для того места, куда он летит (4) спрогнозировать вероятность дополнительных покупок (отель, страховка, туристический тур) и предложить наиболее вероятные пользователю	средний

<p>УК-1 УК-6 ОПК-4</p>	<p>15</p>	<p>С некоторой периодичностью персонал предприятия списывает группы расходных материалов на различных участках учета. Для выявления ошибок, акты списания выборочно проверяются аудитором. Как бы в данном случае формулировалась задача классификации?</p>	<p>(1) определить характерные признаки ошибочных списаний (2) научиться автоматически выявлять ошибочные списания с ожидаемой ошибкой не ниже 97% (3) классифицировать типичные ошибки и составить их список (4) определить три категории: "ошибочные", "под сомнением", "безошибочные" и найти правило отнесения к этим категориям</p>	<p>средний</p>
<p>УК-1 УК-6 ОПК-4</p>	<p>16</p>	<p>На плоскости даны 8 точек с координатами А(1;1), В(2;2), С(2;4), D(3;3), Е(4;2), F(4;4), G(5;5), Н(6;6). Изначально центроиды находятся в точках (1;3), (6;4). Применив алгоритм 2-средних с евклидовой метрикой, определите точки, принадлежащие тому же кластеру, что и точка А, при достижении стационарного состояния.</p>	<p>(1) Кроме А точек нет (2) А, В (3) А, В, С (4) А, В, С, D (5) А, В, С, D, Е (6) А, В, С, D, Е, F</p>	<p>высокий</p>
<p>УК-1 УК-6 ОПК-4</p>	<p>17</p>	<p>Дано исследование самооощущения собственного материального положения среди 200 человек. 100 из них оценили свое материальное положение как "ниже среднего" ("нс"), другие 100 - как "выше среднего" ("вс"). Также эти 200 человек ответили на вопросы о наличии у них дачи, автомобиля и детей. Дача – есть/нет. Дача есть: "нс" - 35 человек, "вс" – 75 человек. Автомобиль – есть/нет. Автомобиль есть: "нс" – 20 человек, "вс" – 70 человек. Дети - нет/1/больше 1. Нет детей: "нс" - 20 человек, "вс" – 40 человек; 1 ребенок: "нс" – 50 человек, "вс" – 50 человек.</p>	<p>(1) Дача, автомобиль, дети (2) Дача, дети, автомобиль (3) Автомобиль, дача, дети (4) Автомобиль, дети, дача (5) Дети, дача, автомобиль (6) Дети, автомобиль, дача</p>	<p>высокий</p>

		Требуется построить дерево по алгоритму CART, нечистота (impurity) вычисляется по Джини. Расположите признаки в порядке убывания по качеству разбиения в корне дерева:		
УК-1 УК-6 ОПК-4	18	В документе d слово "кластер" встречается с частотой $TF("кластер",d)=0,0125$ . Мы имеем возможность программным образом изучить миллион документов, и выяснить, что указанное слово встречается только в 100 из них. Вычислите TF-IDF слова "кластер" в документе d с точностью до двух знаков после запятой:		высокий
УК-1 УК-6 ОПК-4	19	Какие проблемы решают задачи кластеризации, отыскивая "скрытую структуру" исследуемых данных и не имея опорной целевой переменной?	(1) разметка данных "в ручную" очень дорого и трудозатратно (2) построение признаков из очень большого количества данных (3) возможность отслеживать эволюционные изменения (4) поиск выбросов и шумов в исследуемых данных (5) исследование и визуализация больших данных	высокий
УК-1 УК-6 ОПК-4	20	Дома на четной стороне улицы имеют номера 2, 4, 6, ... . Номер дома – это признак:	(1) Бинарный (2) Номинальный (3) Порядковый (4) Количественный (5) Нет правильного ответа	высокий