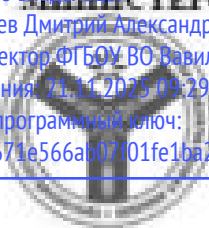


Документ подписан простой электронной подписью
Информация о документе:
ФИО: Соловьев Дмитрий Александрович
Должность: ректор ФГБОУ ВО Вавиловский университет
Дата подписания: 23.11.2025 09:29:20
Уникальный программный ключ:
528682d78e674e566ab07f01fe1ba2172f735a12



МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования

«Саратовский государственный университет генетики, биотехнологии и инженерии имени Н.И. Вавилова»

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ для проверки сформированности компетенций

Дисциплина	Анализ данных
Направление подготовки	09.03.03 Прикладная информатика
Направленность (профиль)	Проектирование информационных систем
Квалификация выпускника	Бакалавр
Нормативный срок обучения	4 года
Форма обучения	Очная, заочная

Разработчик: *доцент Шибайкин В.А.*


(подпись)

Саратов 2024

ОГЛАВЛЕНИЕ

1. Перечень компетенций с указанием этапов их формирования в процессе освоения ОПОП	3
2. Сценарии выполнения заданий	3
3. Система оценивания выполнения заданий	5
4. Описание дополнительных материалов и оборудования, необходимых для выполнения заданий	5
5. Задания для проверки уровня сформированности компетенций с указанием типа заданий (с ключами к оцениванию заданий).....	6

1. Перечень компетенций с указанием этапов их формирования в процессе освоения ОПОП

В результате изучения дисциплины «Анализ данных» обучающиеся, в соответствии с ФГОС ВО по направлению подготовки 09.03.03. Прикладная информатика, утвержденного приказом Министерства образования и науки РФ от 19.09.2017 № 922, формируют следующие компетенции, указанные в таблице:

Код компетенции	Наименование компетенции	Этапы формирования компетенции в процессе освоения ОПОП	
		семестр (очная форма обучения)	курс (заочная форма обучения)
ПК-2	Способность понимать сущность и значение информации в развитии современного информационного общества, сознавать опасности и угрозы, возникающие в этом процессе, перерабатывать большие объёмы информации, анализировать и интерпретировать геопространственные данные, проводить целенаправленный поиск в различных источниках информации по профилю деятельности	6	4
ПК-8	Способен вести базы данных и поддержку информационного обеспечения решения прикладных задач	6	4

2. Сценарии выполнения заданий

№ п/п	Тип задания	Последовательность действий при выполнении задания
1. Задания закрытого типа		
1.1	Задание закрытого типа на установление соответствия	1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидаются пары элементов. 2. Внимательно прочитать оба списка: список 1 – вопросы, утверждения, факты, понятия и т.д.; список 2 – утверждения, свойства объектов и т.д. 3. Сопоставить элементы списка 1 с элементами списка 2, сформировать пары элементов. 4. Записать попарно буквы и цифры (в зависимости от задания) вариантов ответа (например, А1 или Б4).
1.2	Задание закрытого типа на установление последовательности	1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается последовательность элементов. 2. Внимательно прочитать предложенные варианты ответа. 3. Построить верную последовательность из предложенных элементов.

№ п/п	Тип задания	Последовательность действий при выполнении задания
		4. Записать буквы/цифры (в зависимости от задания) вариантов ответа в нужной последовательности без пробелов и знаков препинания (например, БВА или 135).
2. Задания открытого типа		
2.1	Задание открытого типа с кратким ответом	1. Внимательно прочитать текст задания и понять суть вопроса. 2. Продумать краткий ответ. 3. Записать ответ в виде слова, словосочетания или числа. 4. В случае расчетной задачи, записать ответ в виде числа.
2.2	Задание открытого типа с развернутым ответом	1. Внимательно прочитать текст задания и понять суть вопроса. 2. Продумать логику и полноту ответа. 3. Записать ответ, используя четкие компактные формулировки. 4. В случае расчетной задачи, записать решение и ответ.
3. Задания комбинированного типа		
3.1	Задание комбинированного типа с выбором одного верного ответа из предложенных и обоснованием выбора	1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов. 2. Внимательно прочитать предложенные варианты ответа. 3. Выбрать один ответ, наиболее верный. 4. Записать только номер (или букву) выбранного варианта ответа. 5. Записать аргументы, обосновывающие выбор ответа.
3.2	Задание комбинированного типа с выбором нескольких верных ответов из предложенных и обоснованием выбора	1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается несколько из предложенных вариантов. 2. Внимательно прочитать предложенные варианты ответа. 3. Выбрать несколько ответов, наиболее верных. 4. Записать только номера (или буквы) выбранных вариантов ответа. 5. Записать аргументы, обосновывающие выбор ответов.

3. Система оценивания выполнения заданий

№ п/п	Указания по оцениванию	Характеристика правильности ответа
1. Задания закрытого типа		
1.1	Задание закрытого типа на установление соответствия считается верным, если правильно установлены все соответствия (позиции из одного столбца верно сопоставлены с позициями другого).	«верно» / «неверно»
1.2	Задание закрытого типа на установление последовательности считается верным, если правильно указана вся последовательность цифр.	«верно» / «неверно»
2. Задания открытого типа		
2.1	Задание открытого типа с кратким ответом оценивается по следующим критериям: 1) Правильность ответа (отсутствие фактических и грамматических ошибок). 2). Сопоставимость с эталонным ответом в случае расчетной задачи.	«верно» / «неверно»
2.2	Задание открытого типа с развернутым ответом оценивается по следующим критериям. 1) Правильность ответа (отсутствие фактических ошибок). 2) Полнота ответа (раскрытие объема используемых понятий). 3) Обоснованность ответа (наличие аргументов). 4) Логика изложения ответа (грамотная последовательность излагаемого материала). 5. Сопоставимость с эталонным ответом.	«верно» / «неверно»
3. Задания комбинированного типа		
3.1	Задание комбинированного типа с выбором одного верного ответа из предложенных с обоснованием выбора ответа считается верным, если правильно указана цифра (буква) и приведены корректные аргументы, используемые при выборе ответа	«верно» / «неверно»
3.2	Задание комбинированного типа с выбором нескольких вариантов ответа из предложенных с обоснованием выбора ответов считается верным, если правильно указаны цифры (буквы) и приведены корректные аргументы, используемые при выборе ответа.	«верно» / «неверно»

4. Описание дополнительных материалов и оборудования, необходимых для выполнения заданий

Для выполнения заданий дополнительные материалы и оборудование не требуются.

5. Задания для проверки уровня сформированности компетенций с указанием типа заданий (с ключами к оцениванию заданий)

Номер задания	Формулировка задания	Тип задания	Ключ к оцениванию задания
6 семестр (очная форма обучения) // 4 курс (заочная форма обучения)			
ПК-2 Способность понимать сущность и значение информации в развитии современного информационного общества, сознавать опасности и угрозы, возникающие в этом процессе, перерабатывать большие объёмы информации, анализировать и интерпретировать геопространственные данные, проводить целенаправленный поиск в различных источниках информации по профилю деятельности			
1.	<i>Прочитайте текст, выберите один правильный вариант ответа и запишите аргументы, обосновывающие выбор ответа:</i> Выберите правильное утверждение, которое верно описывает суть анализа данных: 1) Анализ данных — это процесс решения задач в строго определенных математических рамках, используя только классические методы статистики. 2) Анализ данных — это исключительно техническая дисциплина, не требующая понимания предметной области. 3) Анализ данных — это прикладная наука, которая требует изобретения новых техник и подходов для решения индивидуальных задач, а также умения интерпретировать результаты. 4) Анализ данных — это только обработка больших объемов данных с помощью компьютерных программ.	Задание комбинированного типа с выбором одного верного ответа из предложенных и обоснованием выбора	3 Объяснение: Анализ данных является прикладной наукой, которая требует не только технических навыков, но и понимания предметной области, а также умения интерпретировать результаты. Это область, в которой часто приходится разрабатывать новые методы и подходы для решения индивидуальных задач.

Номер задания	Формулировка задания	Тип задания	Ключ к оцениванию задания
	5) Анализ данных — это процесс, который не требует использования статистических методов и математических моделей.		
2.	<i>Прочитайте текст, выберите один правильный вариант ответа и запишите аргументы, обосновывающие выбор ответа:</i> Выберите одно понятие, которое не является основным понятием или методом анализа данных: 1) Машинное обучение 2) Интеллектуальный анализ данных (ИАД) 3) Эконометрика 4) Теория относительности 5) Статистический анализ	Задание комбинированного типа с выбором одного верного ответа из предложенных и обоснованием выбора	4 Объяснение: Теория относительности — это физическая теория, которая не имеет прямого отношения к анализу данных. Основные понятия и методы анализа данных включают машинное обучение, интеллектуальный анализ данных, эконометрику и статистический анализ.
3.	<i>Прочитайте текст, выберите один правильный вариант ответа и запишите аргументы, обосновывающие выбор ответа:</i> Какой из методов используется для проверки гипотезы о том, что средние значения двух выборок относятся к одной и той же совокупности: 1) Метод наименьших квадратов 2) Критерий Стьюдента (t-критерий) 3) Метод максимального правдоподобия 4) Метод моментов 5) Метод Байеса	Задание комбинированного типа с выбором одного верного ответа из предложенных и обоснованием выбора	2 Объяснение: Критерий Стьюдента (t-критерий) используется для проверки гипотезы о том, что средние значения двух выборок относятся к одной и той же совокупности. Этот метод позволяет оценить вероятность того, что разница между средними значениями двух выборок является статистически значимой.
4.	<i>Прочитайте текст, выберите один правильный вариант ответа и запишите аргументы, обосновывающие выбор ответа:</i> Из перечисленных методов анализа данных выберите один, который обладая высокой степенью	Задание комбинированного типа с выбором одного верного ответа из	3 Объяснение: Объяснение: Метод опорных векторов (SVM) решает задачи классификации и регрессии путем

Номер задания	Формулировка задания	Тип задания	Ключ к оцениванию задания
	гибкости при решении задач различного уровня сложности используется для классификации и регрессии: 1) Метод k-Ближайших Соседей (k-Nearest Neighbours, KNN) 2) Метод Байеса 3) Метод опорных векторов (Support Vector Machines, SVM) 4) Линейная регрессия 5) Кластерный анализ	предложенных и обоснованием выбора	построения нелинейной плоскости, разделяющей решения, что обеспечивает высокую степень гибкости.
5.	<i>Прочитайте текст, выберите один правильный вариант ответа и запишите аргументы, обосновывающие выбор ответа:</i> Выберите раздел математики, который является наиболее важным для специалистов в области Data Science, занимая 90-95% времени в процессе анализа данных: 1) Алгебра 2) Геометрия 3) Теория вероятностей 4) Статистика 5) Математический анализ	Задание комбинированного типа с выбором одного верного ответа из предложенных и обоснованием выбора	4 Объяснение: Статистика является наиболее важным разделом математики для специалистов в области Data Science, так как разведочный анализ данных, их очистка, подготовка и конструирование новых признаков занимают основную часть времени и являются чистой статистикой.
6.	<i>Прочитайте текст, выберите один правильный вариант ответа и запишите аргументы, обосновывающие выбор ответа:</i> Выберите метод машинного обучения используется для обнаружения закономерностей в данных без предварительной информации о классах или категориях:	Задание комбинированного типа с выбором одного верного ответа из предложенных и обоснованием выбора	3 Объяснение: Кластеризация (например, k-средних) используется для обнаружения закономерностей в данных без предварительной информации о классах или категориях. Это метод обучения без учителя, который группирует данные на основе их сходства. Другие методы, такие как логистическая регрессия, метод

Номер задания	Формулировка задания	Тип задания	Ключ к оцениванию задания
	1) Логистическая регрессия 2) Метод опорных векторов (SVM) 3) Кластеризация (например, k-средних) 4) Линейная регрессия 5) Нейронные сети		опорных векторов, линейная регрессия и нейронные сети, в основном используются для задач классификации и регрессии, которые обычно требуют известных категорий или меток.
7.	<i>Прочитайте текст, выберите один правильный вариант ответа и запишите аргументы, обосновывающие выбор ответа:</i> Выберите инструмент, который используется в первую очередь используется для хранения и управления табличными данными: 1) SQL 2) Tableau 3) Python 4) R 5) Jupyter Notebook	Задание комбинированного типа с выбором одного верного ответа из предложенных и обоснованием выбора	1 Объяснение: SQL в первую очередь используется для хранения и управления данными. Это стандартный язык используемый для управления реляционными базами данных. Tableau – программный продукт используется для визуализации данных. Python и R - языки программирования, которые, хотя и могут использоваться для обработки данных, в первую очередь не используются для хранения и управления данными. Jupyter Notebook - интерактивная среда программирования для анализа и визуализации данных, но не инструмент для хранения данных.
8.	<i>Прочитайте текст, выберите один правильный вариант ответа и запишите аргументы, обосновывающие выбор ответа:</i> Выберите из следующих критериев один, используемый в анализе данных для проверки гипотезы о том, что распределение данных является нормальным: 1) Критерий Стьюдента (t-критерий) 2) Критерий хи-квадрат 3) Критерий Шапиро-Уилка 4) Критерий Манна-Уитни	Задание комбинированного типа с выбором одного верного ответа из предложенных и обоснованием выбора	3 Объяснение: Объяснение: Критерий Шапиро-Уилка используется для проверки гипотезы о том, что распределение данных является нормальным. Этот критерий проверяет, насколько распределение данных соответствует нормальному распределению. Другие критерии, такие как критерий Стьюдента, критерий хи-квадрат, критерий Манна-Уитни и критерий Крускала-Уоллиса, используются для других целей, таких как проверка

Номер задания	Формулировка задания	Тип задания	Ключ к оцениванию задания
	5) Критерий Крускала-Уоллиса		равенства средних, независимости переменных и сравнение распределений.
9.	<i>Прочитайте текст, выберите все правильные варианты ответов и запишите аргументы, обосновывающие выбор ответа:</i> Выберите языки программирования, которые часто используются для визуализации данных и обладает широкими возможностями для анализа данных: 1) Java 2) C++ 3) Python 4) SQL 5) R	Задания комбинированного типа с выбором нескольких верных ответов из предложенных и обоснованием выбора	3,5 Объяснение: Python широко используется для визуализации данных и обладает множеством библиотек, таких как Matplotlib, Seaborn и Plotly, которые позволяют создавать разнообразные графики и диаграммы для анализа данных. Язык программирования R часто используется в анализе данных из-за своей специализированности в области статистики, гибкости, мощных инструментов визуализации, возможности интеграции с другими технологиями, активного сообщества и доступности как открытого и бесплатного инструмента.
10.	<i>Прочитайте текст, выберите один правильный вариант ответа и запишите аргументы, обосновывающие выбор ответа:</i> Выберите из следующих компонентов один, который является ключевым элементом сверточной нейронной сети (CNN) и позволяет извлекать локальные особенности из входных данных: 1) Полносвязный слой 2) Сверточный слой 3) Пуллинг слой 4) Рекуррентный слой 5) LSTM-блок	Задание комбинированного типа с выбором одного верного ответа из предложенных и обоснованием выбора	2 Объяснение: Сверточный слой (Convolutional Layer) является ключевым элементом сверточной нейронной сети (CNN). Он использует фильтры (ядра) для извлечения локальных особенностей из входных данных, таких как изображений. Сверточные слои позволяют сети обнаруживать простые особенности на низких уровнях (например, края и углы на изображениях) и более сложные особенности на более высоких уровнях. Полносвязный слой (Fully Connected Layer): Используется в конце сети для классификации или регрессии на основе извлеченных особенностей. Пуллинг слой (Pooling Layer): Служит для уменьшения пространственного размера данных, сохраняя основные особенности.

Номер задания	Формулировка задания	Тип задания	Ключ к оцениванию задания
			Рекуррентный слой (Recurrent Layer): Используется в рекуррентных нейронных сетях (RNN) для обработки последовательных данных. LSTM-блок (Long Short-Term Memory block): Это тип рекуррентного слоя, используемый в LSTM-сетях для обработки временных рядов и последовательных данных.
6 семестр (очная форма обучения) // 4 курс (заочная форма обучения)			
ПК-8 Способен вести базы данных и поддержку информационного обеспечения решения прикладных задач			
11	<i>Прочитайте текст, выберите один правильный вариант ответа и запишите аргументы, обосновывающие выбор ответа:</i> Выберите одно из перечисленных требований к платформе для ведения базы знаний, которое наиболее важно для обеспечения её эффективного использования в организации: 1) Наличие графического логотипа платформы 2) Поддержка только текстового формата 3) Возможность интеграции с другими системами (например, CRM) 4) Ограничение прав доступа до одного пользователя 5) Минимальный срок тестового периода	Задание комбинированного типа с выбором одного верного ответа из предложенных и обоснованием выбора	3 Объяснение: Ведение базы данных и обеспечение информационной поддержки прикладных задач требует тесной интеграции с существующими корпоративными системами, такими как CRM, ERP или системы управления проектами. Это позволяет автоматизировать обмен данными, повышает точность информации и упрощает анализ. Интеграционные возможности являются ключевым фактором при выборе платформы для хранения и обработки знаний.
12	<i>Прочитайте текст, выберите один правильный вариант ответа и запишите аргументы, обосновывающие выбор ответа:</i> Выберите команду в SQL, с помощью которой можно выбрать все записи из таблицы patients, где значение столбца age больше 30: 1) SELECT * FROM patients WHERE age > 30 *	Задание комбинированного типа с выбором одного верного ответа из предложенных и	1 Объяснение: Для выборки всех записей из таблицы patients, где возраст (age) больше 30, используется команда: SELECT * FROM patients WHERE age > 30; SELECT * — выбирает все столбцы.

Номер задания	Формулировка задания	Тип задания	Ключ к оцениванию задания
	2) SELECT age FROM patients IF age = 30 3) GET * FROM patients WHEN age >= 30 4) FETCH ALL FROM patients WHERE age < 30 5) LIST patients WHERE age > 30	обоснованием выбора	FROM patients — указывает источник данных. WHERE age > 30 — фильтрует строки по условию. Это стандартный синтаксис языка SQL для выполнения фильтрации данных в реляционной базе данных SQLite.
13	<i>Прочитайте текст, выберите один правильный вариант ответа и запишите аргументы, обосновывающие выбор ответа:</i> Выберите из перечисленных способов корректно реализующий условную конструкцию в Python для проверки, является ли число положительным: 1) if number > 0: print("Число положительное") 2) if number > 0 then: print("Число положительное") 3) if (number > 0) { print("Число положительное") } 4) IF number > 0: PRINT("Число положительное") 5) if number > 0: print("Число положительное")	Задание комбинированного типа с выбором одного верного ответа из предложенных и обоснованием выбора	1 Объяснение: Обоснование правильного ответа: В Python синтаксис условной конструкции if строится следующим образом: Условие записывается после ключевого слова if. После условия ставится двоеточие :. Блок кода, который должен выполняться при истинности условия, пишется с отступом (рекомендуется 4 пробела). Вариант 5 формально работает, но не рекомендуется к использованию из-за отсутствия отступа, что нарушает стандарты читаемости кода.
14	<i>Прочитайте текст, выберите все правильные варианты ответов и запишите аргументы, обосновывающие выбор ответа:</i> Выберите один из перечисленных модулей стандартной библиотеки Python, который позволяет взаимодействовать с базой данных через SQL запросы: 1) sqlite3	Задания комбинированного типа с выбором нескольких верных ответов из предложенных и обоснованием выбора	1,2,3,4 Объяснение: Обоснование правильного ответа: Модуль sqlite3 является частью стандартной библиотеки Python и предназначен для работы с локальной базой данных SQLite. Он позволяет создавать, читать и обновлять данные без необходимости установки дополнительных пакетов.

Номер задания	Формулировка задания	Тип задания	Ключ к оцениванию задания
	2) sqlalchemy 3) psycpg2 4) mysql.connector 5) pymongo		sqlalchemy — это ORM-библиотека, не входит в стандартную библиотеку. psycpg2 — используется для работы с PostgreSQL. mysql.connector — применяется для MySQL. pymongo — работает с документной NoSQL БД MongoDB.
15	<i>Прочитайте текст, выберите один правильный вариант ответа и запишите аргументы, обосновывающие выбор ответа:</i> Выберите из следующих отраслей одну, для которой наиболее эффективно используется анализ данных приводящий к повышению производительности и оптимизации процессов: 1) Варианты ответов: 2) Сельское хозяйство 3) Образование 4) Телекоммуникации (мульти сервисные платформы) 5) Культура 6) Спорт	Задание комбинированного типа с выбором одного верного ответа из предложенных и обоснованием выбора	3 Объяснение: Телекоммуникации. В телекоммуникационной отрасли анализ данных используется для множества целей, таких как прогнозирование оттока клиентов, оптимизация сетевых ресурсов, предотвращение мошенничества и улучшение качества обслуживания. Эти приложения напрямую влияют на производительность и экономические показатели компании. В других отраслях, таких как сельское хозяйство, образование, культура и спорт, анализ данных также важен, но его влияние на производительность и оптимизацию процессов может быть менее очевидным или менее критичным.
16	<i>Прочитайте текст, выберите один правильный вариант ответа и запишите аргументы, обосновывающие выбор ответа:</i> Какая из перечисленных услуг относится к концепции "Инфраструктура как сервис" (IaaS)? 1) Google Docs 2) Microsoft Azure Virtual Machines * 3) Salesforce 4) Adobe Creative Cloud	Задание комбинированного типа с выбором одного верного ответа из предложенных и обоснованием выбора	2 Объяснение: Microsoft Azure Virtual Machines. IaaS предоставляет пользователям доступ к вычислительным ресурсам, хранилищам и сетям в облаке, позволяя им развертывать и управлять операционными системами и приложениями. Azure Virtual Machines — это сервис, который предоставляет виртуальные машины, которые являются ключевым компонентом IaaS. Google Docs,

Номер задания	Формулировка задания	Тип задания	Ключ к оцениванию задания
	5) Dropbox.		Salesforce, Adobe Creative Cloud и Dropbox — это примеры SaaS, где пользователям предоставляется готовое программное обеспечение
17	<i>Прочитайте текст, выберите один правильный вариант ответа и запишите аргументы, обосновывающие выбор ответа:</i> Ваша компания решила перенести локальную базу данных SQL Server в российское облако, чтобы обеспечить выполнение требований 152-ФЗ. Какой подход будет наиболее эффективным для такого переноса с минимальным временем простоя? 1) Ручное создание базы данных "с нуля" в облаке и написание скриптов импорта данных. 2) Использование физического резервного копирования (bak-файлы) ручного восстановления в облачной СУБД. 3) Применение встроенного сервиса миграции данных, предоставляемого российской облачной платформой (например, VK Cloud Solutions, СберОблако, Яндекс.Облако) 4) Экспорт данных в CSV-файлы с последующей загрузкой их через облачное хранилище (Object Storage). 5) Использование зарубежного сервиса миграции Azure Database Migration Service.	Задание комбинированного типа с выбором одного верного ответа из предложенных и обоснованием выбора	З Объяснение: Специализированные сервисы: Крупные российские облачные провайдеры (VK Cloud Solutions, СберОблако, Яндекс.Облако) предлагают встроенные сервисы миграции данных. Эти сервисы разработаны специально для переноса баз данных с минимальным простоем. Они могут реплицировать изменения с локального сервера в облачный экземпляр в реальном времени, создавая его "клон" без остановки основной системы. Соответствие требованиям: Использование сервиса от российского облачного провайдера напрямую обеспечивает выполнение требований 152-ФЗ о хранении и обработке данных на территории РФ. Почему не другие варианты: Ручное создание: Крайне трудоемко, высокая вероятность ошибок, большой простой. Ручное восстановление bak: Обычно требует остановки источника на время создания bak-файла и пока идет перенос/восстановление, что создает длительный простой. Требует совместимости версий СУБД. CSV: Потеря типов данных, связей, индексов, триггеров. Требует полной остановки на время экспорта/импорта. Высок риск потери данных и целостности. Зарубежный сервис: Использование зарубежного сервиса миграции потенциально нарушает требования 152-ФЗ, так как данные во время переноса могли бы обрабатываться сервисом, находящимся за пределами

Номер задания	Формулировка задания	Тип задания	Ключ к оцениванию задания
			РФ. Кроме того, его интеграция с российскими облаками может быть сложной или невозможной.
18	<i>Прочитайте текст, выберите один правильный вариант ответа и запишите аргументы, обосновывающие выбор ответа:</i> Для анализа больших объёмов исторических данных о продажах сети магазинов (например, для выявления долгосрочных трендов и сезонности) оптимальным решением будет использование: 1) Реляционной базы данных (например, PostgreSQL или Microsoft SQL Server), оптимизированной для транзакционной обработки (OLT) 2) Файловых таблиц в Microsoft Excel или Google Sheets 3) Системы управления хранилищем данных (Data Warehouse, DWH), например, на базе ClickHouse, Greenplum или Yandex DataLens 4) Системы потоковой обработки данных (например, Apache Kafk 5) Озера данных (Data Lake) в неструктурированном виде (например, набор файлов .txt в облачном хранилище)	Задание комбинированного типа с выбором одного верного ответа из предложенных и обоснованием выбора	3 Объяснение: Система управления хранилищем данных (DWH)** или специализированная аналитическая платформа — наиболее эффективный и производительный инструмент для поставленной задачи анализа исторических данных на предмет трендов и сезонности.
19	<i>Прочитайте текст, выберите один правильный вариант ответа и запишите аргументы, обосновывающие выбор ответа:</i> Команда аналитиков электронной коммерции хочет автоматически выявлять группы покупателей со схожими паттернами поведения (например, "часто покупают со скидками", "любят премиум-бренды",	Задание комбинированного типа с выбором одного верного ответа из предложенных и	3 Объяснение: Решается цель — выявление групп (сегментов) покупателей со схожими характеристиками. Кластеризация — это основной метод Data Mining для решения задач сегментации без предварительно заданных меток. Она группирует объекты (клиентов) на

Номер задания	Формулировка задания	Тип задания	Ключ к оцениванию задания
	"редко совершают повторные покупки") на основе истории транзакций, демографии и просмотров товаров. Цель — интегрировать эти группы в существующую CRM-систему для персонализации маркетинга. Какой метод Data Mining будет наиболее практичным и интерпретируемым для первоначального решения этой задачи? 1) Ассоциативные правила (например, алгоритм Apriori для поиска часто покупаемых вместе товаров. 2) Деревья решений для прогнозирования вероятности оттока конкретного клиента. 3) Кластеризация (например, алгоритм K-means или DBSCAN для сегментации клиентов на группы по схожести признаков. * 4) Глубокое обучение (нейронные сети) для создания комплексной модели предсказания поведения. 5) Анализ временных рядов для прогнозирования будущих продаж по категориям.	обоснованием выбора	основе меры сходства их признаков (история покупок, демография, активность). Также решается Практичность и интерпретируемость.
20	<i>Прочитайте текст, выберите один правильный вариант ответа и запишите аргументы, обосновывающие выбор ответа:</i> Сеть супермаркетов хочет внедрить систему автоматического прогнозирования ежедневного спроса на скоропортящиеся товары (молоко, хлеб) для оптимизации закупок и минимизации потерь. Исторические данные включают ежедневные продажи за 2 года, информацию о	Задание комбинированного типа с выбором одного верного ответа из предложенных и обоснованием выбора	2 Объяснение: Решается прямая цель — прогнозирование ежедневного спроса (количественная величина основе временных рядов (исторические данные с временными метками). Методы анализа временных рядов (ARIMA, экспоненциальное сглаживание специально разработаны для моделирования и прогнозирования данных, зависящих от времени.

Номер задания	Формулировка задания	Тип задания	Ключ к оцениванию задания
	выходных/праздниках и среднюю температуру дня. Какой метод анализа данных будет наиболее эффективным и интерпретируемым для построения первоначальной прогнозной модели в рамках этой прикладной задачи? 1) Кластеризация (K-meap для группировки дней со схожим спросом. 2) Модель временных рядов (например, ARIMA или экспоненциальное сглаживание с учетом сезонности и внешних факторов (праздники, температура). * 3) Ассоциативные правила (Apriori) для поиска связей между покупками разных товаров. 4) Дерево решений для классификации дней на "высокий спрос" / "низкий спрос".		