

Документ подписан простой электронной подписью  
Информация о владельце:  
ФИО: Таскаев Сергей Валерьевич  
Должность: Ректор  
Дата подписания: 05.09.2025 11:19:49  
Уникальный программный ключ  
04c19ed8bfb98f5b6cb77a486b9a8788b8322523



МИНОБРНАУКИ РОССИИ  
Федеральное государственное бюджетное  
образовательное учреждение высшего образования  
«Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)

Фонд оценочных средств для промежуточной аттестации по дисциплине (модулю) «Анализ данных» по  
направлению подготовки 38.04.05 «Бизнес-информатика» направленности «Информационная бизнес-  
аналитика» ФГБОУ ВО «ЧелГУ»

стр. 1

Фонд оценочных средств для промежуточной аттестации по дисциплине (модулю)  
**«Анализ данных»**

Направление подготовки (специальность)  
**38.04.05 «Бизнес-информатика»**

Направленность (профиль)  
**«Информационная бизнес-аналитика»**

Присваиваемая квалификация  
**Магистр**

Форма обучения  
**Очная**

Год набора  
**2025**

Челябинск, 2025 г.

**38.04.05 Бизнес-информатика, Информационная бизнес-аналитика, магистр,  
Анализ данных, 2025, очная**

**Фонд оценочных средств дисциплины (модуля) одобрен и рекомендован**

Проректор по учебной работе      утверждено 24.02.2025      А.А. Саламатов

Ученым советом института информационных технологий

Протокол заседания № 6 от 20.02.2025

Председатель Ученого совета  
института информационных  
технологий

согласовано

Ю. В. Петриченко

**Заседанием кафедры информационных технологий и экономической  
информатики**

Протокол заседания № 6 от 20.02.2025

И. о. заведующего кафедрой

согласовано

С.А. Скрипов

Автор (составитель)

А.В. Мельников

**Структура рабочей программы соответствует приказу ректора ФГБОУ ВО  
«ЧелГУ» от «13» апреля 2021 г. № 247-1**



## Содержание

1. Паспорт фонда оценочных средств .....	3
2. Перечень формируемых компетенций .....	4
3. Содержание оценочных средств по дисциплине .....	6
3.1. Виды оценочных средств .....	6
3.2. Содержание оценочных средств .....	7
4. Порядок проведения и критерии оценивания промежуточной аттестации .....	44
4.1. Порядок проведения промежуточной аттестации .....	44
4.2. Критерии оценивания промежуточной аттестации по видам оценочных средств .....	44
4.3. Результаты промежуточной аттестации и уровни сформированности компетенций.....	44



## 1. Паспорт фонда оценочных средств

Направление подготовки: 38.04.05 Бизнес-информатика.

Направленность: Информационная бизнес-аналитика

Дисциплина: Анализ данных.

Семестр: 2

Форма промежуточной аттестации: зачёт

Для оценивания результатов обучения используется балльно-рейтинговая система.



## 2. Перечень формируемых компетенций

Изучение дисциплины «Анализ данных» направлено на формирование компетенций, приведённых в 1.

Таблица 1. Результаты обучения по дисциплине.

Коды компетенции и согласно ФГОС (ОПОП ВО)	Содержание компетенций согласно ФГОС (ОПОП ВО)	Индикаторы достижения компетенции согласно ОПОП	Перечень планируемых результатов обучения по дисциплине
1	2	3	4
ПК-2	Способность собирать, обрабатывать и интерпретировать данные, используя методы машинного обучения и технологии бизнес-аналитики	ПК-2.1. Обладает знаниями методов и алгоритмов машинного обучения, средств и технологий сбора, обработки и представления данных ПК-2.2. Демонстрирует умения выбора методов машинного обучения и технологий анализа данных исходя из требований к решению прикладных задач ПК-2.3. Имеет практический опыт разработки новых алгоритмических решений в прикладных задачах анализа данных	Знать: Подходы и методы анализа информационно-технологической инфраструктуры предприятия Уметь: Работать с различными информационными системами для решения задач Владеть: Навыками обработки данных при решении задач с использованием информационных систем
ПК-3	Способен управлять проектно-внедренческими коллективами в процессе выработки структурных и технологических решений в области информационно-коммуникационных технологий	ПК-3.1 Уметь идентифицировать ключевые методы и методологии управления проектами, а также основные принципы работы в области информационно-коммуникационных технологий. ПК-3.2 Уметь организовывать и координировать деятельность проектно-внедренческих команд, а также эффективно распределять задачи и сроки среди участников команды с учетом их навыков и ресурсов. ПК-3.3 Владеть навыками лидерства, позволяющими мотивировать и вдохновлять команду для достижения поставленных целей,	Знать: возможности анализа информации на основе методов ИАД, основные понятия, задачи, стадии и методы ИАД, построение и использование моделей для анализа данных, применения подходов ИАД в современных областях экономики, науки и образования. Уметь: выполнять комплексный анализ информации с целью принятия решений на основе использования современных методов и систем для ИАД(интеллектуальный анализ данных). Владеть: методами ИАД с целью поддержки принятия решений в современных областях экономики, науки и образования.



МИНОБРНАУКИ РОССИИ  
Федеральное государственное бюджетное  
образовательное учреждение высшего образования  
«Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)

Фонд оценочных средств для промежуточной аттестации по дисциплине (модулю) «Анализ данных» по направлению подготовки 38.04.05 «Бизнес-информатика» направленности «Информационная бизнес-аналитика» ФГБОУ ВО «ЧелГУ»

стр. 5

		проводить анализ результатов проекта для дальнейшего его улучшения.	
--	--	---	--



### 3. Содержание оценочных средств по дисциплине

#### 3.1. Виды оценочных средств

Таблица 2. Виды оценочных средств.

№ п/п	Код компетенции/ планируемые результаты обучения	Контролируемые темы/ разделы	Наименование оценочного средства для текущего контроля	Наименование оценочного средства на промежуточной аттестации/№ задания
1	ПК-2.1. Обладает знаниями методов и алгоритмов машинного обучения, средств и технологий сбора, обработки и представления данных Знать: Подходы и методы анализа информационно-технологической инфраструктуры предприятия	Понятие Data Mining Деревья принятия решений, классификация Кластеризация	Опрос по практической работе, проверка отчета по выполнению практической работы	Задания теста № 1-355
2	ПК-2.2. Демонстрирует умения выбора методов машинного обучения и технологий анализа данных исходя из требований к решению прикладных задач Уметь: Работать с различными информационными системами для решения задач	Понятие Data Mining Деревья принятия решений, классификация Кластеризация	Опрос по практической работе, проверка отчета по выполнению практической работы	Задания теста № 1-355
3	ПК-2.3. Имеет практический опыт разработки новых алгоритмических решений в прикладных задачах анализа данных Владеть: Навыками обработки данных при решении задач с использованием информационных систем	Понятие Data Mining Деревья принятия решений, классификация Кластеризация	Опрос по практической работе, проверка отчета по выполнению практической работы	Задания теста № 1-355
4	ПК-3.1 Уметь идентифицировать ключевые методы и методологии управления проектами, а также основные принципы работы в области информационно-коммуникационных технологий. Знать: возможности анализа	Понятие Data Mining Деревья принятия решений, классификация Кластеризация	Опрос по практической работе, проверка отчета по выполнению практической работы	Задания теста № 1-355



	информации на основе методов ИАД, основные понятия, задачи, стадии и методы ИАД, построение и использование моделей для анализа данных, применения подходов ИАД в современных областях экономики, науки и образования.			
5	ПК-3.2 Уметь организовывать и координировать деятельность проектно-внедренческих команд, а также эффективно распределять задачи и сроки среди участников команды с учетом их навыков и ресурсов. Уметь: выполнять комплексный анализ информации с целью принятия решений на основе использования современных методов и систем для ИАД(интеллектуальный анализ данных).	Понятие Data Mining Деревья принятия решений, классификация Кластеризация	Опрос по практической работе, проверка отчета по выполнению практической работы	Задания теста № 1-355
6	ПК-3.3 Владеть навыками лидерства, позволяющими мотивировать и вдохновлять команду для достижения поставленных целей, проводить анализ результатов проекта для дальнейшего его улучшения. Владеть: методами ИАД с целью поддержки принятия решений в современных областях экономики, науки и образования.	Понятие Data Mining Деревья принятия решений, классификация Кластеризация	Опрос по практической работе, проверка отчета по выполнению практической работы	Задания теста № 1-355

Типовые задания, критерии и показатели оценивания в рамках текущего контроля представлены в рабочей программе дисциплины (модуля). Полные комплекты оценочных средств и контрольно-измерительных материалов хранятся на кафедре.

### 3.2. Содержание оценочных средств

#### База тестовых вопросов

№ п/п	Формулировка вопроса	Варианты ответов
1.	Data Mining — это процесс	(1) ранее сформулированных гипотез



	обнаружения в сырых данных	(2) неочевидных закономерностей (3) практических закономерностей (4) объективных закономерностей (5) большого количества закономерностей
2.	Какая из перечисленных дисциплин более сосредоточена на теории проверки гипотез?	(1) Data Mining (2) статистика (3) визуализация
3.	Подготовка данных в процессе Data Mining является:	(1) необязательным этапом работы (2) существенным этапом работы (3) может вообще отсутствовать
4.	Data Mining — это процесс обнаружения в сырых данных знаний, необходимых для:	(1) принятия решений в различных сферах человеческой деятельности (2) замены аналитика в процессе принятия решений (3) увеличения стоимости анализа данных
5.	В процессе работы Data Mining программы пользователь может получить такие результаты:	(1) большой процент ложных, недостоверных или бессмысленных результатов (2) только верные результаты, ложные выводы исключены (3) только статистически достоверные результаты
6.	Назовите факторы, обусловившие возникновение и развитие Data Mining:	(1) совершенствование аппаратного и программного обеспечения (2) совершенствование технологий хранения и записи данных (3) накопление большого количества ретроспективных данных (4) совершенствование алгоритмов обработки информации (5) необходимость замены аналитика информационной технологий
7.	Если сравнивать Data Mining, машинное обучение и статистику, какая из дисциплин сконцентрирована на едином процессе анализа данных, включает очистку данных, обучение, интеграцию и визуализацию результатов:	(1) Data Mining (2) машинное обучение (3) статистика
8.	Инструменты Data Mining:	(1) могут самостоятельно строить гипотезы о взаимосвязях в данных (2) не могут самостоятельно строить гипотезы о взаимосвязях в данных (3) могут самостоятельно строить гипотезы о взаимосвязях в данных, которые обязательно подтверждаются
9.	Закономерности, найденные в процессе использования технологии Data Mining должны обладать такими свойствами:	(1) быть очевидными (2) быть неочевидными (3) быть практически полезными (4) быть объективными (5) чем больше найдено закономерностей, тем лучше



10.	В результате использования инструментов Data Mining пользователь может ...	(1) получить гипотезы о взаимосвязях в данных, самостоятельно выдвинутые инструментом Data Mining (2) получить подтверждение или опровержение гипотез, выдвинутых пользователем (3) проверить гипотезы о взаимосвязях в данных, самостоятельно выдвинутые пользователем инструмента Data Mining (4) все ответы верны
11.	Оцените правильность утверждения: "Data Mining может заменить аналитика"	(1) утверждение верно (2) утверждение неверно. Технология не может дать ответы на те вопросы, которые не были заданы (3) утверждение неверно. Технология всего лишь дает аналитику инструмент для облегчения и улучшения его работы
12.	Какая из перечисленных ниже групп методов достаточно часто использует для выявления взаимосвязей в данных концепцию усреднения по выборке?	(1) Data Mining (2) статистические методы (3) OLAP
13.	Выберите характеристику, наиболее подходящую для Data Mining	(1) подходит для понимания ретроспективных данных (2) опирается на ретроспективные данные для получения ответов на вопросы о будущем (3) подходит для обобщения ретроспективных данных
14.	Оцените правильность утверждения: "Извлечение полезных сведений невозможно без хорошего понимания сути данных"	(1) утверждение верно (2) утверждение неверно. Технологии Data Mining не нужен аналитик, поэтому понимание кем-либо данных - излишне (3) утверждение неверно. Технологии не нужно понимание данных
15.	Данные представляют собой:	(1) факты и графики (2) текст (3) картинки, звуки, аналоговые или цифровые видеосегменты (4) все вместе
16.	Объектом не является:	(1) запись (2) случай (3) пример (4) строка таблицы (5) переменная
17.	Номинальная шкала – это шкала,	(1) содержащая только категории, которые не могут упорядочиваться (2) содержащая категории, которые могут упорядочиваться (3) содержащая только две категории
18.	Данные могут быть получены в результате:	(1) измерений (2) экспериментов (3) арифметических и логических операций (4) всего вместе



19.	Порядковая шкала – это шкала, содержащая	(1) категории, которые могут упорядочиваться (2) категории, которые не могут упорядочиваться (3) только две категории
20.	Данные — это ...	(1) необработанный материал, предоставляемый поставщиками данных и используемый потребителями для формирования информации на основе данных (2) готовый материал для формирования информации (3) синоним информации
21.	Строка таблицы также известна как:	(1) запись (2) атрибут (3) пример (4) переменная
22.	Интервальная шкала – это шкала,	(1) содержащая категории, которые могут упорядочиваться, однако разности не имеют смысла (2) разности между значениями которой могут быть вычислены, однако их отношения не имеет смысла (3) содержащая только категории, которые не могут упорядочиваться
23.	Объект описывается как ...	(1) набор атрибутов (2) свойство, характеризующее объект (3) поле таблицы
24.	Такие данные как температура воздуха относятся к ...	(1) непрерывным данным (2) дискретным данным (3) Оба ответа неверны
25.	Для какой шкалы применимы только такие операции как равно и не равно?	(1) номинальная шкала (2) порядковая шкала (3) интервальная шкала
26.	Атрибут – это:	(1) свойство, характеризующее объект (2) строка таблицы (3) случай или пример
27.	К какой категории данных относится вес измеряемых объектов:	(1) непрерывным данным (2) дискретным данным (3) оба ответа неверны
28.	Для какой шкалы применимы только такие операции как равно, не равно, больше, меньше?	(1) номинальная шкала (2) порядковая шкала (3) интервальная шкала
29.	Какие из перечисленных ниже пунктов являются названиями стадий Data Mining?	(1) свободный поиск (2) прогностическое моделирование (3) анализ исключений (4) индукция правил
30.	Стадия свободного поиска представлена действиями:	(1) выявление закономерностей условной логики (2) выявление закономерностей ассоциативной логики (3) выявление трендов и колебаний (4) предсказание неизвестных значений



31.	На какие две группы подразделяются методы Data Mining по принципу работы с исходными обучающими данными?	(1) непосредственное использование данных или сохранение данных (2) выявление и использование формализованных закономерностей (3) статистические методы (4) кибернетические методы
32.	Какой из перечисленных ниже пунктов не является названием стадии Data Mining?	(1) свободный поиск (2) прогностическое моделирование (3) анализ исключений (4) индукция правил
33.	Прогностическое моделирование включает такие действия:	(1) выявление трендов и колебаний (2) предсказание неизвестных значений (3) прогнозирование развития процессов
34.	Нейронные сети относятся к группам ...	(1) статистических методов (2) методов на основе уравнений (3) методов кросс-табуляции
35.	На стадии свободного поиска осуществляется ...	(1) выявление закономерностей (2) использование выявленных закономерностей для предсказания неизвестных значений (3) анализ исключений
36.	Частью какой из перечисленных стадий является валидация закономерностей?	(1) свободный поиск (2) прогностическое моделирование (3) анализ исключений
37.	Деревья решений относятся к группе (-ам) ...	(1) статистических методов (2) кибернетических методов (3) логических методов (4) методов кросс-табуляции
38.	Большинство аналитических методов, используемые в технологии Data mining – это ...	(1) известные математические алгоритмы и методы (2) новейшие математические алгоритмы и методы (3) классические статистические методы
39.	Какая из перечисленных ниже стадий может считаться дополнительной или частью одной из основных стадий Data mining:	(1) выявление закономерностей (свободный поиск) (2) использование выявленных закономерностей для предсказания неизвестных значений (прогностическое моделирование) (3) валидация
40.	Регрессионный и дискриминантный анализ ...	(1) относятся к статистическим методам Data mining (2) относятся к кибернетическим методам Data mining (3) не являются методами Data mining
41.	Большинство методов Data mining были разработаны в рамках ...	(1) теории искусственного интеллекта (2) классического анализа данных (3) теории баз данных
42.	Нечеткая логика и деревья решений ...	(1) относятся к статистическим методам Data mining (2) относятся к кибернетическим методам Data mining (3) не являются методами Data mining



43.	В ходе решения какой из перечисленных задач устанавливаются закономерности между связанными событиями в наборе данных?	(1) задачи поиска ассоциативных правил (2) задачи поиска последовательных ассоциативных правил (3) задачи анализа отклонений
44.	К классу описательных задач Data Mining относятся такие задачи:	(1) прогнозирование (2) классификация (3) кластеризация (4) визуализация
45.	Любые, неизвестные ранее сведения о каком-либо событии, сущности, процессе и т.п., являющиеся объектом некоторых операций, для которых существует содержательная интерпретация, являются...	(1) данными (2) информацией (3) знаниями
46.	В ходе решения каких из перечисленных задач устанавливаются закономерности между событиями, связанными во времени?	(1) задачи поиска ассоциативных правил (2) задачи поиска последовательных ассоциативных правил (3) задачи анализа отклонений
47.	К классу прогнозирующих задач Data Mining относятся такие задачи:	(1) прогнозирование (2) классификация (3) кластеризация (4) визуализация
48.	Формирование какой из перечисленных категорий происходит в процессе сбора и передачи данных, т.е. их обработки?	(1) информации (2) знаний (3) и того, и другого
49.	Правильна ли такая формулировка: "Ассоциация является частным случаем последовательности с временным лагом, равным нулю"?	(1) формулировка верна (2) нет; последовательность является частным случаем ассоциации (3) нет; ни последовательность, ни ассоциация не являются частными случаями друг друга
50.	Продолжите фразу: "Кластеризация и классификация относятся к..."	(1) стратегии обучения с учителем" (2) стратегии обучения без учителя" (3) к двум разным стратегиям: обучения без учителя и обучения с учителем"
51.	Совокупность фактов, закономерностей и эвристических правил, с помощью которых решается поставленная задача, – это ...	(1) данные (2) информация (3) знания
52.	Согласно классификации по стратегиям, задачи Data Mining подразделяются на:	(1) обучение с учителем (2) обучение без учителя (3) дескриптивные (4) прогнозирующие
53.	Заполните пропуск в формулировке: "Формирование ... происходит в процессе сбора и передачи, т.е. обработки данных"	(1) знаний (2) информации (3) данных
54.	У основания так называемой	(1) данные



	информационной пирамиды находится категория ...	(2) знания (3) информация
55.	Задачи Data Mining, в зависимости от используемых моделей подразделяются на:	(1) обучение с учителем (2) обучение без учителя (3) дескриптивные (4) прогнозирующие
56.	В результате использования одних и тех же данных и различных методов...	(1) должна появляться разная информация (2) должна появляться только одинаковая информация (3) может появляться разная информация, это зависит от выбранных методов обработки данных
57.	Информация, данные и знания являются:	(1) частью одного потока (2) частями разных потоков (3) оба ответа неверны
58.	Классификация — это ...	(1) отнесение объектов к одному из заранее известных классов (2) отнесение объектов к одной из заранее неизвестных групп (3) процесс формирования групп и отнесения объектов к одному из них.
59.	Задачу классификации нельзя решить с помощью...	(1) метода деревьев решений (2) метода линейной регрессии (3) алгоритма Apriori
60.	Множество примеров, используемое для конструирования модели, называется...	(1) обучающим множеством (2) тестовым множеством (3) проверочным множеством
61.	Задачей классификации можно назвать предсказание...	(1) категориальной зависимой переменной, основываясь на выборке непрерывных и/или категориальных переменных (2) числовой зависимой переменной, основываясь на выборке непрерывных и/или категориальных переменных (3) порядковой зависимой переменной, основываясь на выборке непрерывных и/или категориальных переменных
62.	Задачи классификации решаются следующими алгоритмами:	(1) нейронные сети (2) линейной регрессии (3) алгоритмом Apriori
63.	Процесс классификации состоит из следующих этапов:	(1) конструирование модели (2) использование модели (3) определение вида модели
64.	Основная характеристика задачи бинарной классификации:	(1) зависимая переменная может принимать только два значения (2) классификация осуществляется по одному признаку (3) классификация осуществляется по двум признакам
65.	Множество примеров, используемое для проверки работы	(1) обучающим множеством (2) тестовым множеством



	сконструированной модели, называется...	(3) тренировочным множеством
66.	Заполните пропуск в формулировке: "Для проведения ... должны присутствовать признаки, характеризующие группу, к которой принадлежит то или иное событие или объект"	(1) классификации (2) кластеризации (3) классификации и кластеризации
67.	Классификация относится к стратегии:	(1) обучения с учителем (2) обучения без учителя (3) оба ответа неверны
68.	Кластер можно охарактеризовать как ...	(1) группу объектов, имеющих общие свойства (2) один объект, изолированный от других (3) группу объектов, имеющую внутреннюю однородность
69.	Метод деревьев решений применяется для решения задач ...	(1) классификации (2) кластеризации (3) классификации и кластеризации
70.	Задачей классификации часто является предсказание ...	(1) числовой зависимой переменной (2) категориальной зависимой переменной (3) категориальной независимой переменной
71.	Классификация относится к:	(1) контролируемому обучению (2) управляемому обучению (3) обучения без учителя
72.	Изначальная предопределенность классов является характеристикой задачи ...	(1) классификации (2) кластеризации (3) классификации и кластеризации
73.	Иерархические алгоритмы применяются для решения задач ...	(1) классификации (2) кластеризации (3) классификации и кластеризации
74.	Какой из параметров является основной единицей времени, на которую делается прогноз?	(1) период прогнозирования (2) горизонт прогнозирования (3) интервал прогнозирования
75.	В чем сходство задач классификации и прогнозирования?	(1) при решении обеих задач используется 2-х этапный процесс построения модели и ее использования для предсказания (2) в результате решения этих задач предсказывается класс независимой переменной (3) результат решения этих задач - предсказание будущих числовых значений зависимой переменной
76.	Возможности визуализации включают:	(1) поддержку интерактивного и согласованного исследования (2) помощь в представлении результатов (3) формализацию задач Data Mining
77.	Какой из параметров является числом	(1) период прогнозирования



	периодов в будущем, которые покрывает прогноз?	(2) горизонт прогнозирования (3) интервал прогнозирования
78.	Продолжите фразу: "Прогнозирование будет иметь смысл, если горизонт прогнозирования ..."	(1) не меньше, чем время, которое необходимо для реализации решения, принятого на основе прогноза" (2) не больше, чем время, которое необходимо для реализации решения, принятого на основе прогноза" (3) оба ответа верны
79.	Продолжите фразу: "Визуализация ..."	(1) направлена исключительно на совершенствование техники анализа" (2) может самостоятельно выполнять функции анализа" (3) может самостоятельно выполнять функции анализа, но пользователь должен иметь специальную подготовку"
80.	Какой из параметров является частотой, с которой делается новый прогноз?	(1) период прогнозирования (2) горизонт прогнозирования (3) интервал прогнозирования
81.	Решение задачи прогнозирования ...	(1) возможно без обучающей выборки данных (2) требует некоторой обучающей выборки данных (3) является решением задачи "обучения без учителя"
82.	Преимуществом визуализации является:	(1) простота ее использования (2) возможность решать самые разнообразные задачи (3) отсутствие необходимости специальной подготовки пользователя
83.	В чем состоит основное сходство задач прогнозирования и классификации?	(1) при решении обеих задач используется двухэтапный процесс построения модели на основе обучающего набора и ее использования для предсказания неизвестных значений зависимой переменной (2) сходство заключается в том, что при решении обеих задач предсказываются числовые значения зависимой переменной (3) оба ответа верны
84.	Временной ряд — последовательность наблюдаемых значений какого-либо признака,...	(1) упорядоченных в неслучайные моменты времени (2) упорядоченных в случайные моменты времени (3) не обязательно упорядоченных, но зафиксированных в неслучайные моменты времени
85.	Параметрами прогнозирования являются:	(1) период прогнозирования (2) горизонт прогнозирования (3) интервал прогнозирования (4) тренд
86.	Характеристики визуализации:	(1) может помочь в представлении результатов Data Mining (2) иногда может ввести пользователя в заблуждение (3) всегда дает верное представление о данных
87.	В чем состоит основное отличие задач прогнозирования и классификации?	(1) отличие заключается в этапах процесса решения задач (2) отличие задач классификации и прогнозирования состоит в том, что в первой задаче предсказывается класс зависимой переменной, а во второй - числовые значения



		зависимой переменной
88.	Отличием анализа временных рядов от анализа случайных выборок является:	(1) предположение о равных промежутках времени между наблюдениями (2) их хронологический порядок (3) оба варианта верны
89.	Период прогнозирования – это ...	(1) параметр прогнозирования (2) составляющая временного ряда (3) характеристика временного ряда
90.	Оцените правильность утверждения: "Визуализация направлена исключительно на совершенствование техники анализа"	(1) утверждение верно (2) утверждение неверно. Визуализация не направлена на совершенствование техники анализа (3) утверждение неверно. Визуализация может самостоятельно выполнять функции анализа
91.	Выделите два основных направления Web Mining:	(1) Web Content Mining (2) Web Usage Mining (3) Web Text Mining
92.	При использовании какого из перечисленных ниже направлений выделяют подход, основанный на агентах, и подход, основанных на базах данных:	(1) Web Content Mining (2) Web Usage Mining (3) Web Text Mining
93.	Какие задачи возникают перед разработчиками при построении системы Web Mining:	(1) сбора данных (2) использование методов персонификации (3) анализ полученного знания
94.	Какое из перечисленных ниже направлений подразумевает автоматический поиск и извлечение качественной информации разнообразных источников Интернета, перегруженных "информационным шумом":	(1) Web Content Mining (2) Web Usage Mining (3) Web Text Mining
95.	Какие из перечисленных ниже систем используются в подходе, основанном на агентах:	(1) интеллектуальные поисковые агенты (2) фильтрация информации / классификация (3) персонифицированные агенты сети (4) многоуровневые базы данных (5) системы web-запросов
96.	Выделите категории задач Web Mining:	(1) предварительная обработка данных для Web Mining (2) обнаружение шаблонов и открытие знаний с использованием ассоциативных правил, временных последовательностей, классификации и кластеризации (3) анализ полученного знания (4) использование методов персонификации
97.	Какое из перечисленных ниже направлений подразумевает обнаружение закономерностей в действиях пользователя web-узла или их группы?	(1) Web Content Mining (2) Web Usage Mining (3) Web Text Mining



98.	Какие из перечисленных ниже систем используются в подходе, основанном на базах данных:	(1) интеллектуальные поисковые агенты (2) фильтрация информации / классификация (3) персонифицированные агенты сети (4) многоуровневые базы данных (5) системы web-запросов
99.	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
100.	Область использования Data Mining ...	(1) ничем не ограничена — она везде, где имеются какие-либо данные (2) ничем не ограничена — она везде, не имеет значения, есть ли какие-либо данные (3) достаточно ограничена, в большинстве случаев – это научные исследования
101.	Технология Web mining применяет технологию Data Mining для анализа:	(1) неструктурированной информации (2) структурированной информации (3) неоднородной информации (4) однородной информации (5) распределенной и значительной по объему информации (6) информации, содержащейся на Web-узлах
102.	Web content mining подразумевает ...	(1) автоматический поиск и извлечение качественной информации разнообразных источников Интернета, перегруженных "информационным шумом" (2) обнаружение закономерностей в действиях пользователя Web-узла или их группы (3) обнаружение последовательности просмотра страниц
103.	Выделяют такие основные сферы применения технологии Data Mining:	(1) наука (2) бизнес (3) исследования для правительства (4) Web-направление
104.	Согласно таксономии Web Mining выделяют основные направления:	(1) Web Content Mining и Web Usage Mining (2) Text Mining и Call Mining (3) Web Content Mining, Web Usage Mining, Text Mining и Call Mining
105.	Web Usage Mining подразумевает ...	(1) автоматический поиск и извлечение качественной информации разнообразных источников Интернета, перегруженных "информационным шумом" (2) обнаружение закономерностей в действиях пользователя Web-узла или их группы (3) обнаружение последовательности просмотра страниц
106.	Характеристики измерения центральной тенденции:	(1) среднее (2) медиана (3) минимум (4) дисперсия
107.	Медианой для выборки 1,2,3,7,10,__,16 является:	(1) 7,714286 (2) 7



		(3) 8,5 (4) рассчитать медиану невозможно из-за пропущенных значений
108.	Выберите соответствующую характеристику данному варианту связи: большие значения из одного набора данных связаны с большими значениями другого набора (положительная корреляция), и это...	(1) наличие прямой (линейной) связи (2) наличие отрицательной линейной связи (3) отсутствие линейной связи
109.	Характеристики вариации данных:	(1) среднее (2) медиана (3) минимум (4) дисперсия
110.	Медианой для выборки 1, __, 3, 7, 10, 15, 16, 18 является:	(1) 7,714286 (2) 7 (3) 8,5 (4) рассчитать медиану невозможно из-за пропущенных значений
111.	Выберите соответствующую характеристику данному варианту связи: малые значения одного набора связаны с большими значениями другого (отрицательная корреляция), и это...	(1) наличие прямой (линейной) связи (2) наличие отрицательной линейной связи (3) отсутствие линейной связи
112.	Какая из перечисленных характеристик не является числом, описывающим определенным способом все значения признака набора данных?	(1) среднее (2) медиана (3) минимум
113.	Медианой для выборки 1, 2, 3, __, __, 15, 16, 18 является:	(1) рассчитать медиану невозможно из-за пропущенных значений (2) 7 (3) 8,5
114.	Выберите соответствующую характеристику данному варианту связи: данные двух диапазонов никак не связаны (нулевая корреляция), и это...	(1) наличие прямой (линейной) связи (2) наличие отрицательной линейной связи (3) отсутствие линейной связи
115.	Размах и дисперсия являются:	(1) характеристиками измерения центральной тенденции (2) характеристиками вариации данных (3) определяют наличие выбросов в данных
116.	Заполните пропуск в формулировке: "Корреляционный анализ применяется для ... оценки взаимосвязи двух наборов данных, представленных в безразмерном виде"	(1) количественной (2) качественной (3) количественной и качественной
117.	Основные особенности регрессионного анализа заключаются в том, что при его помощи можно получить конкретные сведения о том:	(1) какую форму имеет зависимость между исследуемыми переменными (2) какой характер имеет зависимость между исследуемыми переменными (3) какую количественную взаимосвязь имеют два набора данных



118.	Среднее и медиана являются:	(1) характеристиками центральной тенденции (2) характеристиками вариации данных (3) определяют наличие выбросов в данных
119.	Заполните пропуск в формулировке: "Коэффициент корреляции Пирсона, который является безразмерным индексом в интервале ... включительно, отражает степень ... зависимости между двумя множествами данных"	(1) от -1 до 1, линейной (2) от 0 до 1, линейной (3) от 0 до 100, нелинейной
120.	Основные задачи регрессионного анализа включают:	(1) установление формы зависимости (2) определение функции регрессии (3) оценку неизвестных значений зависимой переменной (4) все ответы верны
121.	Если зависимая переменная принимает дискретные значения, при помощи метода дерева решений решается задача:	(1) классификации (2) численного прогнозирования (3) классификации и численного прогнозирования
122.	Внутренний узел дерева решений называют также ...	(1) узлом проверки (2) конечным узлом (3) вершиной (4) листом
123.	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 использованы для построения дерева
124.	Если зависимая переменная принимает непрерывные значения, то дерево решений решает задачу:	(1) классификации (2) численного прогнозирования (3) классификации и численного прогнозирования
125.	Конечный узел дерева решений называют также ...	(1) узлом проверки (2) узлом решения (3) листом
126.	Алгоритмы конструирования деревьев решений ...	(1) имеют возможность обработки пропущенных значений вне зависимости от используемого алгоритма (2) в принципе не могут работать с пропущенными значениями (3) большинство алгоритмов конструирования деревьев решений имеют возможность обработки пропущенных значений
127.	При помощи метода деревьев решений могут решаться задачи:	(1) классификации (2) численного прогнозирования



		(3) классификации и численного прогнозирования
128.	Лист дерева решений является ...	(1) конечным узлом (2) узлом проверки (3) узлом решения
129.	Какие модели строят деревья решений?	(1) непараметрические модели (2) параметрические модели (3) и те, и другие
130.	При помощи метода деревьев решений решаются задачи ...	(1) классификации и прогнозирования (2) кластеризации и прогнозирования (3) классификации и кластеризации (4) кластеризации, классификации и прогнозирования
131.	Заполните пропуски в формулировке: "Каждая ветвь дерева, идущая от внутреннего узла, отмечена ... , который может относиться лишь к одному ... данного узла"	(1) атрибутами расщепления, критерию расщепления (2) предикатом расщепления, атрибуту расщепления (3) критерием расщепления, атрибуту расщепления
132.	Процесс создания дерева ...	(1) происходит сверху вниз, т.е. является нисходящим (2) происходит снизу вверх, т.е. является восходящим (3) может быть как нисходящим, так и восходящим
133.	Заполните пропуски в формулировке: "Если зависимая переменная принимает ... значения, при помощи метода дерева решений ..."	(1) дискретные, решается задача классификации (2) дискретные, решаются задачи классификации и прогнозирования (3) непрерывные, решаются задачи классификации и кластеризации
134.	Внутренние узлы дерева решений называют ...	(1) атрибутами расщепления (2) предикатом расщепления (3) критерием расщепления
135.	Процесс отсечения ветвей или замена некоторых ветвей поддеревом ...	(1) происходит снизу вверх, т.е. является восходящим (2) происходит сверху вниз, т.е. является нисходящим (3) может быть как нисходящим, так и восходящим
136.	В методе опорных векторов для классификации используется ...	(1) не все множество образцов, а лишь их небольшая часть, которая находится на границах (2) все множество образцов (3) часть образцов, которая находится вне границ
137.	Решаются ли задачи классификации и регрессии при помощи метода "ближайшего соседа"?	(1) да (2) нет, только задачи классификации (3) нет, только задачи регрессии
138.	Все переменные являются одинаково важными и статистически независимыми, т.е. значение одной переменной ничего не говорит о значении другой. Это свойства:	(1) наивной байесовской классификации (2) метода "ближайшего соседа" (3) метода опорных векторов
139.	Классификация методом опорных векторов считается хорошей, если область между границами ...	(1) пуста (2) минимально заполнена (3) максимально заполнена



140.	Метод "ближайшего соседа":	(1) может создавать модели и правила. (2) может создавать модели (3) может создавать правила (4) не может создавать модели и правила
141.	На результат классификации в наивно-байесовском подходе влияют:	(1) только индивидуальные значения входных переменных (2) комбинированное влияние пар или троек значений разных атрибутов (3) индивидуальные значения входных переменных, комбинированное влияние пар или троек значений разных атрибутов
142.	Если область между границами пуста, классификация ...	(1) считается хорошей (2) считается ненадежной (3) невозможной
143.	С помощью метода "ближайшего соседа" возможно решение задач:	(1) классификации и регрессии (2) классификации и кластеризации (3) классификации
144.	Использование байесовских сетей имеет следующие преимущества:	(1) позволяет избежать проблемы переучивания (2) определяет зависимости между всеми переменными (3) на результат классификации влияют только индивидуальные значения входных переменных
145.	Назовите метод, недостаток которого приведен ниже: "Существует сложность выбора меры "близости", от этой меры главным образом зависит объем множества записей, которые нужно хранить в памяти для достижения удовлетворительной классификации или прогноза"	(1) метод байесовской классификации (2) метод "к-ближайших соседей" (3) метод опорных векторов
146.	Назовите свойства наивной байесовской классификации:	(1) использование всех переменных и определение всех зависимостей между ними (2) наличие предположения относительно того, что все переменные являются одинаково важными (3) наличие предположения относительно того, что все переменные являются статистически независимыми, т.е. значение одной переменной ничего не говорит о значении другой
147.	В основе метода опорных векторов лежит ...	(1) понятие плоскостей решений (2) предположение о взаимной независимости признаков (3) предположение о взаимной зависимости признаков
148.	Метод, который делает заключения относительно данной ситуации по результатам поиска аналогий, хранящихся в базе прецедентов относится к категории ...	(1) "обучение без учителя" (2) "обучение с учителем" (3) самообучающейся системы
149.	Назовите метод, недостаток которого приведен ниже: "Перемножать условные вероятности корректно только тогда, когда все входные	(1) метод байесовской классификации (2) метод "к-ближайших соседей" (3) метод опорных векторов



	переменные действительно статистически независимы"	
150.	Группа синапсов нейрона – это ...	(1) однонаправленные входные связи, соединенные с выходами других нейронов (2) выходная связь данного нейрона, с которой сигнал (возбуждения или торможения) поступает на синапсы следующих нейронов (3) один или несколько нейронов, на входы которых подается один и тот же общий сигнал
151.	В синхронных нейронных сетях в каждый момент времени свое состояние меняет...	(1) лишь один нейрон (2) целая группа нейронов, как правило, весь слой (3) возможен и тот и другой вариант
152.	Явление переобучения характеризуется ...	(1) чрезмерно точным соответствием нейронной сети конкретному набору обучающих примеров, при котором сеть теряет способность к обобщению (2) возникновением, в случае слишком долгого обучения, недостаточного числа обучающих примеров или слишком сложной структуры нейронной сети (3) возникновением, в случае слишком долгого обучения, слишком большого числа обучающих примеров или слишком сложной структуры нейронной сети
153.	Нейрон имеет аксон, который представляет собой ...	(1) выходную связь данного нейрона, с которой сигнал (возбуждения или торможения) поступает на синапсы следующих нейронов (2) однонаправленные входные связи, соединенные с выходами других нейронов (3) один или несколько нейронов, на входы которых подается один и тот же общий сигнал
154.	В асинхронных сетях в каждый момент времени свое состояние меняет ...	(1) лишь один нейрон (2) целая группа нейронов, как правило, весь слой (3) возможен и тот и другой вариант
155.	Многослойный персептрон – это сеть ...	(1) прямого распространения сигнала (без обратных связей) (2) обратного распространения сигнала (с обратными связями) (3) в которой входной сигнал преобразуется в выходной, проходя последовательно через несколько слоев
156.	Слой нейронной сети – это ...	(1) один или несколько нейронов, на входы которых подается один и тот же общий сигнал (2) выходная связь данного нейрона, с которой сигнал (возбуждения или торможения) поступает на синапсы следующих нейронов (3) однонаправленные входные связи, соединенные с выходами других нейронов
157.	Синхронные и асинхронные сети отличаются:	(1) принципом изменения состояния нейронов (состояние изменяется либо у одного нейрона, либо у их группы) (2) принципом обработки информации (последовательно либо всем нейронам сети)



		(3) оба варианта верны
158.	В многослойном персептроне ...	(1) должен быть хотя бы один скрытый слой (2) может быть какое угодно количество скрытых слоев, они также могут вообще отсутствовать (3) присутствие нескольких скрытых слоев оправдано лишь в случае использования нелинейных функций активации
159.	Главная функция искусственного нейрона - ...	(1) формирование выходного сигнала в зависимости от сигналов, поступающих на его входы (2) преобразование функции активации (3) передача входных сигналов на обработку адаптивному сумматору
160.	Заполните пропуски в формулировке: "... - выходная связь нейрона, с которой сигнал (возбуждения или торможения) поступает на... следующих нейронов"	(1) аксон, синапсы (2) синапс, аксоны (3) аксон, точку ветвления
161.	Ошибкой обучения нейронной сети называется ...	(1) разность между желаемым и полученным на выходе сигналами (2) целевая функция, требующая минимизации в процессе управляемого обучения нейронной сети (3) переобучение нейронной сети
162.	Заполните пропуски в формулировке: "В самой распространенной конфигурации входные сигналы обрабатываются ..., затем выходной сигнал сумматора поступает в ..., где преобразуется функцией активации, и результат подается на ..."	(1) адаптивным сумматором, нелинейный преобразователь, выход (2) нелинейным преобразователем, адаптивный сумматор, выход (3) входным сумматором, нелинейный преобразователь, адаптивный сумматор
163.	Однонаправленные входные связи, соединенные с выходами других нейронов – это ...	(1) синапсы (2) аксоны (3) слои сети
164.	Целевая функция, требующая минимизации в процессе управляемого обучения нейронной сети – это ...	(1) функция ошибок (2) ошибка обучения (3) функция переобучения
165.	Сети без обратных связей - это ...	(1) персептрон (2) сети Хопфилда (задачи ассоциативной памяти) (3) сети Кохонена (задачи кластерного анализа)
166.	Сеть Кохонена представляет собой ...	(1) два слоя: входной и выходной (2) три слоя: входной, скрытый и выходной (3) входной, выходной и какое угодно количество скрытых слоев
167.	Правило формирования окрестности (несколько нейронов, которые окружают нейрон-победитель):	(1) сначала к окрестности принадлежит большое число нейронов, далее ее размер постепенно уменьшается (2) сначала к окрестности принадлежит малое число нейронов, далее ее размер постепенно увеличивается (3) размер окрестности в процессе обучения не



		изменяется
168.	Сети с обратными связями – это...	(1) персептрон (2) сети Хопфилда (задачи ассоциативной памяти) (3) сети Кохонена (задачи кластерного анализа)
169.	Наиболее распространенное применение сетей Кохонена:	(1) разведочный анализ данных (2) обнаружение новых явлений (3) прогнозирование числовых значений
170.	Самоорганизующиеся сети в процессе обучения подстраиваются ...	(1) под закономерности во входных данных (2) под эталонное значение выхода (3) под закономерности в выходных данных
171.	Сети Кохонена относятся к классу:	(1) сети с обратными связями (2) сети без обратных связей (3) рекуррентных сетей
172.	Уникальность метода самоорганизующихся карт состоит в ...	(1) преобразовании n-мерного пространства в двухмерное (2) возможности преобразования n-мерного пространства в пространство с любым количеством измерений (3) преобразовании двухмерного пространства в n-мерное
173.	Обучение самоорганизующихся сетей заключается ...	(1) в минимизации ошибки (2) в подстройке весов (внутренних параметров нейросети) для наибольшего совпадения с входными данными (3) в подстройке весов (внутренних параметров нейросети) для наибольшего совпадения с выходными данными
174.	Наличие блоков динамической задержки и обратных связей - характерная особенность ...	(1) рекуррентных сетей (2) сетей прямого распространения (3) и тех, и других
175.	При ... для каждого обучающего входного примера требуется знание правильного ответа или функции оценки качества ответа	(1) "обучении с учителем" (2) "обучении без учителя" (3) оба варианта верны
176.	Какое количество слоев имеет сеть Кохонена?	(1) один слой: только входной (2) два слоя: входной и выходной (3) три слоя: входной, выходной и скрытый (4) какое угодно количество слоев (5) она вообще не имеет слоев
177.	Закончите фразу: "В процессе обучения сетей Кохонена на входы подаются данные, сеть при этом подстраивается..."	(1) не под закономерности во входных данных, а под эталонное значение выхода (2) не под эталонное значение выхода, а под закономерности во входных данных (3) не под закономерности во входных данных, а под коэффициенты весов
178.	Сети Кохонена относятся к классам:	(1) сети с обратными связями (2) сети без обратных связей (3) сети прямого распространения



179.	При... раскрывается внутренняя структура данных или корреляции между образцами в наборе данных	(1) обучении с учителем (2) обучении без учителя (3) оба варианта верны
180.	Традиционно темно-синие участки на карте Кохонена соответствуют ...	(1) наименьшим значениям показателя (2) самым высоким значениям показателя (3) средним значениям показателя
181.	Обучение сетей Кохонена заключается ....	(1) не в минимизации ошибки, а в подстройке весов (2) не в подстройке весов, а в минимизации ошибки (3) не в подстройке весов, а в минимизации их коэффициентов
182.	Спорный объект кластеризации — это объект, который по мере сходства ...	(1) может быть отнесен к нескольким кластерам (2) не может быть отнесен ни к одному кластеру (3) может быть отнесен более чем к двум кластерам
183.	Работа кластерного анализа опирается на предположения:	(1) рассматриваемые признаки объекта в принципе допускают желательное разбиение объектов на кластеры (2) отнесение всех объектов к одному из предопределенных признаков (3) о сравнимости шкал
184.	Процедура, которая приводит значения всех преобразованных переменных к единому диапазону значений путем выражения через отношение этих значений к некоей величине, отражающей определенные свойства, это – ...	(1) стандартизация (2) нормирование (3) оба ответа верны
185.	Работа кластерного анализа опирается на следующие предположения (выберите неверный ответ):	(1) рассматриваемые признаки объекта в принципе допускают желательное разбиение объектов на кластеры (2) правильность выбора масштаба или единиц измерения признаков (3) отнесение всех объектов к одному из предопределенных признаков
186.	Иерархические агломеративные методы характеризуются ...	(1) последовательным объединением исходных элементов и соответствующим уменьшением числа кластеров (2) делением одного кластера на меньшие кластеры, в результате образуется последовательность расщепляющих групп (3) сопоставлением фиксированного числа кластеров наблюдения кластерам так, что средние в кластере максимально возможно отличаются друг от друга
187.	Объект относится к кластеру, если ...	(1) расстояние от объекта до центра кластера меньше радиуса кластера (2) расстояние от объекта до центра кластера меньше диаметра кластера (3) расстояние от объекта до центра кластера больше радиуса кластера
188.	Иерархические дивизимные методы характеризуются ...	(1) последовательным объединением исходных элементов и соответствующим уменьшением числа



		кластеров (2) делением одного кластера на меньшие кластеры, в результате образуется последовательность расщепляющих групп (3) сопоставлением фиксированного числа кластеров наблюдения кластерам так, что средние в кластере максимально возможно отличаются друг от друга
189.	Назовите характеристики кластерного анализа:	(1) позволяет сокращать размерность данных (2) позволяет делать данные более наглядными (3) имеет в своем арсенале около десяти алгоритмов (4) имеет в своем арсенале около сотни алгоритмов
190.	При применении кластерного анализа переменные ...	(1) должны измеряться в сравнимых шкалах (2) могут измеряться в каких угодно шкалах (3) должны быть только числовыми
191.	Характеристикой каких групп методов являются последовательное объединение исходных элементов и соответствующее уменьшение числа кластеров?	(1) иерархические агломеративные методы (2) иерархические дивизимные (делимые) методы (3) и тех, и других
192.	Кластерный анализ ...	(1) может применяться к совокупностям временных рядов (2) не может применяться к совокупностям временных рядов (3) может определять группы временных рядов со схожей динамикой
193.	Деление одного кластера на меньшие кластеры, в результате чего образуется последовательность расщепляющих групп. Характеристика каких групп методов описана выше?	(1) иерархические агломеративные методы (2) иерархические дивизимные (делимые) методы (3) и тех, и других
194.	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
195.	К какой группе методов относится метод ближнего соседа?	(1) иерархический КА (2) быстрый КА (3) оба ответа неверны
196.	Процесс последовательного укрупнения кластеров лежит в основе работы...	(1) иерархического КА (2) быстрого (3) всех методов
197.	Основные меры расстояния между объектами при использовании иерархического метода КА:	(1) евклидово расстояние (2) квадрат евклидова расстояния (3) манхэттенское расстояние (4) расстояние Чебышева
198.	Какой метод требует априорной информации о количестве кластеров?	(1) метод k-средних (2) метод ближнего соседа (3) вся группа иерархических методов (4) все ответы неверны



199.	К какой группе методов относится метод k-средних?	(1) иерархический (2) быстрый (3) оба ответа неверны
200.	Пересчет кластерных центров и перераспределение объектов между кластерными центрами – это шаги...	(1) иерархического (2) быстрого (3) всех методов
201.	Перед началом кластеризации все объекты считаются отдельными кластерами, которые в ходе алгоритма объединяются. Это характеристика...	(1) иерархического (2) быстрого (3) всех методов
202.	Какой метод рекомендуется использовать при небольших объемах выборки?	(1) метод k-средних (2) метод ближнего соседа (3) вся группа иерархических методов (4) все ответы неверны
203.	К какой группе методов относится алгоритм PAM (partitioning around Medoids)?	(1) иерархический КА (2) КА, основанный на разделении данных (3) оба ответа неверны
204.	Чувствительность к выбросам – это недостаток...	(1) иерархического КА (2) быстрого КА (3) всех методов КА
205.	Дендрограмма – результат работы ...	(1) дивизимного кластерного анализа (2) быстрого кластерного анализа (3) агломеративного кластерного анализа
206.	Назовите достоинства алгоритма кластеризации k-средних	(1) простота использования (2) быстрота использования (3) понятность и прозрачность алгоритма (4) нечувствительность к выбросам
207.	Какие методы выявляют более высокую устойчивость по отношению к шумам и выбросам, некорректному выбору метрики, включению незначимых переменных в набор, участвующий в кластеризации?	(1) неиерархические методы (2) иерархические методы (3) оба ответа верны
208.	Какие методы отказываются от определения числа кластеров, а строят полное дерево вложенных кластеров?	(1) иерархические методы (2) неиерархические методы (3) самоорганизующиеся карты
209.	Назовите недостатки алгоритма быстрой кластеризации	(1) сложность использования (2) чувствительность к выбросам (3) алгоритм может медленно работать на больших базах данных (4) все ответы верны
210.	Назовите сложности иерархических методов кластеризации:	(1) ограничение объема набора данных (2) выбор меры близости (3) негибкость полученных классификаций (4) наличие предположений относительно числа кластеров.



211.	Преимуществом какой группы методов кластеризации является их наглядность и возможность получить детальное представление о структуре данных	(1) иерархические методы (2) неиерархические методы (3) оба варианта верны
212.	Набор называют часто встречающимся (frequent), если:	(1) его поддержка выше определенного пользователем минимального значения (2) его поддержка ниже определенного пользователем максимального значения (3) его поддержка равна определенному пользователем значению
213.	С помощью алгоритма Apriori определите часто встречающиеся наборы в базе данных D, состоящие из трех товаров с минимальной поддержкой, равной 2	(1) n,l,o (2) k,l,m (3) k,l (4) n,m,o
214.	Выберите задачу, которая не решается при помощи поиска ассоциативных правил:	(1) определение товаров, которые стоит продвигать совместно (2) выбор местоположения товара в магазине (3) классификация клиентов фирмы на однотипные группы (4) анализ потребительской корзины
215.	Транзакция – это множество событий, которые произошли ...	(1) одновременно (2) одно за другим (3) оба ответа неверны
216.	Достоверность ассоциативного правила определяет...	(1) количество транзакций, содержащих определенный набор данных (2) какая вероятность того, что из события A следует событие B (3) процент транзакций, содержащих определенный набор данных
217.	Поддержка ассоциативного правила определяет...	(1) количество транзакций, содержащих определенный набор данных (2) какая вероятность того, что из события A следует событие B (3) процент транзакций, содержащих определенный набор данных
218.	Выберите правильное утверждение:	(1) чем больше значение поддержки правила, тем лучше правило (2) чем ниже значение поддержки правила, тем лучше правило (3) если поддержка правила слишком велика, в результате будут найдены правила очевидные и хорошо известные
219.	Набор ассоциативных правил представляет интерес, если его поддержка ...	(1) выше определенного пользователем минимального значения (2) ниже определенного пользователем минимального значения (3) равна определенному пользователем значению



220.	Вероятность того, что из события А следует событие В. Это - ...	(1) достоверность правила (2) поддержка правила (3) обеспечение правила
221.	Назовите алгоритмы, при помощи которых осуществляется поиск ассоциативных правил:	(1) алгоритм AIS (2) алгоритм SETM (3) алгоритм Apriori (4) алгоритм PAM
222.	Каждый этап работы алгоритма Apriori состоит из таких шагов:	(1) формирование кандидатов (2) кодирование кандидатов (3) подсчет кандидатов
223.	Количество транзакций, содержащих определенный набор данных. Это - ...	(1) достоверность набора (2) поддержка набора (3) обеспечение набора
224.	Назовите алгоритм, который не осуществляет поиск ассоциативных правил:	(1) алгоритм DHP (2) алгоритм PAM (3) алгоритм DIC
225.	Традиционные методы визуализации могут находить следующее применение:	(1) представлять пользователю информацию в наглядном виде (2) компактно описывать закономерности, присущие исходному набору данных (3) снижение размерности или сжатие информации (4) восстановление пробелов в наборе данных (5) все ответы верны
226.	Способы визуального представления могут ...	(1) быть иллюстрацией построения модели (2) помочь интерпретировать полученный результат (3) быть средством оценки качества построенной модели
227.	Основные тенденции в области визуализации:	(1) разработка сложных видов диаграмм (2) увеличение размеров и сложности структур данных, представляемых визуализацией (3) уменьшение роли пользователя (4) все ответы верны
228.	Нахождение шумов и выбросов в данных ...	(1) возможно при помощи средств визуализации (2) невозможно при помощи средств визуализации (3) не является функцией визуализации
229.	Визуализация в виде параллельных координат является представлением информации в ...	(1) двухмерном измерении (2) трехмерном измерении (3) более чем в трехмерном измерении
230.	Основными тенденциями в области визуализации являются:	(1) повышение уровня взаимодействия с визуализацией пользователя (2) уменьшения уровня взаимодействия с визуализацией пользователя (3) сведение роли пользователя во взаимодействии в визуализацией к минимальной
231.	Какие из перечисленных средств визуализации помогают	(1) дерево решений (2) представление графа нейронной сети



	интерпретировать полученный результат?	(3) дендрограмма
232.	Компактное описание закономерностей, присущих исходному набору данных ...	(1) возможно при помощи средств визуализации (2) невозможно при помощи средств визуализации (3) не является функцией визуализации
233.	Визуализация в виде "лиц Чернова" является представлением информации в ...	(1) двухмерном измерении (2) трехмерном измерении (3) более, чем в трехмерном измерении
234.	Назовите характеристики одной из основных тенденций в области визуализации:	(1) увеличение размеров структур данных, представляемых визуализацией (2) усложнение структур данных, представляемых визуализацией (3) уменьшение размеров структур данных, представляемых визуализацией
235.	Какие из перечисленных средств визуализации служат средством оценки качества построенной модели?	(1) карты входов самоорганизующихся сетей Кохонена (2) представление графа нейронной сети (3) таблица сопряженности
236.	Существенными концепциями системы поддержки принятия решений являются:	(1) компьютерная интерактивная (2) поддержка принятия решений (3) слабоструктурированных и неструктурированных проблем (4) структурированных и слабоструктурированных проблем
237.	Исходные данные при использовании MOLAP архитектуры хранятся:	(1) в многомерной БД или в многомерном локальном кубе (2) в реляционных БД (3) в плоских локальных таблицах на файл-сервере
238.	Предметная ориентация хранилища данных означает, что ...	(1) данные объединены в категории и сохраняются соответственно областям, которые они описывают, а не применениям, их использующим (2) данные удовлетворяют требованиям всего предприятия, а не одной функции бизнеса (3) хранилище можно рассматривать как совокупность "исторических" данных: возможно восстановление данных в любой момент времени
239.	Охарактеризуйте неструктурированные задачи	(1) имеют только качественное описание, основанное на суждениях ЛПР, количественные зависимости между основными характеристиками задачи не известны (2) характеризуются существенными зависимостями, которые могут быть выражены количественно (3) сочетают количественные и качественные зависимости, причем малоизвестные и неопределенные стороны задачи имеют тенденцию доминировать (4) именно такими проблемами занимаются руководители
240.	Исходные данные при использовании ROLAP архитектуры хранятся...	(1) в многомерной БД или в многомерном локальном кубе (2) в реляционных БД или в плоских локальных таблицах на файл-сервере (3) в реляционных БД, а агрегаты размещаются в



		многомерной БД
241.	Интегрированность хранилища данных означает, что ...	(1) данные объединены в категории и сохраняются соответственно областям, которые они описывают, а не применениям, их использующим (2) данные удовлетворяют требованиям всего предприятия, а не одной функции бизнеса (3) хранилище можно рассматривать как совокупность "исторических" данных: возможно восстановление данных в любой момент времени
242.	Охарактеризуйте слабоструктурированные задачи	(1) имеют только качественное описание, основанное на суждениях ЛПР, количественные зависимости между основными характеристиками задачи не известны (2) характеризуются существенными зависимостями, которые могут быть выражены количественно (3) сочетают количественные и качественные зависимости, причем малоизвестные и неопределенные стороны задачи имеют тенденцию доминировать (4) именно такими проблемами занимаются руководители
243.	Исходные данные при использовании HОLAP архитектуры хранятся:	(1) в многомерной БД или в многомерном локальном кубе (2) в реляционных БД или в плоских локальных таблицах на файл-сервере (3) в реляционной базе, а агрегаты размещаются в многомерной
244.	Привязка ко времени хранилища данных означает, что ...	(1) данные объединены в категории и сохраняются соответственно областям, которые они описывают, а не применениям, их использующим (2) данные удовлетворяют требованиям всего предприятия, а не одной функции бизнеса (3) хранилище можно рассматривать как совокупность "исторических" данных: возможно восстановление данных на любой момент времени
245.	СППР — система, предназначенная для поддержки принятия решений в ... проблемах различных видов человеческой деятельности, существенная концепцией которой ...	(1) слабоструктурированных и неструктурированных, не обуславливает обязательного непосредственного использования ЛПР системы поддержки принятия решений (2) структурированных и слабоструктурированных, обуславливает обязательное непосредственное использование ЛПР системы поддержки принятия решений (3) неструктурированных и структурированных, не обуславливает обязательного непосредственного использования ЛПР системы поддержки принятия решений
246.	EIS (Execution Information System) или информационные системы руководства в большинстве ориентированы на ..., основаны на ...	(1) неподготовленного пользователя, на запросах, количество которых ограничено (2) подготовленного пользователя, глубокой проработке данных (3) неподготовленного пользователя, глубокой проработке данных.



247.	MOLAP является:	<ol style="list-style-type: none"><li>(1) способом хранения данных в OLAP-системах</li><li>(2) архитектурой OLAP-серверов, при которой исходные и многомерные данные хранятся в многомерной БД или в многомерном локальном кубе</li><li>(3) методом Data Mining</li><li>(4) интеграцией Data Mining и OLAP</li><li>(5) архитектурой OLAP-серверов, при которой исходные данные хранятся в реляционных БД или в плоских локальных таблицах на файл-сервере</li></ol>
248.	Назовите основные концепции хранилища данных:	<ol style="list-style-type: none"><li>(1) предметная ориентация</li><li>(2) интегрированность</li><li>(3) привязка ко времени</li><li>(4) отсутствие привязки ко времени</li><li>(5) периодическая изменяемость данных</li></ol>
249.	Назовите существенные характеристики СППР:	<ol style="list-style-type: none"><li>(1) решение принимает человек</li><li>(2) решение принимает система</li><li>(3) предназначена для решения слабоструктурированных задач</li><li>(4) предназначена для решения неструктурированных задач</li><li>(5) предназначена для решения структурированных задач</li></ol>
250.	DSS (Decision Support System) ориентированы на ..., основаны на ...	<ol style="list-style-type: none"><li>(1) неподготовленного пользователя, на запросах, количество которых ограничено</li><li>(2) подготовленного пользователя, глубокой проработке данных</li><li>(3) неподготовленного пользователя, глубокой проработке данных</li></ol>
251.	ROLAP является:	<ol style="list-style-type: none"><li>(1) архитектурой OLAP-серверов, при которой исходные данные хранятся в реляционных БД или в плоских локальных таблицах на файл-сервере</li><li>(2) интеграцией Data Mining и OLAP</li><li>(3) методом Data Mining</li><li>(4) архитектурой OLAP-серверов, при которой исходные данные остаются в реляционной базе, а агрегаты размещаются в многомерной</li></ol>
252.	Как называется характеристика хранилища данных описанная ниже: "Данные объединены в категории и сохраняются соответственно областям, которые они описывают, а не применениям, их использующим"	<ol style="list-style-type: none"><li>(1) предметная ориентация</li><li>(2) интегрированность</li><li>(3) привязка ко времени</li><li>(4) неизменность данных</li></ol>
253.	Для какого вида набора данных важно определение наличия сезонной компоненты:	<ol style="list-style-type: none"><li>(1) для упорядоченных данных</li><li>(2) для неупорядоченных данных</li><li>(3) для тех и других</li></ol>
254.	Выберите неверную формулировку	<ol style="list-style-type: none"><li>(1) все типы грязных данных в базе данных могут быть автоматически обнаружены и очищены</li><li>(2) появление некоторых грязных данных может быть предотвращено</li><li>(3) некоторые грязные данные непригодны для автоматического обнаружения и очистки</li></ol>



		(4) появление некоторых грязных данных невозможно предотвратить
255.	Наличие дубликатов в наборе данных может быть:	(1) результатом ошибок при подготовке данных (2) способом повышения значимости некоторых записей (3) оба ответа верны
256.	Если набор данных упорядочен и в нем присутствует сезонная или цикличная компонента, то каково минимальное количество данных, которое необходимо иметь для возможности анализа?	(1) данные за один сезон/цикл (2) данные за половину сезона/цикла (3) данные за два сезона/цикла
257.	Выберите верную(-ые) формулировку(-и).	(1) все типы грязных данных в базе данных могут быть автоматически обнаружены и очищены (2) появление некоторых грязных данных может быть предотвращено (3) некоторые грязные данные непригодны для автоматического обнаружения и очистки (4) появление некоторых грязных данных невозможно предотвратить
258.	При наличии дубликатов в наборе данных следует использовать такой вариант их обработки:	(1) удалить всю группу записей, содержащую дубликаты (2) заменить группу дубликатов на одну уникальную запись (3) оба варианта можно использовать при обработке дубликатов
259.	Если набор данных не упорядочен, то количество данных, которое необходимо иметь для возможности анализа данных:	(1) не имеет значения (2) должно быть больше одного цикла (3) желательно, чтобы количество записей в наборе данных было значительно больше количества переменных
260.	Все типы грязных данных в базе данных могут быть автоматически обнаружены и очищены	(1) формулировка неверна (2) формулировка верна (3) в зависимости от метода очистки данных
261.	Ваши действия при обнаружении выбросов в наборе данных:	(1) их следует сразу же исключить из дальнейшего анализа (2) оценить степень их влияния на результаты дальнейшего анализа (3) такой набор данных вообще не поддается анализу
262.	Какой из перечисленных этапов является первым в процессе Data Mining?	(1) анализ предметной области (2) подготовка данных (3) построение модели
263.	Если данные являются неупорядоченными, это означает ...	(1) невозможность процесса Data Mining (2) возможность процесса Data Mining (3) необходимость их упорядочения перед проведением анализа
264.	Качество данных – это критерий, определяющий такие качества данных	(1) полноту (2) точность



	как:	(3) своевременность (4) возможность их интерпретации (5) все варианты верны
265.	Постановка задачи ...	(1) является необходимым этапом процесса Data Mining (2) является необязательным этапом процесса Data Mining (3) не является этапом процесса Data Mining
266.	Репрезентативность выборки означает, что ...	(1) выборка должна представлять как можно больше возможных ситуаций (2) выборка должна включать более ста записей (3) число записей выборки должно соответствовать числу переменных
267.	Существуют такие типы грязных данных:	(1) данные, которые могут быть автоматически обнаружены и очищены (2) данные, которые не могут быть автоматически обнаружены и очищены (3) данные, появление которых можно было предотвратить (4) данные, появление которых невозможно было предотвратить
268.	Ошибки, которые возникают в процессе использования инструментов очистки (являющиеся двумя крайностями очистки данных) — это:	(1) решение инструментом очистки данных проблемы, которой на самом деле не существует (2) ошибки, возникающие, когда инструменты очистки полностью упускают существующую проблему (3) ошибки, возникающие, когда инструменты очистки не могут обнаружить существующую проблему
269.	Качественная программа очистки данных должна:	(1) не затрагивать правильные данные (2) исправлять неверные данные (3) создавать небольшой по объему отчет о подозрительных записях (4) требовать серьезного процесса установки и обслуживания
270.	Специальные средства очистки служат:	(1) для работы с конкретными областями (имена и адреса) (2) для работы по исключению дубликатов (3) для реализации возможности сложных преобразований и большей части технологического процесса преобразования и очистки данных
271.	Ошибка Типа 1 возникает в случае, когда ...	(1) инструмент очистки данных пытается решить проблему, которой на самом деле не существует (2) инструмент очистки данных полностью упускает существующую проблему (3) инструмент очистки данных не может обнаружить существующую проблему
272.	Инструменты ETL предназначены:	(1) для работы с конкретными областями (имена и адреса) (2) для работы по исключению дубликатов (3) для реализации возможности сложных преобразований и большей части технологического процесса преобразования и очистки данных



273.	Ошибка Типа 2 возникает в случае, когда ...	(1) инструмент очистки данных пытается решить проблему, которой на самом деле не существует (2) инструмент очистки полностью упускает существующую проблему (3) инструмент очистки не может обнаружить существующую проблему
274.	Качественная программа очистки данных должна иметь такие характеристики:	(1) исправлять неверные данные (2) создавать небольшой по объему отчет о подозрительных записях (3) требовать минимальных затрат на установку, обслуживание и ручные проверки (4) может частично затрагивать правильные данные
275.	Когда речь идет о создании банков данных всего предприятия и, соответственно, о сплошной очистке данных, имеет смысл пользоваться следующими средствами очистки данных:	(1) универсальными системами, предназначенными для обслуживания всей базы данных целиком (2) верификаторами имени/адреса для очистки только данных о клиентах (3) специальными средствами очистки данных
276.	Оцените правильность формулировки: "Инструменты Data Mining служат средством очистки данных"	(1) формулировка верна (2) формулировка неверна. Задача инструментов Data Mining совершенно другая. (3) некоторые инструменты Data Mining могут быть средством очистки данных
277.	Инструменты очистки данных обычно выполняют такие функции:	(1) парсинг (2) стандартизация (3) проверка допустимости (4) улучшение (5) согласование и консолидация (6) все ответы верны
278.	Автоматизированный процесс очистки данных ... к ошибкам в данных, которых ранее в них не было	(1) иногда может приводить (2) всегда приводит (3) не может приводить
279.	Согласно классификации средств очистки данных инструменты Data Mining относятся к классу ...	(1) средств анализа и модернизации данных (2) специальных средств очистки (3) очистки специфической области
280.	Согласно классификации ошибок в данных, которые возникают в результате использования средств очистки, выделяют такие их классы:	(1) ошибки, возникающие, когда инструмент очистки пытается решить проблему, которой на самом деле не существует (2) ошибки, возникающие, когда инструменты очистки полностью упускают существующую проблему (3) ошибки, требующие немедленного исправления (4) ошибки, не поддающиеся исправлению
281.	Построение моделей Data Mining осуществляется с целью:	(1) исследования или изучения моделируемого объекта и получения новых знаний, необходимых для принятия решений (2) выбора наиболее быстродействующей модели (3) исследования всех возможных свойств и характеристик изучаемого объекта



282.	Простота модели в сравнении с исследуемым объектом является ...	(1) преимуществом использования моделей (2) недостатком использования моделей (3) признаком невозможности использования модели
283.	Прогнозирующие модели Data Mining ...	(1) позволяют на основе выявленных закономерностей предсказывать будущее поведение объекта (2) описывают общие закономерности предметной области (3) решают задачи кластеризации, группировки, обобщения (4) все ответы неверны
284.	Характеристиками модели являются ...	(1) простота модели в сравнении с исследуемым объектом (2) выделение в объекте наиболее существенных факторов (3) абстрактность модели
285.	Преимуществом модели является возможность выделить в объекте ...	(1) наиболее существенные факторы, с точки зрения цели исследования, и не отвлекаться на маловажные детали (2) абсолютно все факторы, как существенные, так и маловажные (3) абстрактные факторы
286.	Классификационные модели Data Mining ...	(1) осуществляют прогнозирование класса объекта (2) описывают общие закономерности предметной области (3) решают задачи кластеризации, группировки, обобщения (4) все ответы верны
287.	Использование моделей Data Mining позволяет:	(1) определить наилучшее решение в конкретной ситуации (2) определить как существенные, так и незначительные факторы (3) оба ответа верны
288.	Модель обладает свойством неполноты.	(1) утверждение верно (2) утверждение неверно (3) утверждение неверно. Модель обладает свойством упрощать объект.
289.	Дескриптивные модели ...	(1) описывают общие закономерности предметной области (2) осуществляют прогнозирование класса объекта (3) решают задачи кластеризации, группировки, обобщения (4) все ответы верны
290.	Создание каких моделей Data Mining означает поиск правил, которые объясняют зависимость выходных параметров от входных?	(1) моделей классификации и прогнозирования (2) моделей кластеризации и классификации (3) моделей правил ассоциаций
291.	Экзогенные переменные — это переменные, которые ...	(1) задаются вне модели, они известны заранее (2) определяются по ходу расчетов в модели, они не



		задаются извне (3) задаются внутри модели, они известны заранее
292.	Назовите причины, из-за которых следует переобучать или обучать модель заново:	(1) изменились входящие данные или их поведение (2) появились дополнительные данные для обучения (3) изменились требования к форме и количеству выходных данных (4) изменились цели бизнеса, которые повлияли на критерии принятия решений (5) изменились внешнее окружение или среда (6) все ответы верны
293.	Какие модели используются для классификации объектов, при условии, что набор целевых классов неизвестен?	(1) модели кластеризации (2) модели кластеризации и классификации (3) модели правил ассоциаций
294.	Эндогенные переменные — это переменные, которые ...	(1) задаются вне модели, они известны заранее (2) определяются по ходу расчетов в модели, они не задаются извне (3) задаются внутри модели, они известны заранее
295.	Если модель с успехом используется определенное время, это означает, что ...	(1) ее не следует считать абсолютно верной на все времена (2) она проверена, и ее можно считать верной на все времена (3) она уже устарела, и нужна новая модель
296.	Data Mining это ... , который должен быть интегрирован в бизнес.	(1) не только инструмент, но также процесс (2) инструмент (3) процесс
297.	На каком этапе пересекается работа специалиста предметной области и специалиста по добыче данных?	(1) анализ бизнес-процессов (2) анализ данных (3) подготовка данных (4) все ответы неверны
298.	Специалист по анализу данных, который имеет, как минимум, основы статистических знаний и способен применять технологии Data Mining, а также интерпретировать полученные результаты - это...	(1) специалист по добыче данных (2) специалист предметной области (3) администратор баз данных (4) программист (5) все ответы неверны
299.	Data Mining по стандарту CRISP-DM включает следующие фазы:	(1) осмысление бизнеса (2) осмысление данных (3) подготовка данных (4) исследование отношений в данных
300.	Data Mining — это не только инструмент, но также процесс, который...	(1) должен быть интегрирован в бизнес (2) может существовать отдельно от бизнеса (3) должен предшествовать бизнесу
301.	На каком(-их) этапе(-ах) пересекается работа администратора баз данных и специалиста по добыче данных	(1) анализ бизнес-процессов (2) анализ данных (3) сбор данных (4) все ответы неверны



302.	Специалист, имеющий знания о окружении бизнеса, процессах, заказчиках, клиентах, потребителях, а также конкурентах - это...	(1) специалист по добыче данных (2) специалист предметной области (3) администратор баз данных (4) менеджер проекта (5) все ответы неверны
303.	Стандарт PMML относится к группе:	(1) стандартов по хранению и передаче моделей Data Mining (2) стандартов, относящиеся к унификации интерфейсов (3) стандартов, направленных на разработку надстройки над языком SQL
304.	Data Mining — это не только инструмент, но также ...	(1) процесс, который должен быть интегрирован в бизнес (2) процесс, который, однако, не может быть интегрирован в бизнес (3) процесс интеграции в бизнес
305.	На каком этапе пересекается работа специалиста предметной области и администратора баз данных?	(1) анализ бизнес-процессов (2) анализ данных (3) сбор данных (4) все ответы неверны
306.	Специалист, имеющий знания о том, где и каким образом хранятся данные, как получить к ним доступ, и как связать между собой эти данные - это...	(1) специалист по добыче данных (2) специалист предметной области (3) администратор баз данных (4) программист (5) все ответы неверны
307.	Стандарт CWM (Common Warehouse Metamodel) относится к группе:	(1) стандартов по хранению и передаче моделей Data Mining (2) стандартов, относящихся к унификации интерфейсов (3) стандартов, направленных на разработку надстройки над языком SQL
308.	Анализ предметной области и интерпретация результатов, полученных в результате Data Mining - это точки соприкосновения таких специалистов как:	(1) специалиста предметной области (2) специалиста по добыче данных (3) администратора баз данных (4) всех вместе
309.	Шаги какой из методологий Data Mining здесь описаны: осмысление бизнеса; осмысление данных; подготовка данных; моделирование; оценка результатов; внедрение?	(1) CRISP-DM (2) SEMMA (3) Two Crow
310.	Какой стандарт обеспечивает возможности обмена моделями данных между программным обеспечением разных разработчиков?	(1) PMML (2) CWM Data Mining (3) JDM
311.	Анализ требований к данным и сбор данных- это точки соприкосновения таких специалистов как:	(1) специалиста предметной области (2) специалиста по добыче данных (3) администратора баз данных (4) всех вместе
312.	Шаги какой из методологий Data Mining здесь описаны:	(1) CRISP-DM (2) SEMMA



	отбор данных, исследование отношений в данных, модификация данных, моделирование взаимозависимостей, оценка полученных моделей и результатов?	(3) Two Crow
313.	Какой стандарт обеспечивает поддержку наиболее распространенных прогнозных моделей, созданных при помощи алгоритмов и методов анализа данных?	(1) PMML (2) CWM Data Mining (3) JDM
314.	Существуют следующие варианты решений по внедрению инструментов Data Mining:	(1) только покупка готового инструмента, собственная разработка системы Data Mining практически невозможна (2) разработка Data Mining-продукта на заказ фирмой-разработчиком (3) оба варианта неверны
315.	На рынке инструментов Data Mining в последние годы наблюдается:	(1) спад (2) рост (3) ситуация на рынке за последние годы почти не меняется
316.	Цены на инструменты Data Mining уровня предприятия находятся в диапазоне:	(1) US \$10,000 и больше (2) от \$1,000 до \$9,999 (3) от \$1 до \$999
317.	На рынке инструментов Business Intelligence в последние годы наблюдается:	(1) значительный рост, в том числе инструментов Data Mining (2) значительный рост только сегмента инструментов Data Mining (3) значительный рост за исключением сегмента инструментов Data Mining
318.	Цены на инструменты Data Mining уровня отдела находятся в диапазоне:	(1) US \$10,000 и больше (2) от \$1,000 до \$9,999 (3) от \$1 до \$999
319.	Могут ли отличаться цены на инструменты Data Mining для различных категорий пользователей?	(1) да (2) нет (3) только в виде исключения
320.	Для использования технологии Data Mining ...	(1) возможна только покупка готового программного обеспечения, собственная разработка практически невозможна (2) возможна как покупка готового программного обеспечения, так и разработка собственными силами (3) наиболее правильный и выгодный вариант – использование программного обеспечения, адаптированного под конкретный бизнес
321.	Инструмент SPSS относится к категории:	(1) бесплатного программного обеспечения (2) достаточно недорогого программного обеспечения (3) достаточно дорогого программного обеспечения
322.	Рынок Business Intelligence, в том числе рынок инструментов Data Mining, ...	(1) насколько широк и разнообразен, что любая компания может выбрать для себя инструмент, который подойдет ей



		<p>по функциональности и по возможностям бюджета (2) является узким, и малое число компаний может позволить себе выбрать инструмент по требующейся ей функциональности (3) является насколько узким на сегодняшний день, что лишь большие компании могут позволить себе пользоваться инструментами по требующейся им функциональности</p>
323.	Инструменты Data Mining могут решать ...	<p>(1) только одну задачу Data Mining (2) несколько задач Data Mining (3) все задачи Data Mining (4) это зависит от конкретного инструмента</p>
324.	Инструмент Weka относится к категории:	<p>(1) бесплатного программного обеспечения (2) достаточно недорогого программного обеспечения (3) достаточно дорогого программного обеспечения.</p>
325.	Охарактеризуйте рынок программного обеспечения Data Mining:	<p>(1) представлен множеством инструментов (2) представлен достаточно небольшим количеством инструментов (3) на нем идет постоянная конкурентная борьба за потребителя (4) на нем практически нет конкуренции (5) он постоянно развивается (6) он уже достиг достаточного уровня развития и в ближайшее время предвидится спад</p>
326.	Постановка задачи, построение оптимальной модели, понимание модели, применение результатов. Перечисленные выше этапы являются этапами:	<p>(1) традиционного процесса Data Mining (2) подхода KXEN (3) и того, и другого (4) ни того, ни другого</p>
327.	Подготовка данных в KXEN включает следующие этапы :	<p>(1) преобразование данных (2) оптимальное кодирование указанных атрибутов для их наилучшего анализа в рамках выбранных алгоритмов (3) разделение атрибутов на символьные и числовые</p>
328.	Какой компонент KXEN используется в случаях, когда "сырые" данные содержат одновременно статическую информацию (например, возраст, пол или профессия индивида) и динамические переменные (например, шаблоны покупок или транзакции по кредитной карте)?	<p>(1) компонент Агрегирования Событий (KXEN Event Log – KEL) (2) компонент Согласованного Кодирования (KXEN Consistent Coder – K2C) (3) компонент Интеллектуальной Сегментации (KXEN Smart Segmenter – K2S)</p>
329.	Укажите, какие из перечисленных этапов являются этапами подхода KXEN к анализу данных:	<p>(1) постановка задачи (2) построение оптимальной модели (3) понимание модели (4) применение результатов</p>
330.	Существует ли необходимость временного или постоянного копирования данных для анализа в системе KXEN?	<p>(1) да (2) нет (3) по запросу</p>
331.	Какой компонент позволяет	<p>(1) компонент Согласованного Кодирования (KXEN</p>



	автоматически подготовить данные и трансформировать их в формат, подходящий для использования аналитическими приложениями KXEN?	Consistent Coder – K2C) (2) компонент Агрегирования Событий (KXEN Event Log – KEL) (3) компонент Интеллектуальной Сегментации (KXEN Smart Segmenter – K2S)
332.	Выберите пропущенный этап аналитического процесса KXEN: Постановка задачи...Понимание моделиПрименение результатов	(1) построение и тестирование модели (2) построение оптимальной модели (3) построение модели (4) тестирование модели
333.	Рейнжиниринг аналитического процесса KXEN ...	(1) позволяет автоматизировать процесс построения моделей (2) позволяет увеличить скорость проводимого анализа (3) позволяет заменить аналитика
334.	Какой компонент KXEN позволяет выявить естественные группы (кластеры) в наборе данных?	(1) компонент Интеллектуальной Сегментации (KXEN Smart Segmenter – K2S) (2) компонент Согласованного Кодирования (KXEN Consistent Coder – K2C) (3) компонент Агрегирования Событий (KXEN Event Log – KEL)
335.	Какие задачи позволяет решать инструмент KXEN?	(1) задачи регрессии и классификации (2) задачи кластеризации (3) анализ временных рядов. (4) поиск ассоциативных правил (5) все ответы верны
336.	В чем заключается основная особенность инструмента KXEN?	(1) в практически полной автоматизации процесса построения моделей (2) в возможности использования малого количества ретроспективных данных (3) в сложности построенных моделей (4) все ответы верны
337.	Охарактеризуйте квалификацию, которой требуется обладать пользователю для работы с KXEN	(1) пользователю не требуется обладать специальной квалификацией и знаниями в области анализа и статистики (2) пользователю требуется обладать специальной квалификацией и знаниями в области анализа и статистики (3) пользователю требуется обладать специальной квалификацией и знаниями в области анализа, статистики и искусственного интеллекта
338.	На решение каких трудностей направлен усовершенствованный аналитический процесс KXEN?	(1) трудоемкость подготовки данных (2) сложность выбора переменных, включенных в модель (3) требования к квалификации аналитиков (4) сложность интерпретации полученных результатов (5) сложность построения моделей (6) все ответы верны
339.	Построение модели в KXEN можно охарактеризовать как ...	(1) функцию предсказательного анализа в режиме on-line в формате "вопрос-ответ" (2) итеративный процесс



		(3) набор таких шагов: подготовка модели, построение модели, тестирование модели
340.	Что требуется от пользователя при работе с KXEN?	(1) данные, которые необходимо проанализировать (2) определение типа задачи, которую нужно решить (3) выбор лучшей модели (4) тестирование модели
341.	Data Mining-услуги могут предоставляться ...	(1) на определенных территориях (2) в определенных предметных областях (3) с использованием определенных методов
342.	Преимуществами использования готового программного обеспечения являются:	(1) готовые алгоритмы (2) техническая поддержка производителя (3) простота подготовки данных
343.	Постановка бизнес-задачи – это этап, который ...	(1) формулирует конкретные бизнес-задачи, и они уже не могут быть изменены (2) формулирует конкретные бизнес-задачи, и они не могут быть изменены в ходе прохождения именно этого цикла (3) формулирует конкретные бизнес-задачи, и они могут быть изменены в ходе прохождения именно этого цикла
344.	Data Mining консультирование может включать следующие услуги:	(1) образовательные услуги (2) услуги по разработке и внедрению программного обеспечения Data Mining (3) услуги по адаптации программного обеспечения Data Mining
345.	Слабыми сторонами использования готового программного обеспечения могут быть:	(1) высокая стоимость (2) необходимость наличия высококвалифицированных кадров (3) сложность подготовки данных (4) полная конфиденциальность информации
346.	На этапе первичного исследования данных ...	(1) со стороны заказчика может потребоваться лишь минимальное участие (2) со стороны заказчика может потребоваться максимальное участие (3) всю работу осуществляет заказчик
347.	Data Mining консультирование может включать следующие услуги:	(1) публикация отчетности Data Mining (2) проведение образовательных семинаров (3) консультации пользователей и разработчиков Data Mining
348.	Преимущества использования адаптированного программного обеспечения Data Mining по сравнению с готовыми программными продуктами и их самостоятельным использованием является:	(1) адаптированность (2) сложность подготовки данных (3) наличие терминов предметной области (4) полная конфиденциальность информации (5) не требуется дописывать программный код
349.	На этапе подготовки данных...	(1) специалисты компании Разработчика готовят данные для их дальнейшего анализа



		(2) специалисты компании Заказчика готовят данные для их дальнейшего анализа (3) специалисты компании Разработчика и Заказчика готовят данные для их дальнейшего анализа
350.	Вариант использования адаптированного программного обеспечения Data Mining ...	(1) имеет как сильные, так и слабые стороны (2) имеет неоспоримые преимущества перед использованием готового программного обеспечения (3) всегда проигрывает перед использованием готового программного обеспечения
351.	Какое решение в большей мере требует наличия высококвалифицированных специалистов при внедрении и использования инструмента Data Mining?	(1) использование готового программного обеспечения (2) заказ готового решения у фирмы-разработчика (3) адаптация программного обеспечения под конкретную задачу
352.	Гибкость инструмента Data Mining означает ...	(1) возможность выбора наиболее удобных понятий, в терминах которых должны быть сформулированы знания или термины предметной области (2) получение осмысленных и понятных знаний в естественной форме (3) оба ответа верны
353.	Готовые алгоритмы, полная конфиденциальность информации, техническая поддержка производителя, общение с другими пользователями пакета - это преимущества использования ...	(1) готового программного обеспечения (2) заказ готового решения у фирмы-разработчика (3) адаптация программного обеспечения под конкретную задачу (4) все ответы верны
354.	Достаточно высокая стоимость, невозможность добавлять свои функции, сложность подготовки данных, практическое отсутствие в интерфейсе терминов предметной области – это слабые стороны ...	(1) готового программного обеспечения (2) заказ готового решения у фирмы-разработчика (3) адаптация программного обеспечения под конкретную задачу (4) все ответы верны
355.	Какую часть мирового рынка Data Mining занимают услуги или консультации по эффективному внедрению этой технологии для решения актуальных бизнес-задач?	(1) менее 10% рынка (2) около половины рынка (3) более 75% рынка



## 4. Порядок проведения и критерии оценивания промежуточной аттестации

### 4.1. Порядок проведения промежуточной аттестации

Зачёт проводится в виде тестирования. Студент должен ответить на вопросы закрытого типа, которые предполагают выбор вариантов ответа, а также на вопросы открытого типа, которые не предполагают вариантов ответа, правильный ответ требуется написать самостоятельно. Всего 40 тестовых вопросов. Продолжительность теста – 60 минут.

### 4.2. Критерии оценивания промежуточной аттестации по видам оценочных средств

Тест формируется в системе электронного обучения MOODLE.

Максимальный балл за тест — 100 баллов.

Оценка	Отлично/ Зачтено	Хорошо/ зачтено	Удовлетворитель но/зачтено	Неудовлетворительно/ Не зачтено
Баллы	100-90 баллов	89-76 баллов	75-60 баллов	59-0 баллов
Уровень освоения проверяемых компетенций	высокий	средний	базовый	недостаточный

### 4.3. Результаты промежуточной аттестации и уровни сформированности компетенций

Для получения «зачтено» в конце семестра необходимо пройти итоговый тест на 60 и более баллов и сдать на оценку не ниже удовлетворительно первую практическую работу

Особенности проведения процедуры оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья обозначены в рабочей программе дисциплины (модуля).

Уровни сформированности компетенций определяется следующим образом:

1. **Высокий уровень сформированности компетенций** соответствует оценке отлично:
  - предполагает формирование компетенций на высоком уровне;
  - знание теоретических разделов изучаемой дисциплины на уровне не ниже оценки отлично;
  - студент умеет применять на практике знания, полученные в рамках изучения дисциплины
  - формируются навыки использования теоретических и практических разделов дисциплины для решения задач профессиональной деятельности;
2. **Средний уровень** соответствует оценке хорошо:
  - предполагает формирование компетенций на среднем уровне;



- знание теоретических разделов изучаемой дисциплины на уровне не ниже оценки хорошо;
  - студент умеет применять знания, полученные в рамках изучения дисциплины, для решения задач профессиональной деятельности;
3. Базовый уровень соответствует оценке удовлетворительно:
- предполагает формирование компетенций на базовом уровне;
  - знание теоретических разделов изучаемой дисциплины на уровне не ниже оценки удовлетворительно;
4. Недостаточный уровень соответствует оценке неудовлетворительно.