

МИНОБРНАУКИ РОССИИ  
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ  
ВЫСШЕГО ОБРАЗОВАНИЯ  
«ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»  
(ФГБОУ ВО «ВГУ»)

**УТВЕРЖДАЮ**

Заведующий кафедрой  
информационных систем



Борисов Д. Н.  
10.04.2024 г.

**РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ**  
Б1.В.ДВ.03.01 Data mining в бизнес-журналистике

- 1. Код и наименование специальности:** 42.04.02. Журналистика
- 2. Профиль:** Бизнес-журналистика и корпоративные медиакоммуникации
- 3. Квалификация выпускника:** журналист
- 4. Форма обучения:** заочная
- 5. Кафедра, отвечающая за реализацию дисциплины:** информационных систем
- 6. Составители программы:** Сычев Александр Васильевич, кандидат физико-математических наук, доцент
- 7. Рекомендована:** НМС факультета компьютерных наук, протокол №5 от 05.03.2024 г.
- 8. Учебный год:** 2024-2025 **Курс(ы):** 1

## 9. Цели и задачи учебной дисциплины

Целями освоения учебной дисциплины являются:

практическое знакомство студентов с современными технологиями анализа многомерных данных, включая модели, алгоритмы и программные средства, используемые для решения основных задач анализа многомерных данных: классификации, кластеризации и других.

Задачи учебной дисциплины:

- изучение структуры систем поддержки принятия решений (СППР), многомерной модели данных OLAP;
- изучение концепции Data Mining и основных задач, реализующих эту концепцию: классификацию, кластеризацию, поиск ассоциативных правил и других;
- изучение структуры процесса Data Mining и связанных с ним стандартов;
- практическое изучение задач Data Mining и методов их решения с помощью программных пакетов RapidMiner и других.

## 10. Место учебной дисциплины в структуре ОПОП:

Учебная дисциплина относится к части, формируемой участниками образовательных отношений, блока Б1 Дисциплины (модули) учебного плана по направлению магистратуры 42.04.02 Журналистика (дисциплины по выбору).

## 11. Планируемые результаты обучения по дисциплине/модулю (знания, умения, навыки), соотнесенные с планируемыми результатами освоения образовательной программы (компетенциями) и индикаторами их достижения:

Код	Название компетенции	Код(ы)	Индикатор(ы)	Планируемые результаты обучения
УК-1	Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, выработать стратегию действий	УК-1.1	Определяет пробелы в информации, необходимой для решения проблемной ситуации	Знать: принципы системного подхода Уметь: определять пробелы в информации, необходимой для решения проблемной ситуации. Владеть: навыками критического анализа проблемных ситуаций на основе системного подхода.
ПК-1	Способен осуществлять авторскую деятельность любого характера и уровня сложности с учетом специфики разных типов СМИ и других медиа	ПК-1.3.	Анализирует релевантную информацию из доступных документальных источников	Знать: доступные документальные источники Уметь: анализировать релевантную информацию из доступных документальных источников.
		ПК-1.4.	Контролирует достоверность и полноту полученной информации, систематизирует факты и мнения	Уметь: контролировать достоверность и полноту полученной информации, систематизировать факты и мнения. Владеть: приемами и техникой систематизации фактов и мнений.

12. Объем дисциплины в зачетных единицах/час.:2/72.

Форма промежуточной аттестации: экзамен

### 13. Трудоемкость по видам учебной работы

Вид учебной работы	Трудоемкость	
	Всего	По семестрам
		Курс 1 сессия 2
Аудиторные занятия	14	14
в том числе: практические	14	14
Самостоятельная работа	58	58
Форма промежуточной аттестации – Зачет	72	72

#### 13.1. Содержание дисциплины

п/п	Наименование раздела дисциплины	Содержание раздела дисциплины	Реализация раздела дисциплины с помощью онлайн-курса, ЭУМК *
<b>1. Практические занятия</b>			
1.1	Системы поддержки принятия решений и хранилища данных	Системы поддержки принятия решений (СППР). Классы задач анализа данных в СППР. Обобщенная архитектура СППР. OLTP и СППР - сравнение. Понятие об интеллектуальном анализе данных и Data Mining. Концепция хранилища данных (ХД). Свойства ХД. Архитектура систем ХД. Структура СППР с физическим ХД. Структура СППР с виртуальным ХД. Витрина данных (ВД). Структура СППР с самостоятельными ВД. Архитектура ХД.	
1.2	OLAP-системы	Многомерная модель данных. Основные элементы OLAP. Представление данных в виде гиперкуба. Операции над гиперкубом. Определение OLAP-систем. Архитектура OLAP-систем: MOLAP, ROLAP, HOLAP.	
1.3	Задачи Data Mining	Правила классификации. Деревья решений. Конструирование дерева решений. Критерий расщепления. Алгоритмы: байесовский, CART, C4.5. Алгоритмы классификации: метод "ближайшего соседа", метод построения математических функций, метод опорных векторов (SVM). Постановка задачи. Методы прогнозирования временных рядов. Методы поиска ассоциативных	

		правил. Алгоритм Apriori и его разновидности. Понятие кластерного анализа. Характеристики кластеров. Методы кластерного анализа: иерархические и неиерархические. Иерархические методы кластеризации. Агломеративные и дивизимные методы. Дендрограммы. Метрики подобия объектов. Проверка качества кластеризации. Достоинства и недостатки иерархических алгоритмов. Алгоритм k-средних. Достоинства и недостатки алгоритма. Проверка качества кластеризации. Этапы кластерного анализа. Сложности и проблемы кластерного анализа. Сравнение иерархических и неиерархических методов кластеризации. Этапы визуального анализа данных. Характеристики средств визуализации данных. Типы методов визуализации. Визуализация Data Mining моделей. Методы визуализации. Параллельные координаты. "Лица Чернова". Рекомендации по использованию средств визуализации. Качество визуализации. Основные тенденции в области визуализации.	
--	--	---	--

### 13.2. Темы (разделы) дисциплины и виды занятий

№ п/п	Наименование темы (раздела) дисциплины	Виды занятий (количество часов)			
		Лекции	Практические	Самостоятельная работа	Всего
1	Системы поддержки принятия решений и хранилища данных	-	2	10	12
2	OLAP-системы	-	2	10	12
3	Задачи Data Mining	-	10	38	48
Итого:		-	14	58	72

### 14. Методические указания для обучающихся по освоению дисциплины

1) При изучении дисциплины рекомендуется использовать следующие средства:

- рекомендуемую основную и дополнительную литературу;
- методические указания и пособия;
- контрольные задания для закрепления теоретического материала;
- электронные версии учебников и методических указаний для выполнения лабораторно - практических работ.

2) Для лучшего усвоения дисциплины рекомендуется проведение письменного опроса(тестирование, решение задач) студентов по материалам лекций и практических работ. Подборка вопросов для тестирования осуществляется на основе изученного теоретического материала.

Такой подход позволяет повысить мотивацию студентов при конспектировании лекционного материала.

3) При проведении лабораторных занятий обеспечивается практическая демонстрация материалов лекционных занятий и осуществляется экспериментальная проверка методов, алгоритмов и технологий обработки многомерных данных.

4) При переходе на дистанционный режим обучения для создания электронных курсов, проведения практических занятий используются информационные ресурсы образовательного портала "Электронный университет ВГУ (<https://edu.vsu.ru>), базирующегося на системе дистанционного обучения Moodle, развернутой в университете

## 15. Перечень основной и дополнительной литературы, ресурсов интернет, необходимых для освоения дисциплины

а) основная литература:

№ п/п	Источник
1	Чубукова, И. А. Data Mining : учебное пособие / И. А. Чубукова. — 2-е изд. — Москва : ИНТУИТ, 2016. — 470 с. — ISBN 978-5-94774-819-2. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/book/100582">https://e.lanbook.com/book/100582</a>
2	Макшанов, А. В. Современные технологии интеллектуального анализа данных : учебное пособие для спо / А. В. Макшанов, А. Е. Журавлев, Л. Н. Тындыкарь. — Санкт-Петербург : Лань, 2020. — 228 с. — ISBN 978-5-8114-5451-8. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/book/149343">https://e.lanbook.com/book/149343</a>
3	Макшанов, А. В. Системы поддержки принятия решений : учебное пособие / А. В. Макшанов, А. Е. Журавлев, Л. Н. Тындыкарь. — Санкт-Петербург : Лань, 2020. — 108 с. — ISBN 978-5-8114-5492-1. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/book/147094">https://e.lanbook.com/book/147094</a>

б) дополнительная литература:

№ п/п	Источник
4	Боровиков, В.П. Популярное введение в современный анализ данных в системе STATISTICA. Методология и технология современного анализа данных / В.П. Боровиков .— Москва : Горячая линия - Телеком, 2013 .— 288 с.
5	Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP : учебное пособие / А.А. Барсегян [и др.] .— 2-е изд. — СПб : БХВ-Петербург, 2007 .— 375 с.
6	Юре, Л. Анализ больших наборов данных / Л. Юре, Р. Ананд, Д. У. Джеффри ; перевод с английского А. А. Слинкин. — Москва : ДМК Пресс, 2016. — 498 с. — ISBN 978-5-97060-190-7. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/book/93571">https://e.lanbook.com/book/93571</a>

в) информационные электронно-образовательные ресурсы (официальные ресурсы интернет):

№ п/п	Ресурс
1.	Электронная библиотека ЗНБ ВГУ. - <a href="https://lib.vsu.ru">https://lib.vsu.ru</a>
2.	ЭБС "Университетская библиотека online". - <a href="http://biblioclub.ru/">http://biblioclub.ru/</a>
3.	ЭБС "Консультант студента". - <a href="http://www.studmedlib.ru">http://www.studmedlib.ru</a>
4.	Ресурсы «Электронного университета». – <a href="http://edu.vsu.ru">http://edu.vsu.ru</a>

## 16. Перечень учебно-методического обеспечения для самостоятельной работы

№ п/п	Источник
1	Электронный курс на образовательном портале «Электронный университет ВГУ».– ( <a href="https://edu.vsu.ru/">https://edu.vsu.ru/</a> )

## 17. Образовательные технологии, используемые при реализации учебной дисциплины, включая дистанционные образовательные технологии (ДОТ), электронное обучение (ЭО), смешанное обучение):

При реализации дисциплины используются практические занятия (проблемные, дискуссионные и т. д.), применяются дистанционные образовательные технологии в части освоения учебного материала, проведения текущей аттестации, самостоятельной работы по дисциплине или отдельным ее разделам и т. д.

## 18. Материально-техническое обеспечение дисциплины:

Аудитории для проведения занятий лекционного типа. Типовое оснащение, оборудование: мультимедиапроектор View Sonic; ПК (i5/4Gb/HDD 1Tb); экран настенный с электроприводом CS 244\*244; акустическая система BEHRINGER B115D, микшер UB 1204 FX, микрофон B-1. Программное обеспечение: WinPro 8 RUS Upgrd OLP NL Acdm; OfficeSTd 2013 RUS OLP NL Acdmc; Неисключительные права на ПО Dr. Web Enterprise Security Suite Комплексная защита Dr. Web Desktop Security Suite; СПС «ГАРАНТ-Образование».

Аудитории для проведения занятий семинарского типа, текущего контроля и промежуточной аттестации. Типовое оснащение, оборудование: мультимедиапроектор BenQ, экран настенный CS 244\*244; переносной ноутбук 15\*Packard Bell. Программное обеспечение: WinPro 8 RUS Upgrd OLP NL Acdm; OfficeSTd 2013 RUS OLP NL Acdmc; Неисключительные права на ПО Dr. Web Enterprise Security Suite Комплексная защита Dr. Web Desktop Security Suite; СПС «ГАРАНТ- Образование».

Аудитории для самостоятельной работы студентов. Используются компьютерные классы: ауд. 115 ([Воронеж, ул. Хользунова, 40-а](#)). Типовое оснащение, оборудование: мультимедиапроектор BenQ MX511; экран настенный CS 244\*244; интерактивная доска Promethean, ПК (i5/4Gb/HDD 1Tb) (11 шт.);

ауд. 126 ([Воронеж, ул. Хользунова, 40-а](#)). Типовое оснащение, оборудование: мультимедиапроектор BenQ MX511; ПК (Razer 5/4Gb/1Tb) (10 шт.); экран настенный CS 244\*244, интерактивная доска Promethean.

Программное обеспечение: WinPro 8 RUS Upgrd OLP NL Acdm; OfficeSTd 2013 RUS OLP NL Acdmc; Неисключительные права на ПО Dr. Web Enterprise Security Suite Комплексная защита Dr. Web Desktop Security Suite; СПС «ГАРАНТ- Образование».

## 19. Оценочные средства для проведения текущей и промежуточной аттестаций

Порядок оценки освоения обучающимися учебного материала определяется содержанием следующих разделов дисциплины:

№ п/п	Наименование раздела дисциплины (модуля)	Компетенция(и)	Индикатор(ы) достижения компетенции	Оценочные средства
1.	Системы поддержки принятия решений	УК-1	УК-1.1	КИМы

№ п/п	Наименование раздела дисциплины (модуля)	Компетенция(и)	Индикатор(ы) достижения компетенции	Оценочные средства
	и хранилища данных			
2.	OLAP-системы	ПК-1	ПК-1.3, ПК-1.4	КИМы, практические задания
3.	Задачи Data Mining	ПК-1	ПК-1.3, ПК-1.4	КИМы, практические задание
Промежуточная аттестация форма контроля: зачет				Перечень вопросов Практическое задание

## 20 Типовые оценочные средства и методические материалы, определяющие процедуры оценивания

### 20.1 Текущий контроль успеваемости

Тестовые задания - 1 балл за каждый правильный тест (максимум).

*Компетенция УК-1.*

1. Выберите правильное соответствие для задач анализа данных в СППР:

- a) СППР осуществляет поиск необходимых данных. Выполняются заранее определенные запросы
- b) СППР производит группирование и обобщение данных в любом виде, необходимом аналитику
- c) СППР осуществляет поиск функциональных и логических закономерностей в накопленных данных, построение моделей и правил, которые объясняют найденные закономерности и/или прогнозируют развитие некоторых процессов

- a) Оперативно-аналитический
- b) Информационно-поисковый
- c) Интеллектуальный

**Правильное соответствие:**

a	b
b	a
c	c

2. Выберите правильное соответствие для свойств знаний:

- a) Знания должны быть "разложены по полочкам".
- b) Для человека - это способность быстро понять и запомнить или, наоборот, вспомнить; для компьютерных знаний - средства доступа к знаниям
- c) Позволяет быстро осваивать и перерабатывать знания и повышает "коэффициент полезного содержания".
- d) Знания нужны для того, чтобы их использовать. Одно из главных свойств знаний - возможность их передачи другим и способность делать выводы на их основе.

- a) Удобство доступа и усвоения
- b) Структурированность
- c) Процедуры обработки
- d) Лаконичность

**Правильное соответствие:**

a	b
b	a
c	d
d	c

*Компетенция ПК-1*

3. Выберите правильное соответствие для задач Data Mining:

- a) Состоит в определении класса объекта по его характеристикам. Множество классов известно заранее.
  - b) Позволяет определить по известным характеристикам объекта значение некоторого его параметра. Значением параметра в отличие от задачи классификации является не конечное множество классов, а множество действительных чисел.
  - c) Сводится к нахождению частных зависимостей между объектами или событиями. Зависимости представляются в виде правил.
  - d) Заключается в поиске независимых групп и их характеристик во всем множестве анализируемых данных.
  - e) На основе особенностей исторических данных оцениваются пропущенные или же будущие значения целевых численных показателей.
  - f) Предсказание непрерывных значений признака объекта.
  - g) Создается графический образ анализируемых данных.
- 
- a) Поиск ассоциативных правил
  - b) Задачи прогнозирования
  - c) Задача регрессии
  - d) Задача оценивания
  - e) Задача кластеризации
  - f) Задача классификации
  - g) Задача визуализации

**Правильное соответствие:**

a	f
b	c
c	a



d	e
e	b
f	d
g	g

4. Выберите правильное соответствие для методов и задач Data Mining:

- a) Метод К-средних.
- b) Метод опорных векторов.
- c) Метод "ближайших соседей".
- d) Метод деревьев решений.
- e) Метод Apriori.
- d) "Лица Чернова".

- a) Поиск ассоциативных правил
- b) Задача кластеризации
- c) Задача классификации
- d) Задача классификации
- e) Задача классификации
- f) Задача визуализации

**Правильное соответствие:**

a	b
b	e
c	d
d	c
e	a
f	f

5. Выберите правильное соответствие для средств визуализации данных:

- a) Одномерные массивы, временные ряды
- b) Точки двумерных графиков, географические координаты
- c) Финансовые показатели, результаты экспериментов
- d) Газетные статьи, web-документы
- e) Структуры подчиненности в организации, электронная переписка, гиперссылки документов
- f) Информационные потоки, отладочные операции

- a) Тексты и гипертексты
- b) Одномерные данные
- c) Многомерные данные
- d) Иерархические и связанные данные
- e) Двумерные данные
- f) Алгоритмы и программы

**Правильное соответствие:**

a	b
b	e
c	c
d	a
e	d
f	f

6. Выберите правильные утверждения, относящиеся к временным рядам и их прогнозированию:

- a) Неслучайная функцию, которая формируется под действием общих или долговременных тенденций, влияющих на временной ряд.
- b) Частота, с которой делается новый прогноз.
- c) Число периодов в будущем, которые покрывает прогноз.
- d) Периодически повторяющаяся компонента временного ряда. Через примерно равные промежутки времени форма кривой, которая описывает поведение зависимой переменной, повторяет свои характерные очертания.
- e) Основная единица времени, на которую делается прогноз.

- a) Тренд
- b) Сезонная составляющая временного ряда
- c) Период прогнозирования
- d) Интервал прогнозирования
- e) Горизонт прогнозирования

**Правильное соответствие:**

a	a
b	d
c	e
d	b
e	c

7. Выберите правильное соответствие:

- a) Множество событий (операций), которые произошли одновременно.
- b) Вероятность того, что из события А следует событие В.
- c) Набор, у которого поддержка элементов выше определенного пользователем минимального значения поддержки
- d) Процент транзакций из всего набора, которые содержат одновременно наборы элементов А и В.

- a) Часто встречающийся набор 3
- b) Транзакция 1

- c) Поддержка правила 4
- d) Достоверность 2

**Правильное соответствие:**

a	b
b	d
c	a
d	c

8. Выберите правильное соответствие для характеристик кластеров:

- a) Среднее геометрическое место точек в пространстве переменных.
  - b) Максимальное расстояние точек от центра кластера.
  - c) Объект, который по мере сходства может быть отнесен к нескольким кластерам.
- 
- a) Радиус кластера
  - b) Спорный объект
  - c) Центр кластера

**Правильное соответствие:**

a	c
b	a
c	b

9. Выберите правильные утверждения, относящиеся к задаче классификации:

- a) Может быть многомерной (по двум и более признакам)
- b) Может быть одномерной (по одному признаку)
- c) Может быть только трехмерной (по трем признакам)
- d) Набор исходных данных (или выборку данных) разбивают на два множества: обучающее и тестовое
- e) Относится к стратегии обучения с учителем
- f) Решение задачи состоит из двух этапов: конструирования и модели и ее использования
- g) Целью задачи является разбиение совокупности объектов на однородные группы.

**Правильные ответы: a,b,d,e,f**

10. Выберите утверждения, не относящиеся к задаче кластеризации:

- a) В данной задаче классы изучаемого набора данных заранее не predetermined.
- b) Набор исходных данных (или выборку данных) разбивают на два множества: обучающее и тестовое.
- c) Относится к стратегии обучения с учителем.

- d) Решение задачи состоит из двух этапов: конструирования и модели и ее использования.
- e) Синонимами для названия задачи являются "обучение без учителя" и "таксономия".
- f) Целью задачи является разбиение совокупности объектов на однородные группы.

**Правильные ответы: *b,c,d***

11. Выберите правильные утверждения, относящиеся к задаче прогнозирования:

- a) В данной задаче предсказывается класс зависимой переменной.
- b) В данной задаче предсказываются числовые значения зависимой переменной, пропущенные или неизвестные (относящиеся к будущему).
- c) Задача заключается в установлении функциональной зависимости между зависимыми и независимыми переменными.
- d) Основой для решения задачи является историческая информация, хранящаяся в базе данных в виде временных рядов.
- e) Решение задачи направлено на определение тенденций динамики конкретного объекта или события на основе ретроспективных данных, т.е. анализа его состояния в прошлом и настоящем.

**Правильные ответы: *b,c,d,e***

12. Выберите правильные утверждения, относящиеся к методу опорных векторов:

- a) Для классификации используется не все множество образцов, а лишь их небольшая часть, которая находится на границах.
- b) При помощи данного метода не могут решаться задачи бинарной классификации.
- c) Метод сильно подвержен проблеме переобучения.
- d) Метод работает с любым количеством измерений.
- e) Метод относится к группе граничных методов. Он определяет классы при помощи границ областей.
- f) Для метода характерны устойчивые решения, нет проблем с локальными минимумами.

**Правильные ответы: *a,d,e,f***

13. Выберите, какие модели используются для представления полученных знаний в Data Mining:

- a) Файлы
- b) Тексты
- c) Таблицы
- d) Правила
- e) Математические функции

- f) Кластеры
- g) Инструкции
- h) Деревья решений

**Правильные ответы: c,d,e,f,g,h**

14. Укажите, какие из приведенных ниже задач относятся к основным задачам Data Mining?

- a) Управление
- b) Тестирование
- c) Регрессия
- d) Проектирование
- e) Программирование
- f) Прогнозирование и оценивание
- g) Поиск ассоциативных правил
- h) Контроль
- i) Кластеризация
- j) Визуализация

**Правильные ответы: c,f,g,i,j**

15. Опишите правильную последовательность действий для метода "ближайшего соседа"

- a) Сопоставление имеющейся информации о задаче с деталями прецедентов, хранящихся в базе, для выявления аналогичных случаев
- b) Сбор подробной информации о поставленной задаче
- c) Проверка корректности каждого вновь полученного решения
- d) Занесение детальной информации о новом прецеденте в базу прецедентов
- e) Выбор прецедента, наиболее близкого к текущей проблеме, из базы прецедентов
- f) Адаптация выбранного решения к текущей проблеме, если это необходимо

**Правильные ответы: b,a,e,f,c,d**

**Вопросы с кратким (вычисляемым) ответом - 1 балл за каждый правильный тест (максимум)**

*Компетенция ПК-1.*

1. База данных содержит 8 транзакций на основе элементов {A, B, C, D, E, F}:

<i>tid</i>	<i>items</i>
1	ABC
2	BCD
3	CDE

4	BC
5	CD
6	ABCD
7	ABD
8	EF

укажите значение поддержки для ассоциативного правила A -> C.

**Правильный ответ: 2**

2. База данных содержит 8 транзакций на основе элементов {A, B, C, D, E, F}:

<i>tid</i>	<i>items</i>
1	ABC
2	BCD
3	CDE
4	BC
5	CD
6	ABCD
7	ABD
8	EF

укажите значение поддержки для ассоциативного правила A -> B.

**Правильный ответ: 3**

3. База данных содержит 8 транзакций на основе элементов {A, B, C, D, E, F}:

<i>tid</i>	<i>items</i>
1	ABC
2	BCD
3	CDE
4	BC
5	CD
6	ABCD
7	ABD
8	EF

При минимальной поддержке, равной 2, сколько будет найдено часто встречающихся 3-элементных наборов элементов?

**Правильный ответ: 3**

4. Набор данных включает атрибут "возраст". Значения атрибута приведены в порядке возрастания: 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70.

Рассчитайте значения медианы ***M*** и среднего ***A*** (с округлением до целого значения) для приведенного массива значений. Рассчитанные значения введите, разделяя пробелом: ***M A***.

**Правильный ответ: 25 30**

5. Кластер содержит 3 объекта:  $A_1(2, 10)$ ;  $A_2(2, 5)$ ;  $A_3(8, 4)$ .

Рассчитайте координаты центроида для данного кластера.

Ответ необходимо привести в виде ***C(x,y)***. Значения ***x*** и ***y*** должны быть округлены до целых значений.

**Правильный ответ: C(4,6)**

6. Кластер содержит 4 объекта:  $A_1(3,8)$ ;  $A_2(4,5)$ ;  $A_3(8,5)$ ;  $A_4(5,2)$ .

Рассчитайте координаты центроида для данного кластера.

Ответ необходимо привести в виде ***C(x,y)***. Значения ***x*** и ***y*** должны быть округлены до целых значений.

**Правильный ответ: C(4,6)**

**Вопросы с развернутым ответом**

*Компетенция ПК-1.*

1. Опишите алгоритм поиска ассоциативных правил Apriori. Какие характеристики используются для оценки полученных правил?
2. Опишите алгоритм кластеризации K-средних (по шагам). Каким образом определяется ожидаемое количество кластеров?
3. Какие инструменты используются для визуализации многомерных (4+) данных? Кратко охарактеризуйте наиболее известные из них.
4. Опишите кратко основные этапы процесса решения задачи кластеризации.
5. Кратко опишите основные этапы классификации на основе дерева решений. Чем сокращение дерева полезно при построении дерева решений?
6. Опишите метод "ближайшего соседа" для решения задачи классификации. Что представляет собой модель классификации в рамках данного метода?
7. Опишите, в чем заключается проблема переобучения? Какие подходы используются для устранения данной проблемы?
8. Охарактеризуйте кратко этапы процесса Data Mining. Какие из этапов являются наиболее трудоемкими? Объясните почему?

Критерии оценивания развернутого ответа:

В зависимости от степени полноты ответа на вопрос максимально можно получить за развернутый ответ на вопрос 3 балла.

## 20.2 Промежуточная аттестация

Контрольно-измерительные материалы промежуточной аттестации включают в себя вопросы, позволяющие оценить уровень полученных знаний и практические задания, позволяющие оценить степень сформированности умений и навыков

Примерный перечень теоретических вопросов к зачету.

<p>1. Системы поддержки принятия решений (СППР). Классы задач анализа данных в СППР. Обобщенная архитектура СППР. OLTP и СППР - сравнение. Понятие об интеллектуальном анализе данных и Data Mining. Концепция хранилища данных (ХД). Свойства ХД. Архитектура систем ХД. Структура СППР с физическим ХД. Структура СППР с виртуальным ХД. Витрина данных (ВД). Структура СППР с самостоятельными ВД. Архитектура ХД.</p>
<p>2. Многомерная модель данных. Основные элементы OLAP. Представление данных в виде гиперкуба. Операции над гиперкубом. Определение OLAP-систем. Двенадцать правил Кодда для OLAP. Дополнительные правила Кодда. Тест FASMI. Архитектура OLAP-систем: MOLAP, ROLAP, HOLAP.</p>
<p>3. Правила классификации. Деревья решений. Конструирование дерева решений. Критерий расщепления. Алгоритмы: байесовский, CART, C4.5.</p>
<p>4. Алгоритмы классификации: метод "ближайшего соседа", метод построения математических функций, метод опорных векторов (SVM).</p>
<p>6. Постановка задачи. Методы прогнозирования временных рядов. Методы поиска ассоциативных правил. Алгоритм Apriori и его разновидности.</p>
<p>5. Понятие кластерного анализа. Характеристики кластеров. Методы кластерного</p>



анализа: иерархические и неиерархические. Иерархические методы кластеризации. 6. Агломеративные и дивизимные методы. Дендрограммы. Метрики подобия объектов. Проверка качества кластеризации. Достоинства и недостатки иерархических алгоритмов. Алгоритм k-средних. Достоинства и недостатки алгоритма. Проверка качества кластеризации. Этапы кластерного анализа. Сложности и проблемы кластерного анализа. Сравнение иерархических и неиерархических методов кластеризации. 7. Этапы визуального анализа данных. Характеристики средств визуализации данных. Типы методов визуализации. Визуализация Data Mining моделей. Методы визуализации. Параллельные координаты. "Лица Чернова". Рекомендации по использованию средств визуализации. Качество визуализации. Основные тенденции в области визуализации.

Для оценивания результатов обучения на зачете используется – зачтено, не зачтено по результатам текущей работы в семестре и тестирования. Соотношение показателей, критериев и шкалы оценивания результатов обучения представлено в следующей таблице.

Критерии оценивания компетенций	Уровень сформированности компетенций	Шкала оценок
Обучающийся владеет теоретическими основами методов анализа данных, способен иллюстрировать ответ примерами, фактами, данными научных исследований, применять теоретические знания для решения практических задач. Допустимы незначительные ошибки и неточности в ответах, которые исправляются после замечания преподавателя.	От базового до порогового уровня	Зачтено
Ответ на контрольно-измерительный материал фрагментарный. Обучающийся демонстрирует несистематические, отрывочные знания, допускает грубые, принципиальные ошибки, которые не устранены после дополнительных наводящих вопросов, что соответствует не освоению компетенций.	–	Незачтено