

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РОССИЙСКОЙ ФЕДЕРАЦИИ  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«ДАГЕСТАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»  
Факультет информатики и информационных технологий

## **РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ**

### **Методы анализа данных**

Кафедра информационных систем и технологий программирования

**Образовательная программа**  
09.03.04 Программная инженерия

**Направленность (профиль) программы**  
Разработка программно-информационных систем  
Программное обеспечение вычислительной техники и автоматизированных  
систем

**Уровень высшего образования**  
бакалавриат

Форма обучения  
**Очная**


**Статус дисциплины:**  
дисциплина по выбору

Махачкала, 2022


Рабочая программа дисциплины «Методы анализа данных» составлена в 2022 году в соответствии с требованиями ФГОС ВО – бакалавриат по направлению подготовки 09.03.04 Программная инженерия от «19» сентября 2017г. № 920.

Разработчик(и): кафедра информационных систем и технологий программирования, Магомедгаджиев Ш.М., к.э.н., доцент


Рабочая программа дисциплины одобрена:  
на заседании кафедры ИСиТП от «01» марта 2022г., протокол № 8

Зав. кафедрой  Исмиханов З.Н.  
(подпись)

на заседании Методической комиссии факультета ИиИТ  
от «17» марта 2022г., протокол № 7

Председатель  Бакмаев А.Ш.  
(подпись)

Рабочая программа дисциплины согласована с учебно-методическим управлением «31» марта 2022г.

Начальник УМУ  Гасангаджиева А.Г.  
(подпись)

## Аннотация рабочей программы дисциплины

Дисциплина «Методы анализа данных» входит в часть, формируемая участниками образовательных отношений и является дисциплиной по выбору образовательной программы бакалавриата, по направлению 09.03.04 Программная инженерия.

Дисциплина реализуется на факультете информатики и информационных технологий кафедрой информационных систем и технологий программирования.

Содержание дисциплины охватывает круг вопросов, связанных с изучением основных понятий и жизненного цикла анализа и обработки данных, обзором и применением методов статистического, компьютерного и интеллектуального анализа данных.

Дисциплина нацелена на формирование следующих компетенций выпускника: общепрофессиональных ОПК-7; профессиональных ПК-10.

Преподавание дисциплины предусматривает проведение следующих видов учебных занятий: лекции, лабораторные занятия, самостоятельная работа.

Рабочая программа дисциплины предусматривает проведение следующих видов контроля успеваемости: текущий контроль в форме опроса, тестов, контрольных работ и промежуточный контроль в форме экзамена.

Объем дисциплины 4 зачетные единицы, в том числе 144 в академических часах по видам учебных занятий

форма обучения - очная

Семестр	Учебные занятия					СРС, в том числе экзамен	Форма промежуточной аттестации
	в том числе:						
	всего	Контактная работа обучающихся с преподавателем			СРС, в том числе экзамен		
		всего	из них				
		Лекции	Лабораторные занятия	Практические занятия			
8	144	66	16	28		64	экзамен

## 1. Цели освоения дисциплины

Целями освоения дисциплины «Методы анализа данных» является освоение студентами современных технологий и методов обработки и анализа информации; формирование целостной системы знаний в области создания, накопления, обработки и использования информационных ресурсов; приобретение методологических основ и практических навыков анализа данных.

Задачи курса:

- получение представления об основных понятиях, методах, моделях анализа и обработки данных и наиболее перспективных прикладных сферах их применения;
- освоение математических, статистических и интеллектуальных методов анализа данных;
- освоение навыков работы с современными информационными технологиями анализа и обработки информации;
- получение опыта применения методов и способов построения эффективных алгоритмов и моделей для решения прикладных задач анализа и обработки данных;
- освоение основ участия в разработке и сопровождении информационных хранилищ, технологий оперативного и интеллектуального анализа данных, отражающих деятельность предприятий в различных предметных областях.

## 2. Место дисциплины в структуре ОПОП бакалавриата

Дисциплина «Методы анализа данных» входит в часть, формируемую участниками образовательных отношений и является дисциплиной по выбору образовательной программы по направлению 09.03.04 Программная инженерия.

При изучении дисциплины «Методы анализа данных» предполагается, что студент владеет дисциплинами «Методы анализа и обработки данных», «Программирование», «Моделирование» в объёме, предусмотренном ОПОП подготовки бакалавров.

Данную учебную дисциплину дополняет параллельное или последующее освоение следующих дисциплин: «Методы и алгоритмы обработки изображений», «Имитационное моделирование»

## 3. Компетенции обучающегося, формируемые в результате освоения дисциплины (перечень планируемых результатов обучения и процедура освоения).

Код и наименование компетенции из ОПОП	Код и наименование индикатора достижения компетенций (в соответствии с ОПОП)	Планируемые результаты обучения	Процедура освоения
ОПК-7. Способен применять в практической деятельности основные концепции, принципы, теории и факты, связанные с информатикой	ИПК-7.1. Знает основные языки программирования и работы с базами данных, операционные системы и оболочки, современные программные среды разработки информационных систем и технологий. ИПК-7.2. Умеет применять языки программирования и работы с базами данных, современные программные среды разработки информационных систем и технологий для автоматизации бизнес-процессов, решения	<i>Знает:</i> основные понятия, методы, модели анализа и обработки данных и наиболее перспективные прикладные сферы их применения. <i>Умеет:</i> работать с инструментальными средствами анализа данных и аналитическими библиотеками <i>Владеет:</i> практическими навыками визуализации полученных данных.	Опрос, тестирование, контрольная работа

	<p>прикладных задач различных классов, ведения баз данных и информационных хранилищ. ИПК-7.3. Имеет навыки программирования, отладки и тестирования прототипов программно-технических комплексов задач.</p>		
<p>ПК-10. Владение навыками использования различных технологий разработки программного обеспечения</p>	<p>ИПК-10.1. Знает современные технологии разработки программного обеспечения (структурное, Объектно-ориентированное) ИПК-10.2. Умеет использовать современные технологии разработки программного обеспечения ИПК-10.3. Владеет навыками использования современных технологий разработки программного обеспечения</p>	<p><i>Знает:</i> основные методы сбора и обработки данных <i>Умеет:</i> применять языки программирования для построения эффективных алгоритмов и моделей для решения прикладных задач анализа и обработки данных <i>Владеет:</i> практическими навыками использования языка программирования Python для обработки и анализа данных; навыками работы с разными форматами файлов с данными.</p>	<p>Опрос, тестирование, контрольная работа</p>

#### 4. Объем, структура и содержание дисциплины.

4.1. Объем дисциплины составляет 4 зачетные единицы, 144 академических часов.

4.2. Структура дисциплины.

4.2.1. Структура дисциплины в очной форме

№ п/п	Разделы и темы дисциплины	Семестр	Неделя семестра	Виды учебной работы, включая самостоятельную работу студентов и трудоемкость (в часах)				Самостоятельная работа	Формы текущего контроля успеваемости (по неделям семестра) Форма промежуточной аттестации (по семестрам)
				Лекции	Практические занятия	Лабораторные занятия	Контроль самост. раб.		
Модуль 1. Основы технологии анализа и обработки данных.									
1	Анализ данных, основные понятия	8	1	2	2	2		10	Опрос, тестирование, контрольная работа
2	Первичная обработка данных статистического наблюдения	8	2	2	2	6		10	Опрос, тестирование, контрольная работа
	<i>Итого по модулю 1:</i>			4	4	8		20	
Модуль 2. Компьютерные технологии анализа данных.									
1	Обзор компьютерных технологий анализа данных	8	3-4	2	4	6		6	Опрос, тестирование, контрольная работа
2	Аналитические информационные системы: состав, назначение, архитектура	8	5-6	4	2	6		6	Опрос, тестирование, контрольная работа
	<i>Итого по модулю 2:</i>			6	6	12		12	Зачет
Модуль 3. Основы машинного обучения									
1	Методы классификации и регрессии	8	7-8	2	4	6		6	Опрос, тестирование, контрольная работа
2	Методы кластеризации	8	9	4	2	6		6	Опрос, тестирование, контрольная работа
	<i>Итого по модулю 3:</i>			6	6	12		12	
Модуль 4. Подготовка к экзамену									
	<i>Итого по модулю 4:</i>							36	экзамен
	<b>ИТОГО:</b>			16		28		100	

4.3. Содержание дисциплины, структурированное по темам (разделам).

#### Модуль 1. Основы технологии анализа и обработки данных

Тема 1. Анализ данных, основные понятия

Понятие данных. Структурированные и неструктурированные данные. Наука о данных. Этапы построения систем анализа данных. Основные источники данных. Основные понятия статистики данных. Закон больших чисел. Метод исследования. Статистические показатели, закономерности. Этапы, формы, виды статистического исследования. Ошибки данных.

Тема 2. Первичная обработка данных статистического наблюдения

Абсолютные и относительные статистические величины, средние величины, вариационный анализ. Медиана, квартили, децили, мода. Шкалы измерения. Среднее линейного отклонения, дисперсия, среднее квадратическое отклонение. Функциональная, статистическая и корреляционная связи. Методы измерения тесноты парной корреляционной связи. Ранговые коэффициенты Спирмена и Кендалла.

Понятие об индексах, их классификация. Задачи индексного анализа данных. Общие индексы количественных показателей. Общие индексы качественных показателей. Двухфакторные системы индексов. Индекс Пааше и Ласпейреса. Индексы динамики средних показателей. Индексы переменного состава, постоянного состава, структурных сдвигов.

## **Модуль 2. Компьютерные технологии анализа данных.**

Тема 3. Обзор компьютерных технологий анализа данных

Особенности компьютерного анализа данных. Обзор современных информационных технологий и программных средств анализа и обработки данных. Статистические пакеты обработки данных, электронные таблицы, аналитические платформы. Сравнительные характеристики основных статистических пакетов STATISTICA, SPSS и др. Языки программирования для анализа данных R, Python. Системы анализа данных. Аналитическая платформа Deductor.

Форматирование наборов и предварительная обработка данных в Microsoft Excel. Визуализация данных. Описательная статистика. Сводные таблицы и сводные диаграммы в Microsoft Excel. Выбросы, пропущенные значения и их обработка в Microsoft Excel. Инструментарий в надстройке «Анализ данных». Решение задач оптимизации в Excel, анализ финансовых данных, решение логических задач и решение математических задач в Excel

Тема 4. Аналитические информационные системы: состав, назначение, архитектура

Состав информационно-аналитической системы. Типы инструментальных средств создания и поддержки АИС. Основные принципы построения архитектуры АИС. Средства сбора и доработки данных. Средства преобразования данных. Среды разработки средств сбора и доработки данных и информационных хранилищ. Структура информационных хранилищ. Архитектура средств оперативного (OLAP-анализа). Архитектура комплекса средств интеллектуального анализа (Data mining). Основные принципы построения информационных хранилищ.

## **Модуль 3. Основы машинного обучения**

Тема 5. Методы классификации и регрессии.

Понятие и направления использования машинного обучения. Типы систем машинного обучения. Обучение с учителем и обучение без учителя. Классы задач машинного обучения: регрессия, классификация, кластерный анализ, поиск аномалий. Примеры задач машинного обучения в экономике, управлении и финансах.

Постановка задачи регрессионного анализа. Парная линейная регрессия. Множественная линейная регрессия. Применения линейной регрессии. Кодирование

категориальных признаков. Прогнозирование на основе линейной регрессии. Градиентный спуск. Ускорение градиентного спуска. Измерение качества в задаче регрессии.

#### Тема 6. Задачи классификации и кластеризации

Постановка задачи классификации с обучением. Логистическая регрессия. Понятие о деревьях решений. Кредитный скоринг. Постановка задачи кластерного анализа. Метод К-средних. Сегментирование потребителей. Понятие о методах машинного обучения в задачах поиска аномалий

#### **4.3.2. Содержание лабораторно-практических занятий по дисциплине.**

##### **Лабораторные работы (лабораторный практикум)**

1. Лабораторная работа №1: Основы работы с аналитической платформой Deductor Studio
2. Лабораторная работа №2: Обработка данных. Основные алгоритмы по очистке и трансформации данных в Deductor Studio
3. Лабораторная работа №3: Проектирование и создание хранилища данных
4. Лабораторная работа №4: Наполнение хранилища данных
5. Лабораторная работа №5: Классификация с помощью дерева решений
6. Лабораторная работа №6: Кластеризация методом k-средних
7. Лабораторная работа №7: Поиск ассоциативных правил
8. Лабораторная работа №8: Основы анализа данных на Python
9. Лабораторная работа №9: Библиотека Sklearn и предобработка данных
10. Лабораторная работа №10: Оценка линейной регрессии
11. Лабораторная работа №11: Метод k-ближайших соседей
12. Лабораторная работа №12: Кластеризация с помощью k-средних

#### **5. Образовательные технологии**

Традиционные образовательные технологии - лекции в сочетании с практическими занятиями, семинарами и с лабораторными работами, самостоятельное изучение определенных разделов. Использование персональных компьютеров при выполнении лабораторных работ и сдаче итогового экзамена. Чтение лекций с использованием компьютера и проектора, проведение лабораторных работ в компьютерном классе.

При реализации учебной дисциплины используются электронные практикумы, электронные учебники, презентации средства диагностики и контроля, разработанные специалистами кафедры т.д.

Удельный вес занятий, проводимых в интерактивной форме, составляет 20% аудиторных занятий.

#### **6. Учебно-методическое обеспечение самостоятельной работы студентов.**

##### *Форма контроля и критерий оценок*

В соответствии с учебным планом предусмотрен экзамен в четвертом семестре.

Формы контроля: текущий контроль, промежуточный контроль по модулю, итоговый контроль по дисциплине предполагают следующее распределение баллов.

Текущий контроль

- Выполнение 1 домашней работы 10 баллов
- Активность в системе Moodle 10 баллов

Промежуточный контроль

Примерное распределение времени самостоятельной работы студентов

Вид самостоятельной работы	Примерная трудоёмкость,	Формируемые компетенции
----------------------------	-------------------------	-------------------------



	а.ч.	
	очная	
<b>Текущая СРС</b>		
работа с лекционным материалом, с учебной литературой	4	ПК-6
опережающая самостоятельная работа (изучение нового материала до его изложения на занятиях)	4	ПК-6
самостоятельное изучение разделов дисциплины	8	ПК-6
выполнение домашних заданий, домашних контрольных работ	6	ПК-6
подготовка к лабораторным работам, к практическим и семинарским занятиям	14	ПК-6
подготовка к контрольным работам, коллоквиумам, зачётам	4	ПК-6
<b>Творческая проблемно-ориентированная СРС</b>		
выполнение расчётно-графических работ	14	ПК-6
поиск, изучение и презентация информации по заданной проблеме, анализ научных публикаций по заданной теме	4	ПК-6
исследовательская работа, участие в конференциях, семинарах, олимпиадах	2	ПК-6
анализ данных по заданной теме, выполнение расчётов, составление схем и моделей на основе собранных данных	4	ПК-6
<b>Итого СРС:</b>	<b>64</b>	

Самостоятельная работа студентов (СРС) включает контролируемую и внеаудиторную самостоятельную работу, направлена на повышение качества обучения, углубление и закрепление знаний студента, развитие аналитических навыков по проблематике учебной дисциплины, активизацию учебно-познавательной деятельности студентов и снижение аудиторной нагрузки. Часть программного материала выносится для самостоятельного внеаудиторного изучения с последующим текущим или итоговым контролем знаний на занятиях или экзамене. Контроль СРС и оценка ее результатов организуется как самоконтроль (самооценка) студента, а также как контроль и оценка со стороны преподавателя, например в ходе собеседования. Баллы, полученные по СРС студентом, обязательно учитываются при итоговой аттестации по курсу. Формы контроля СРС включают: тестирование; устную беседу по теме с преподавателем; выполнение индивидуального задания и др.

Роль студента в СРС - самостоятельно организовывать свою учебную работу по предложенному преподавателем, методически обеспеченному плану. СРС по курсу учитывает индивидуальные особенности слушателей и включает не только задания, связанные с решением типовых задач, но также творческие задания, требующие самостоятельно «добывать» знания из разных областей, группировать и концентрировать их в контексте конкретной решаемой задачи. Технология обучения предусматривает выработку навыков презентации результатов выполненного индивидуального задания и создание условий для командной работы над комплексной темой с распределением функций и ответственности между членами коллектива. Оценка результатов выполнения индивидуального задания осуществляется по критериям, известным студентам, отражающим наиболее значимые аспекты контроля за выполнением этого вида работ.

<b>Разделы и темы для</b>	<b>Виды и содержание самостоятельной работы</b>
---------------------------	---

<b>самостоятельного изучения</b>	
Статистические методы анализа данных	<ul style="list-style-type: none"> <li>-конспектирование первоисточников и другой учебной литературы;</li> <li>-проработка учебного материала (по конспектам лекций учебной и научной литературе) и подготовка докладов на семинарах и практических занятиях, к участию в тематических дискуссиях;</li> <li>-поиск и обзор научных публикаций и электронных источников информации, подготовка заключения по обзору;</li> <li>-работа с тестами и вопросами для самопроверки;</li> </ul>
Большие данные, сущность, признаки, особенности обработки и инструментарий	<ul style="list-style-type: none"> <li>-конспектирование первоисточников и другой учебной литературы;</li> <li>-проработка учебного материала (по конспектам лекций учебной и научной литературе) и подготовка докладов на семинарах и практических занятиях, к участию в тематических дискуссиях;</li> <li>-поиск и обзор научных публикаций и электронных источников информации, подготовка заключения по обзору;</li> <li>-работа с тестами и вопросами для самопроверки;</li> <li>- решение домашних контрольных задач.</li> </ul>
Машинное обучение и искусственный интеллект	<ul style="list-style-type: none"> <li>-конспектирование первоисточников и другой учебной литературы;</li> <li>-проработка учебного материала (по конспектам лекций учебной и научной литературе) и подготовка докладов на семинарах и практических занятиях, к участию в тематических дискуссиях;</li> <li>-поиск и обзор научных публикаций и электронных источников информации, подготовка заключения по обзору;</li> <li>-работа с тестами и вопросами для самопроверки;</li> <li>- решение домашних контрольных задач.</li> </ul>
Инструментарий Excel и Google Таблицы для анализа данных.	Изучение документации программного обеспечения
Языки программирования для анализа данных R, Python	Изучение документации программного обеспечения
Изучение перспектив развития теории вычислительных процессов и структур	<ul style="list-style-type: none"> <li>-конспектирование первоисточников и другой учебной литературы;</li> <li>-проработка учебного материала (по конспектам лекций учебной и научной литературе) и подготовка докладов на семинарах и практических занятиях, к участию в тематических дискуссиях;</li> <li>-поиск и обзор научных публикаций и электронных источников информации, подготовка заключения по обзору;</li> <li>-работа с тестами и вопросами для самопроверки;</li> <li>- решение домашних контрольных задач.</li> </ul>

## **7. Фонд оценочных средств для проведения текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины.**

### **7.1. Типовые контрольные задания**

1. Структурированные и неструктурированные данные.

2. Этапы построения систем анализа данных.
3. Первичная обработка данных статистического наблюдения.
4. Методы измерения тесноты парной корреляционной связи.
5. Индексный метод исследования данных.
6. Обзор современных информационных технологий и программных средств анализа и обработки данных.
7. Языки программирования для анализа данных R, Python.
8. Системы анализа данных.
9. Аналитическая платформа Deductor.
10. Аналитические информационные системы: состав, назначение, архитектура.
11. Основные принципы построения информационных хранилищ.
12. Типы систем машинного обучения.
13. Задачи машинного обучения: классификация.
14. Задачи машинного обучения: кластеризация.
15. Задачи машинного обучения: поиск аномальных значений.
16. Задачи машинного обучения: регрессия.

### **Комплект тестовых заданий**

1. Аналитик это ...
  - а) специалист в области анализа и моделирование
  - б) специалист в предметной области;
  - в) человек, решающий определенные задачи;
  - г) человек, который имеет опыт в программировании.
  
2. Переменная X измерена в порядковой шкале. Результаты измерений этой переменной можно представить в:
  - а) номинальной шкале измерений
  - б) количественной шкале измерений
  - г) нельзя представить в какой-либо другой шкале измерений
  
3. Статистика как наука изучает:
  - а) массовые явления;
  - б) единичные явления;
  - в) периодические события.
  
4. Закон больших чисел утверждает, что:
  - а) чем больше единиц охвачено статистическим наблюдением, тем лучше проявляется общая закономерность;
  - б) чем больше единиц охвачено статистическим наблюдением, тем хуже проявляется общая закономерность;
  - в) чем меньше единиц охвачено статистическим наблюдением, тем лучше проявляется общая закономерность.
  
5. Назовите основные виды ошибок регистрации: а) случайные; б) систематические; в) ошибки репрезентативности; г) расчетные:
  - а) а, б, в,
  - б) а, б
  - в) а
  - г) а, б, в, г
  
6. В каких единицах будет выражаться относительный показатель, если база сравнения принимается за единицу?
  - а) в коэффициентах;

- б) в натуральных;
- в) в процентах.

7. Значения признака, повторяющиеся с наибольшей частотой, называется

- а) модой;
- б) медианой.

8. Коэффициент вариации измеряет колеблемость признака в:

- а) относительном выражении;
- б) абсолютном выражении.

9. Абсолютный прирост исчисляется как: а) отношение уровней ряда; б) разность уровней ряда. Темп роста исчисляется как: в) отношение уровней ряда; г) разность уровней ряда:

- а) б, в
- б) а, в
- в) а, г

10. К наиболее простым методам прогнозирования относят:

- а) метод на основе среднего абсолютного прироста;
- б) метод скользящей средней;
- в) индексный метод.

11. Статистический индекс - это:

- а) относительная величина сравнения двух показателей;
- б) сравнительная характеристика двух абсолютных величин;
- в) критерий сравнения относительных величин.

12. Табличные процессоры позволяют строить диаграммы следующих типов:

- а) Гистограмма, линейчатая диаграмма, круговая диаграмма, график;
- б) Диаграмма с областями, поверхностная диаграмма, лепестковая диаграмма;
- и) Всё выше перечисленное;

13. Задача классификации сводится к ...

- а) определения класса объекта по его характеристиками;
- б) нахождения частых зависимостей между объектами или событиями;
- в) определение по известным характеристиками объекта значение некоторого его параметра;
- г) поиска независимых групп и их характеристик в всем множестве анализируемых данных.

14. Задача регрессии сводится к ...

- а) определение по известным характеристиками объекта значение некоторого его параметра;
- б) определения класса объекта по его характеристиками;
- в) нахождения частых зависимостей между объектами или событиями;
- г) поиска независимых групп и их характеристик в всем множестве анализируемых данных.

15. Задача кластеризации заключается в ...

- а) поиска независимых групп и их характеристик в всем множестве анализируемых данных;
- б) определения класса объекта по его характеристиками;
- в) определение по известным характеристиками объекта значение некоторого его параметра;

г) нахождения частых зависимостей между объектами или событиями.

16. Целью поиска ассоциативных правил является ...

- а) нахождения частых зависимостей между объектами или событиями;
- б) определения класса объекта по его характеристиками;

в) определение по известным характеристиками объекта значение некоторого его параметра;

г) поиска независимых групп и их характеристик в всем множестве анализируемых данных.

17. Метаданные - ...

а) высокоуровневые средства отражения информационной модели и описания структуры данных

б) разновидность систем хранения, ориентирована на поддержку процесса анализа данных целостность, обеспечивает, непротиворечивость и хронологию данных, а также высокую скорость выполнения аналитических запросов

в) некоторый набор операций над базой данных, который рассматривается как единственное завершено, с точки зрения пользователя, действие над некоторой информацией, обычно связано с обращением к базе данных

г) это установление зависимости дискретной выходной переменной от входных переменных

18. Классификация - ...

а) это установление зависимости дискретной выходной переменной от входных переменных

б) разновидность систем хранения, ориентирована на поддержку процесса анализа данных целостность, обеспечивает, непротиворечивость и хронологию данных, а также высокую скорость выполнения аналитических запросов

в) высокоуровневые средства отражения информационной модели и описания структуры данных

г) некоторый набор операций над базой данных, который рассматривается как единственное завершено, с точки зрения пользователя, действие над некоторой информацией, обычно связано с обращением к базе данных

19. Регрессия - ...

а) это установление зависимости непрерывной выходной переменной от входных переменных

б) эта группировка объектов (Наблюдений, событий) на основе данных, описывающих свойства объектов

в) выявление закономерностей между связанными событиями

г) это установление зависимости дискретной выходной переменной от входных переменных

20. Кластеризация - ...

а) это установление зависимости непрерывной выходной переменной от входных переменных

б) эта группировка объектов (Наблюдений, событий) на основе данных, описывающих свойства объектов

в) выявление закономерностей между связанными событиями

г) это установление зависимости дискретной выходной переменной от входных переменных.

## 21. Ассоциация - ...

- а) эта группировка объектов (наблюдений, событий) на основе данных, описывающих свойства объектов
- б) это установление зависимости непрерывной выходной переменной от входных переменных
- в) выявление закономерностей между связанными событиями
- г) это установление зависимости дискретной выходной переменной от входных переменных

## 22. Машинное обучение - ...

- а) подразделение искусственного интеллекта изучающий методы построения алгоритмов, способных обучаться на данных
- б) эта группировка объектов (Наблюдений, событий) на основе данных, описывающих свойства объектов
- в) набор данных, каждая запись которого представляет собой учебный пример, содержащего заданный входной влияние, что и отвечает ему правильный выходной результат.
- г) специализированный программный решение (или набор решений), который включает в себя все инструменты для извлечения закономерностей из сырых данных

## 23. Аналитическая платформа - ...

- а) специализированный программный решение (или набор решений), который включает в себя все инструменты для извлечения закономерностей из сырых данных
- б) эта группировка объектов (Наблюдений, событий) на основе данных, описывающих свойства объектов
- в) набор данных, каждая запись которого представляет собой учебный пример, содержащего заданный входной влияние, что и отвечает ему правильный выходной результат.
- г) подразделение искусственного интеллекта изучающий методы построения алгоритмов, способных обучаться на данных

## 24. Обучающая выборка - ...

- а) набор данных, каждая запись которого представляет собой учебный пример, содержащего заданный входной влияние, и соответствующий ему правильный выходной результат
- б) эта группировка объектов (Наблюдений, событий) на основе данных, описывающих свойства объектов
- в) набор данных, каждая запись которого представляет собой учебный пример, содержащего заданный входной влияние, что и отвечает ему правильный выходной результат.
- г) выявление в сырых данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности

## 25. Ошибка обучения - ...

- а) это ошибка, допущенная моделью на учебной множества.
- б) это ошибка, полученная на тестовых примерах, то есть, что вычисляется по тем же формулам, но для тестовой множества
- в) имена, типы, метки и назначения полей исходной выборки данных
- г) набор данных, каждая запись которого представляет собой учебный пример, содержащего заданный входной влияние, и соответствующий ему правильный выходной результат

## **7.2. Методические материалы, определяющие процедуру оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций.**

### **а) Критерии оценивания компетенций (результатов).**

Программой дисциплины в целях проверки прочности усвоения материала предусматривается проведение различных форм контроля:

1. Текущий контроль – это проверка полноты знаний по основному материалу дисциплинарного модуля (ДМ).

2. Промежуточный контроль - итоговая проверка уровня знаний студента по данной дисциплине в конце семестра (в форме устного или письменного экзамена, сетевого компьютерного тестирования.) Промежуточной формой контроля знаний, умений и навыков по дисциплине является экзамен.

Общий результат выводится как интегральная оценка, складывающаяся из текущего контроля - 50% и промежуточного контроля - 50%.

Текущий контроль по дисциплине включает:

- посещение занятий - 10 баллов,
- выполнение лабораторных заданий – 50 баллов,
- выполнение домашних контрольных работ (самостоятельная работа) - 10 баллов.

Текущий контроль по ДМ:

письменная контрольная работа -15 баллов;

тестирование – 15 баллов;

Промежуточный контроль по дисциплине включает:

- устный экзамен (тестирование) - 100 баллов,

*Критерии оценки посещения занятий* – оценка выставляется по 100 бальной системе и соответствует проценту занятий, которые посетил студент из всего количества аудиторных занятий предусмотренных ДМ.

*Критерии оценки выполнения лабораторных заданий.*

Основными показателями оценки выполненной студентом и представленной для проверки работы являются:

1. Степень соответствия выполненного задания поставленным целям, задачам и требованиям;
2. Оформление, структурирование и комментирование лабораторной работы;
3. Уникальность выполнения работы (отличие от работ коллег);
4. Успешные ответы на контрольные вопросы.

Критерии оценки лабораторной работы.

86-100 баллов - оформление соответствует требованиям, критерии выдержаны, защита всего перечня контрольных вопросов.

66-85 баллов - оформление соответствует требованиям, критерии выдержаны, защита только до 85 % контрольных вопросов.

51-65 балл - оформление соответствует требованиям, критерии выдержаны, защита только до 51 % контрольных вопросов.

0-50 баллов – оформление не соответствует требованиям, критерии не выдержаны, защита только менее 51 % контрольных вопросов.

*Критерии оценки выполнения домашних контрольных работ (самостоятельная работа).*

Основными показателями оценки выполненной студентом и представленной для проверки домашней контрольной работы являются:

1. Степень соответствия выполненного задания поставленным целям, задачам и требованиям;
2. Оформление, структурирование и комментирование лабораторной работы;
3. Уникальность выполнения работы (отличие от работ коллег);

#### 4. Успешные ответы на контрольные вопросы.

Критерии оценки домашней контрольной работы.

86-100 баллов - студент правильно выполнил индивидуальное самостоятельное задание. Показал отличное владение навыками применения полученных знаний и умений при решении профессиональных задач в рамках усвоенного учебного материала. Ответил на все дополнительные вопросы на защите.

66-85 баллов - студент выполнил индивидуальное самостоятельное задание с небольшими неточностями. Показал хорошие владения навыками применения полученных знаний и умений при решении профессиональных задач в рамках усвоенного учебного материала. Ответил на большинство дополнительных вопросов на защите.

51-65 балл - студент выполнил индивидуальное самостоятельное задание с существенными неточностями. Показал удовлетворительное владение навыками применения полученных знаний и умений при решении профессиональных задач в рамках усвоенного учебного материала. При ответах на дополнительные вопросы на защите было допущено много неточностей.

0-50 баллов – при выполнении индивидуального самостоятельного задания студент продемонстрировал недостаточный уровень владения умениями и навыками при решении профессиональных задач в рамках усвоенного учебного материала. При ответах на дополнительные вопросы на защите было допущено множество неточностей.

*Критерии оценки текущего контроля по ДМ (письменная контрольная работа и тестирование).*

*Письменная контрольная работа* состоит из двух типов вопросов:

1. Теоретические вопросы из курса лекций и практических работ. - 40 баллов.
2. Практические вопросы и задачи по лекционному и практическому материалу. - 60 баллов.

86-100 баллов - студент, показал всестороннее, систематическое и глубокое знание учебного материала, самостоятельно ответил на вопросы, ответ отличается богатством и точностью использованных терминов, материал излагается последовательно и логично; показал отличное владение навыками применения полученных знаний и умений при решении профессиональных задач.

66-85 баллов - студент, показал полное знание учебного материала, не допускающий в ответе существенных неточностей, самостоятельно выполнивший ответивший на вопросы; показал хорошие владения навыками применения полученных знаний и умений при решении профессиональных задач

51-65 балл - студент, обнаруживший знание основного учебного материала в объёме, необходимом для дальнейшей учебы самостоятельно выполнивший задания, однако допустивший некоторые погрешности при ответе на вопросы; показал удовлетворительное владение навыками применения полученных знаний и умений при решении профессиональных задач.

0-50 баллов – выставляется студенту, обнаружившему пробелы в знаниях или отсутствие знаний по значительной части основного учебного материала, не выполнившего задания, допустившему принципиальные ошибки при ответе на вопросы, продемонстрировавший недостаточный уровень владения умениями и навыками при решении профессиональных задач.

Критерии выставления оценок за *тестирование* Тестовое задание состоит из пятнадцати вопросов. Время выполнения работы: 15-20 мин.

86-100 баллов - оценка «отлично» – 13-15 правильных ответов;

66-85 баллов - оценка «хорошо» – 10-12 правильных ответов;

51-65 балл - оценка «удовлетворительно» – 8-9 правильных ответов;

0-50 баллов – оценка «неудовлетворительно» – менее 8 правильных ответов.

*Критерии оценки устного экзамена*

Экзаменационные билеты включают три типа заданий:

1. Теоретические вопросы из курса лекций и практических работ. - 30 баллов.



2. Практические вопросы по лекционному и практическому материалу. - 40 баллов.

3. Проблемные вопросы и расчетные задачи. - 30 баллов.

В проверка качества подготовки студентов на экзаменах заканчивается выставлением отметок по принятой пятибалльной шкале («отлично», «хорошо», «удовлетворительно», «неудовлетворительно»).

86-100 баллов - оценка «отлично» - студент владеет знаниями по дисциплине «Методы анализа данных» в полном объеме учебной программы, достаточно глубоко осмысливает дисциплину; самостоятельно, в логической последовательности и исчерпывающе отвечает на все вопросы билета, подчеркивая при этом самое существенное, умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделять в нем главное: устанавливать причинно-следственные связи; четко формирует ответы, свободно решает ситуационные задачи повышенной сложности; хорошо знаком с основной литературой; увязывает теоретические аспекты дисциплины с прикладными задачами исследования операций и методов оптимизации; владеет современными информационными технологиями решения прикладных задач.

66-85 баллов - оценка «хорошо» – студент владеет знаниями дисциплины «Методы анализа данных» почти в полном объеме программы (имеются пробелы знаний только в некоторых, особенно сложных разделах); самостоятельно и отчасти при наводящих вопросах дает полноценные ответы на вопросы билета; не всегда выделяет наиболее существенное, не допускает вместе с тем серьезных ошибок в ответах; умеет решать легкие и средней тяжести ситуационные задачи; умеет трактовать выбор тех или иных методов и средств решения прикладных задач.

51-65 балл - оценка «удовлетворительно» - студент владеет основным объемом знаний по дисциплине «Методы анализа данных»; проявляет затруднения в самостоятельных ответах, оперирует неточными формулировками; в процессе ответов допускаются ошибки по существу вопросов. Студент способен решать лишь наиболее легкие задачи, владеет только обязательным минимумом при решении задач исследования операций.

0-50 баллов – оценка «неудовлетворительно» - студент не освоил обязательного минимума знаний дисциплины «Методы анализа данных», не способен ответить на вопросы билета даже при дополнительных наводящих вопросах экзаменатора.

*Критерии оценки экзамена в форме тестирования*

Тестовое задание состоит из тридцати вопросов. Время выполнения работы: 60 мин.

86-100 баллов - оценка «отлично» – 26-30 правильных ответов;

66-85 баллов - оценка «хорошо» – 20-25 правильных ответов;

51-65 балл - оценка «удовлетворительно» – 16-19 правильных ответов;

0-50 баллов – оценка «неудовлетворительно» – менее 16 правильных ответов.

## **8. Учебно-методическое обеспечение дисциплины.**

### **а) адрес сайта курса**

<http://eor.dgu.ru/>.

### **б) основная литература:**

1. Блинков, Ю. В. Основы теории информационных процессов и систем: учебное пособие / Ю. В. Блинков. - Пенза: Пензенский государственный университет архитектуры и строительства, ЭБС АСВ, 2011. — 184 с. - ISBN 978-5-9282-0725-0. - Текст: электронный // Электронно-библиотечная система IPR BOOKS: [сайт]. - URL: <https://www.iprbookshop.ru/23103.html> (дата обращения: 22.06.2022).
2. Егоров, Д. Л. Методы анализа данных и структур: учебное пособие / Д. Л. Егоров. - Казань: Казанский национальный исследовательский технологический университет, 2018. - 92 с. - ISBN 978-5-7882-2378-0. - Текст: электронный //

Электронно-библиотечная система IPR BOOKS: [сайт]. - URL: <https://www.iprbookshop.ru/95042.html> (дата обращения: 22.06.2022).

3. Кузнецов, А. С. Методы анализа данных: учебник / А. С. Кузнецов, Р. Ю. Царев, А. Н. Князьков. - Красноярск: Сибирский федеральный университет, 2015. - 184 с. - ISBN 978-5-7638-3193-1. - Текст: электронный // Электронно-библиотечная система IPR BOOKS: [сайт]. - URL: <https://www.iprbookshop.ru/84154.html> (дата обращения: 22.06.2022).

#### **в) дополнительная литература:**

1. Белов, В. С. Информационно-аналитические системы. Основы проектирования и применения: учебное пособие / В. С. Белов. - Москва: Евразийский открытый институт, 2010. - 112 с. - ISBN 978-5-374-00185-3. - Текст: электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. - Режим доступа: URL: <https://www.iprbookshop.ru/10678.html> (дата обращения: 11.02.2022)
2. Веретельникова, Е. Л. Методы анализа данных. Часть 2. Теория сетей Петри и моделирование систем: учебное пособие / Е. Л. Веретельникова. - Новосибирск: Новосибирский государственный технический университет, 2010. - 61 с. - ISBN 978-5-7782-1340-1. - Текст: электронный // Электронно-библиотечная система IPR BOOKS: [сайт]. - URL: <https://www.iprbookshop.ru/47720.html> (дата обращения: 22.06.2022).
3. Мельниченко, А. С. Математическая статистика и анализ данных: учебное пособие / А. С. Мельниченко. - Москва: Издательский Дом МИСиС, 2018. - 45 с. - ISBN 978-5-906953-62-9. - Текст: электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. - Режим доступа: URL: <https://www.iprbookshop.ru/78563.html> (дата обращения: 14.11.2021). (дата обращения: 22.02.2022).
4. Рязанов, Ю. Д. Методы анализа данных: лабораторный практикум. Учебное пособие / Ю. Д. Рязанов. - Белгород: Белгородский государственный технологический университет им. В.Г. Шухова, ЭБС АСВ, 2011. - 100 с. - Текст: электронный // Электронно-библиотечная система IPR BOOKS: [сайт]. - URL: <https://www.iprbookshop.ru/28402.html> (дата обращения: 22.06.2022).

#### **9. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины.**

1. eLIBRARY.RU [Электронный ресурс]: электронная библиотека. – Режим доступа: <http://elibrary.ru/defaultx.asp> (дата обращения: 22.06.2022). – Яз. рус., англ.
2. IPRbooks [Электронный ресурс]: Электронно-библиотечная система. - Режим доступа: <http://www.iprbookshop.ru/> (дата обращения: 11.02.2020). – Яз. рус., англ.
3. Университетская библиотека онлайн [Электронный ресурс]: Электронно-библиотечная система. - Режим доступа: <http://biblioclub.ru/> (дата обращения: 22.06.2022). – Яз. рус., англ.
4. Электронный каталог НБ ДГУ [Электронный ресурс]: база данных содержит сведения о всех видах лит, поступающих в фонд НБ ДГУ/Дагестанский гос. ун-т. – Махачкала, 2010 – Режим доступа: <http://elib.dgu.ru>, свободный (дата обращения: 22.06.2022).

#### **10. Методические указания для обучающихся по освоению дисциплины.**

Для изучения теоретического курса студентам необходимо использовать лекционный материал, учебники и учебные пособия из списка основной и дополнительной литературы, интернет источники.

По дисциплине «Методы анализа данных» в конце каждого модуля проводится контрольная работа.

В контрольную работу включаются теоретические вопросы и задачи тех типов, которые были разобраны на предшествующих практических занятиях.

Рабочей программой дисциплины «Методы анализа данных» предусмотрена самостоятельная работа студентов в объеме 36 часов. Самостоятельная работа проводится с целью углубления знаний по дисциплине и предусматривает:

– чтение студентами рекомендованной литературы и усвоение теоретического материала дисциплины;

– подготовку к практическим занятиям;

– выполнение индивидуальных заданий;

– подготовку к контрольным работам, зачету и экзаменам.

С самого начала изучения дисциплины студент должен четко уяснить, что без систематической самостоятельной работы успех невозможен. Эта работа должна регулярно начинаться сразу после лекционных и практических занятий, для закрепления только что пройденного материала.

После усвоения теоретического материала можно приступить к самостоятельному решению задач из учебников и пособий, входящих в список основной литературы.

**11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень программного обеспечения и информационных справочных систем.**

Microsoft Office (Excel, Power Point), Anaconda, Python, Jupyter Notebook

**12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине.**

Компьютерный класс, аудитория для проведения лекционных и практических занятий и самостоятельной работы средствами оборудованная оргтехникой, персональными компьютерами, объединенными в сеть с выходом в Интернет; установленное лицензионное и свободное программное обеспечение.