

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

Технологии анализа и обработки данных

Кафедра информационных систем и технологий программирования
факультета информатики и информационных технологий

Образовательная программа бакалавриата
09.03.03 Прикладная информатика

Направленность (профиль) программы
Информационные системы и программирование

Форма обучения
Очная

Статус дисциплины:
входит в часть ОПОП, формируемую участниками образовательных
отношений

Махачкала, 2022

Рабочая программа дисциплины «Технологии анализа и обработки данных» составлена в 2022 году в соответствии с требованиями ФГОС ВО – бакалавриат по направлению подготовки 09.03.03 Прикладная информатика от «19» сентября 2017г. № 922.

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 «Технологии анализа и обработки данных» входит часть, формируемую участниками образовательных отношений, образовательной программы бакалавриата по направлению 09.03.03 Прикладная информатика.

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

Содержание дисциплины охватывает круг вопросов, связанных 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: профессиональных ПК -6, ПК-9.

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 зачета.

Объем дисциплины 2 зачетные единицы, в том числе 72 в академических часах по видам учебных занятий

форма обучения - очная

Семестр	Учебные занятия					СРС	Форма промежуточной аттестации
	в том числе:						
	всего	Контактная работа обучающихся с преподавателем			36		
		всего	из них				
Лекции			Лабораторные занятия	Практические занятия			
1	72	54	18	18		36	зачет

форма обучения - заочная

Семестр	Учебные занятия					СРС	Форма промежуточной аттестации
	в том числе:						
	всего	Контактная работа обучающихся с преподавателем			60		
		всего	из них				
Лекции			Лабораторные занятия	Практические занятия			
1	60	12	6	6		60	зачет

1. Цели освоения дисциплины

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

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 «Технологии анализа и обработки данных» входит в часть, формируемую участниками образовательных отношений, образовательной программы по направлению 09.03.03 Прикладная информатика, профиль подготовки «Информационные системы и программирование».

Данную учебную дисциплину дополняет параллельное или последующее освоение следующих дисциплин: «Математика» и «Алгоритмы и структуры данных», «Эконометрика», «Технологии Big Data», «Разработка систем поддержки принятия решений».

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
ПК-6. Способность программировать приложения и создавать программные прототипы решения прикладных задач.	ИПК- 6.1. Знает основные сведения о методах и способах построения эффективных алгоритмов для решения прикладных задач. ИПК- 6.2. Умеет создавать программные прототипы решения задач предметной области. ИПК- 6.3. Владеет практическими навыками разработки программных прототипов решения прикладных задач.	<i>Знает:</i> основные понятия, методы, модели анализа и обработки данных и наиболее перспективные прикладные сферы их применения. <i>Умеет:</i> применять методы и способы построения эффективных алгоритмов и моделей для решения прикладных задач анализа и обработки данных <i>Владеет:</i> практическими навыками разработки и сопровождении информационных хранилищ, технологий оперативного и	Опрос, тестирование, контрольная работа

		интеллектуального анализа данных, отражающих деятельность предприятий в различных предметных областях.	
ПК-9. Способность применять системный подход и математические методы в формализации решения прикладных задач	ИПК- 11.1. Знать: принципы системного подхода и математические методы в формализации решения прикладных задач, в обосновании правильности выбранной модели информационных процессов и систем; ИПК- 11.2. Уметь: применять системный подход и математические методы в формализации решения прикладных задач; ИПК- 11.3. Владеть: методами построения математической модели профессиональных задач и содержательной интерпретации полученных результатов, навыками разработки информационно-логической, функциональной и объектно-ориентированной модели информационной системы, модели данных информационных систем.	<i>Знает:</i> основные этапы системного подхода при проведении анализа данных <i>Умеет:</i> использовать современные информационные технологии при сборе, обработке и системном анализе данных <i>Владеет:</i> Навыками использования компьютерных технологий для сбора, обработки и системного анализа данных	Опрос, тестирование, контрольная работа

4. Объем, структура и содержание дисциплины.

4.1. Объем дисциплины составляет 2 зачетные единицы, 72 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	1	1-2	2		2		6	Опрос, тестирование, контрольная работа
2	Первичная обработка данных статистического	1	3-4	2		2		6	Опрос, тестирование, контрольная работа

	наблюдения								
3	Индексный метод исследования данных	1	5-6	2		2		6	
4	Обзор компьютерных технологий анализа данных	1	7-8	2		2		6	Опрос, тестирование, контрольная работа
	<i>Итого по модулю 1:</i>			8		8		20	
Модуль 2. Компьютерные технологии анализа данных									
1	Анализ и визуализация данных в Microsoft Excel и Google Таблицах	1	9-10	2		2		6	Опрос, тестирование, контрольная работа
2	Аналитические информационные системы: состав, назначение, архитектура	1	11-12	2		2		6	Опрос, тестирование, контрольная работа
3	Классификация задач машинного обучения.	1	13-14	2		2		6	Опрос, тестирование, контрольная работа
4	Линейная регрессия	1	15-16	2		2		6	Опрос, тестирование, контрольная работа
5	Задачи классификации и кластеризации	1	17-18	2		2		6	Опрос, тестирование, контрольная работа
	<i>Итого по модулю 3:</i>			10		10		16	
	ИТОГО:			16		18		36	

в заочной 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	1	1-2	2				6	Опрос, тестирование, контрольная работа
2	Первичная обработка данных статистического наблюдения	1	3-4			2		8	Опрос, тестирование, контрольная работа
3	Индексный метод	1	5-6	2				6	

	исследования данных								
4	Обзор компьютерных технологий анализа данных	1	7-8					10	Опрос, тестирование, контрольная работа
	<i>Итого по модулю 1:</i>			4		2		30	
Модуль 2. Компьютерные технологии анализа данных									
1	Анализ и визуализация данных в Microsoft Excel и Google Таблицах	1	9-10			2		6	Опрос, тестирование, контрольная работа
2	Аналитические информационные системы: состав, назначение, архитектура	1	11-12					6	Опрос, тестирование, контрольная работа
3	Классификация задач машинного обучения.	1	13-14	2				6	Опрос, тестирование, контрольная работа
4	Линейная регрессия	1	15-16			2		6	Опрос, тестирование, контрольная работа
5	Задачи классификации и кластеризации	1	17-18					6	Опрос, тестирование, контрольная работа
	<i>Итого по модулю 2:</i>			2		4		30	
	ИТОГО:			6		6		36	

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

4.3.1. Содержание лекционных занятий по дисциплине.

Лекционный курс

№ п/п	Наименование темы	Трудоемкость	Содержание	Формируемые компетенции	Результаты освоения (знать, уметь, владеть)	Технологии обучения
1.	Анализ данных, основные понятия	2	Понятие данных. Структурированные и неструктурированные данные. Наука о данных. Этапы построения систем анализа данных. Основные источники данных. Основные понятия статистики данных.	ПК-6. ПК-9.	Знать основные понятия, методы, модели анализа и обработки данных и наиболее перспективные прикладные сферы их применения. Знать основные этапы системного подхода при проведении анализа данных	Опрос, тестирование, контрольная работа

2.	Первичная обработка данных статистического наблюдения	2	Абсолютные и относительные статистические величины, средние величины, вариационный анализ. Шкалы измерения. Функциональная, статистическая и корреляционная связи. Методы измерения тесноты парной корреляционной связи.	ПК-6.	Уметь применять методы первичной обработки данных статистического наблюдения для решения прикладных задач анализа и обработки данных	Опрос, тестирование, контрольная работа, кейс-задача
				ПК-9.	Уметь использовать современные информационные технологии при первичной обработке данных статистического наблюдения	
3.	Индексный метод исследования данных	2	Понятие об индексах, их классификация. Общие индексы количественных показателей. Общие индексы качественных показателей. Двухфакторные системы индексов. Индексы динамики средних показателей.	ПК-6.	Уметь индексный метод исследования данных для решения прикладных задач анализа и обработки данных	Опрос, тестирование, контрольная работа
				ПК-9.	Уметь использовать современные информационные технологии при применении индексных методов исследования данных	
4.	Обзор компьютерных технологий анализа данных	2	Особенности компьютерного анализа данных. Обзор современных информационных технологий и программных средств анализа и обработки данных.	ПК-6.	Знать компьютерных технологий анализа данных и перспективные прикладные сферы их применения	Опрос, тестирование, контрольная работа
5.	Анализ и визуализация данных в Microsoft Excel и Google Таблицах	2	Форматирование наборов и предварительная обработка данных в Microsoft Excel. Визуализация данных Описательная статистика. Сводные таблицы и сводные диаграммы в Microsoft Excel. Инструментарий в надстройке «Анализ данных»	ПК-9.	Владеть навыками использования компьютерных технологий для анализа и визуализации данных в Microsoft Excel и Google Таблицах	Опрос, тестирование, контрольная работа
6.	Аналитические информационные системы: состав, назначение, архитектура	2	Состав информационно-аналитической системы. Типы инструментальных средств создания и поддержки АИС. Основные принципы построения архитектуры АИС. Информационные хранилища.	ПК-6.	Знать состав, назначение, архитектуру аналитических информационных систем и перспективные прикладные сферы их применения	Опрос, тестирование, контрольная работа
					Владеть практическими навыками разработки и сопровождении информационных хранилищ, технологий оперативного и интеллектуального анализа данных, отражающих	

					деятельность предприятий в различных предметных областях.	
7.	Классификация задач машинного обучения.	2	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.	ПК-6.	Знать основные методы машинного обучения и наиболее перспективные прикладные сферы их применения	Опрос, тестирование, контрольная работа
8.	Линейная регрессия	2	Постановка задачи регрессионного анализа. Парная линейная регрессия. Множественная линейная регрессия. Градиентный спуск.	ПК-6.	Владеть практическими навыками разработки методов машинного обучения - регрессионного анализа и градиентного спуска.	Опрос, тестирование, контрольная работа
9.	Задачи классификации и кластеризации	2	Постановка задачи классификации с обучением. Логистическая регрессия. Понятие о деревьях решений. Постановка задачи кластерного анализа. Метод K-средних. Поиск аномалий.	ПК-6.	Владеть практическими навыками разработки методов машинного обучения - логистической регрессии, деревьев решений, k-средних.	

Лабораторные занятия

№ п/п	Наименование темы	Трудоемкость	Содержание	Формируемые компетенции	Результаты освоения (знать, уметь, владеть)	Технологии обучения
1.	Анализ данных, основные понятия	4	Лабораторная работа 1. Основы работы с аналитической платформой Deductor Studio.	ПК-6.	Владеть практическими навыками разработки и сопровождении информационных анализа данных с помощью аналитической платформы Deductor Studio.	Опрос, кейс-задача
				ПК-9.	Владеть навыками использования аналитической платформы Deductor Studio для сбора, обработки и системного анализа данных	
2.	Первичная обработка данных статистического наблюдения	4	Лабораторная работа 2. Обработка данных. Основные алгоритмы по очистке и	ПК-9.	Владеть навыками использования аналитической платформы Deductor Studio для предварительной обработки и интеллектуального анализа	Опрос, кейс-задача

			трансформации данных в Deductor Studio		данных	
3.	Индексный метод исследования данных	4	Лабораторная работа 3. Проектирование и создание хранилища данных	ПК-6.	Владеть практическими навыками разработки и сопровождении информационных хранилищ отражающих деятельность предприятий в различных предметных областях.	Опрос, кейс-задача
4.	Обзор компьютерных технологий анализа данных	4	Лабораторная работа 4. Наполнение хранилища данных	ПК-6.	Владеть практическими навыками разработки и сопровождении информационных хранилищ отражающих деятельность предприятий в различных предметных областях.	Опрос, кейс-задача
5.	Анализ и визуализация данных в Microsoft Excel и Google Таблицах	4	Лабораторная работа 5. Классификация с помощью дерева решений	ПК-6.	Владеть практическими навыками разработки методов машинного обучения с помощью дерева решений.	Опрос, кейс-задача
6.	Аналитические информационные системы: состав, назначение, архитектура	4	Лабораторная работа 6. Кластеризация методом k-средних	ПК-6.	Владеть практическими навыками разработки методов машинного обучения с помощью метода k-средних	Опрос, кейс-задача
7.	Классификация задач машинного обучения.	4	Лабораторная работа 7. Поиск ассоциативных правил	ПК-6.	Владеть практическими навыками разработки методов машинного обучения с помощью ассоциативных правил.	Опрос, кейс-задача
8.	Линейная регрессия	4	Лабораторная работа 8. Трансформация данных скользящим окном. Прогнозирование с помощью нейронной сети	ПК-6.	Владеть практическими навыками разработки нейронных сетей для прогнозирования данных в различных предметных областях	Опрос, кейс-задача
9.	Задачи классификации и кластеризации	4	Лабораторная работа 9. Декомпозиция временного ряда Лабораторная работа 10. Кластерный анализ. Карты Кохонена Лабораторная работа 11. Логистическая регрессия		Владеть практическими навыками разработки методов машинного обучения с помощью методов декомпозиция временного ряда, карт Кохонена и логистической регрессии	

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.

Тема 3. Индексный метод исследования данных

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 4. 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.

Тема 5. Анализ и визуализация данных в Microsoft Excel и Google Таблицах

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 6. 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 7. Классификация задач машинного обучения.

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.

Тема 8. Линейная регрессия

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 9. 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)

Лабораторная работа 1. Основы работы с аналитической платформой Deductor Studio

Лабораторная работа 2. Обработка данных. Основные алгоритмы по очистке и трансформации данных в Deductor Studio

Лабораторная работа 3. Проектирование и создание хранилища данных

Лабораторная работа 4. Наполнение хранилища данных

Лабораторная работа 5. Классификация с помощью дерева решений

Лабораторная работа 6. Кластеризация методом k-средних

Лабораторная работа 7. Поиск ассоциативных правил

Лабораторная работа 8. Трансформация данных скользящим окном.

Прогнозирование с помощью нейронной сети

Лабораторная работа 9. Декомпозиция временного ряда

Лабораторная работа 10. Кластерный анализ. Карты Кохонена

Лабораторная работа 11. Логистическая регрессия

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, а.ч.	Формируемые компетенции
	очная	
Текущая СРС		
работа с лекционным материалом, с учебной литературой	6	ПК-6, ПК-9
опережающая самостоятельная работа (изучение нового материала до его изложения на занятиях)	4	ПК-6, ПК-9
самостоятельное изучение разделов дисциплины	10	ПК-6, ПК-9
выполнение домашних заданий, домашних контрольных работ	8	ПК-6, ПК-9
подготовка к лабораторным работам, к практическим и семинарским занятиям	4	ПК-6, ПК-9
подготовка к контрольным работам, коллоквиумам, зачётам	6	ПК-6, ПК-9
Творческая проблемно-ориентированная СРС		
выполнение расчётно-графических работ	6	ПК-6, ПК-9
поиск, изучение и презентация информации по заданной проблеме, анализ научных публикаций по заданной теме	4	ПК-6, ПК-9
исследовательская работа, участие в конференциях, семинарах, олимпиадах	4	ПК-6, ПК-9
анализ данных по заданной теме, выполнение расчётов, составление схем и моделей на основе собранных данных	4	ПК-6, ПК-9
Итого СРС:	56	

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.

Разделы и темы для самостоятельного изучения	Виды и содержание самостоятельной работы
Статистические методы анализа данных	-конспектирование первоисточников и другой учебной литературы; -проработка учебного материала (по конспектам лекций учебной и научной литературе) и подготовка докладов на семинарах и практических занятиях, к участию в тематических дискуссиях; -поиск и обзор научных публикаций и электронных источников информации, подготовка заключения по обзору; -работа с тестами и вопросами для самопроверки;
Большие данные, сущность, признаки, особенности обработки и инструментарий	-конспектирование первоисточников и другой учебной литературы; -проработка учебного материала (по конспектам лекций учебной и научной литературе) и подготовка докладов на семинарах и практических занятиях, к участию в тематических дискуссиях; -поиск и обзор научных публикаций и электронных источников информации, подготовка заключения по обзору; -работа с тестами и вопросами для самопроверки; - решение домашних контрольных задач.
Машинное обучение и искусственный интеллект	-конспектирование первоисточников и другой учебной литературы; -проработка учебного материала (по конспектам лекций учебной и научной литературе) и подготовка докладов на семинарах и практических занятиях, к участию в тематических дискуссиях; -поиск и обзор научных публикаций и электронных источников информации, подготовка заключения по обзору; -работа с тестами и вопросами для самопроверки; - решение домашних контрольных задач.
Инструментарий Excel и Google Таблицы для анализа данных.	Изучение документации программного обеспечения
Языки программирования для анализа данных R, Python	Изучение документации программного обеспечения

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:

- зачет - 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 владения 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 отличные владения 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 на зачете

В проверка качества подготовки студентов на зачете заканчивается выставлением отметок «зачтено» и «незачтено».

51-100 баллов - оценка «зачтено» выставляется студенту, если студентом дан ответ, свидетельствующий о знании процессов изучаемой дисциплины, отличающийся глубиной и полнотой раскрытия темы, знанием основных вопросов теории, сформированными навыками анализа явлений, процессов, умением давать аргументированные ответы и приводить примеры, свободным владением монологической речью, логичностью и последовательностью ответа. Допускается несколько ошибок в содержании ответа;

0-50 баллов – оценка «не зачтено», если студентом дан ответ, который содержит ряд серьезных неточностей, обнаруживающий незнание процессов изучаемой предметной области, отличающийся неглубоким раскрытием темы, незнанием основных вопросов теории, несформированными навыками анализа явлений, процессов, неумением давать аргументированные ответы, слабым владением монологической речью, отсутствием логичности и последовательности. Выводы поверхностны. Т.е студент не способен ответить на вопросы даже при дополнительных наводящих вопросах преподавателя.

Критерии оценки зачета 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.

. Учебно-методическое обеспечение дисциплины.

а) адрес сайта курса

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

б) основная литература:

1. Александровская, Ю. П. Информационные технологии статистического анализа данных: учебно-методическое пособие / Ю. П. Александровская. - Казань: Казанский национальный исследовательский технологический университет, 2019. - 152 с. - ISBN 978-5-7882-2636-1. - Текст: электронный // Электронно-библиотечная система IPR BOOKS: [сайт]. - Режим доступа: URL: <https://www.iprbookshop.ru/100535.html> (дата обращения: 11.02.2022)
2. 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)
3. Замятин, А. В. Введение в интеллектуальный анализ данных: учебное пособие / А. В. Замятин. - Томск: Издательский Дом Томского государственного университета, 2016. - 119 с. - ISBN 978-5-94621-531-2. - Текст: электронный // Электронно-библиотечная система IPR BOOKS: [сайт]. - Режим доступа: URL: <https://www.iprbookshop.ru/109021.html> (дата обращения: 11.02.2022)

в) дополнительная литература:

1. Александровская, Ю. П. Многомерный статистический анализ в экономике: учебное пособие / Ю. П. Александровская. - Казань: Казанский национальный исследовательский технологический университет, 2017. - 96 с. - ISBN 978-5-7882-2191-5. - Текст: электронный // Электронно-библиотечная система IPR BOOKS: [сайт]. - Режим доступа: URL: <https://www.iprbookshop.ru/79330.html> (дата обращения: 22.02.2022).
2. 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).
3. Прокопенко, Н. Ю. Аналитические информационные системы поддержки принятия решений: учебное пособие / Н. Ю. Прокопенко. - Нижний Новгород: Нижегородский государственный архитектурно-строительный университет, ЭБС АСВ, 2020. - 143 с. - ISBN 978-5-528-00395-5. - Текст: электронный // Электронно-

библиотечная система IPR BOOKS: [сайт]. – Режим доступа: URL: <https://www.iprbookshop.ru/107361.html> (дата обращения: 22.02.2022).

4. Теория и практика машинного обучения: учебное пособие / В. В. Воронина, А. В. Михеев, Н. Г. Ярушкина, К. В. Святков. - Ульяновск: Ульяновский государственный технический университет, 2017. - 291 с. - ISBN 978-5-9795-1712-4. - Текст: электронный // Электронно-библиотечная система IPR BOOKS: [сайт]. - Режим доступа: URL: <https://www.iprbookshop.ru/106120.html>

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: 22.02.2022). – 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: 22.02.2022). – 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: 22.02.2022).

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 «Технологии анализа и обработки 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 «Технологии анализа и обработки данных» предусмотрена самостоятельная работа студентов в объеме 54 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.

Аналитическая платформа «DEDUCTOR» – для выполнения лабораторных работ

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

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