

МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Астраханский государственный университет имени В. Н. Татищева»
(Астраханский государственный университет им. В. Н. Татищева)

СОГЛАСОВАНО
Руководитель ОПОП
_____ А.Н. Марьенков

УТВЕРЖДАЮ
Заведующий кафедрой ИТ
_____ А.Н. Марьенков

«04» апреля 2024 г.

«04» апреля 2024 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

«Обработка и анализ данных»

Составитель(и)	Синельщиков А.В., доцент кафедры информационных технологий
Согласовано с работодателями:	
Направление подготовки / специальность	09.04.04. Программная инженерия
Направленность (профиль) / специализация ОПОП	Проектирование и разработка систем искусственного интеллекта
Квалификация (степень)	магистр
Форма обучения	Очная
Год приёма	2024
Курс	1
Семестр(ы)	1

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)

1.1. Целями освоения дисциплины (модуля) «Обработка и анализ данных» являются получение компетенций для решения цифровых задач в профессиональной деятельности

1.2. Задачи освоения дисциплины (модуля): «Обработка и анализ данных»

- изучение методов работы с данными, их области применения, достоинств и недостатков;
- способствовать формированию культуры представления, описания, интерпретации и оценки выводов над данными.

2. МЕСТО ДИСЦИПЛИНЫ (МОДУЛЯ) В СТРУКТУРЕ ОПОП

2.1. Учебная дисциплина (модуль) «Обработка и анализ данных» относится к части, формируемой участниками образовательных отношений, и осваивается в 1 семестре.

2.2. Для изучения данной учебной дисциплины (модуля) необходимы следующие знания, умения, навыки, формируемые предшествующими учебными дисциплинами (модулями):

- программирование на языке высокого уровня;
- теория вероятностей (уровень бакалавриата);
- базы данных (уровень бакалавриата)

Знания:

- Основы математического анализа (производные, интегралы, пределы).
- Основы линейной алгебры (векторы, матрицы, линейные преобразования).
- Основы теории вероятностей и математической статистики (распределения, гипотезы).

Умения:

- Программировать на одном из языков, используемых для анализа данных (Python, R, etc.).
- Визуализировать данные с помощью графиков и диаграмм.
- Работать с базами данных (SQL).

Навыки:

- Критически мыслить и интерпретировать результаты анализа.
- Формулировать задачи анализа данных и выбирать подходящие методы.
- Работать с большими объемами данных.

2.3. Последующие учебные дисциплины (модули) и (или) практики, для которых необходимы знания, умения, навыки, формируемые данной учебной дисциплиной (модулем):

- Методы машинного обучения
- Архитектура систем ИИ;
- Методология инженерии программных систем ИИ;
- Инфраструктуры больших данных;

3. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

Процесс освоения дисциплины (модуля) направлен на формирование элементов следующей(их) компетенции(ий) в соответствии с ФГОС ВО и ОПОП ВО по данному направлению подготовки / специальности:

- а) профессиональной (ПК).

ПК-3. Способен разрабатывать и модернизировать программное и аппаратное обеспечение технологий и систем искусственного интеллекта с учетом требований информационной безопасности в различных предметных областях

Таблица 1. Декомпозиция результатов обучения

Код компетенции	Код и наименование индикатора достижения компетенции ¹	Планируемые результаты обучения по дисциплине (модулю)		
		Знать (1)	Уметь (2)	Владеть (3)
ПК-3	ПК.3.1. Знать современные тенденции развития информационных технологий.	- логические методы и приемы научного исследования; методологические принципы современной науки, направления, концепции, источники знания и приемы работы с ними; основные особенности научного метода познания	- выбирать и применять методы структурирования знаний для построения концептуальных моделей знаний (онтологий знаний)	- методами решения профессиональные задачи на основе применения новых научных принципов и методов исследования
	ПК.3.2. Уметь разрабатывает программное и аппаратное обеспечение технологий и систем искусственного интеллекта для решения профессиональных задач с учетом требований информационной безопасности в различных предметных областях.	- методологические подходы к выбору и разработке методов получения знаний инженером по знаниям от экспертов; извлечения знаний из данных и текстов и применения соответствующих инструментальных средств	- разрабатывать программное и аппаратное обеспечение технологий и систем искусственного интеллекта для решения профессиональных задач с учетом требований информационной безопасности в различных предметных областях	- навыками использования инструментов извлечения знаний из данных и текстов
	ПК.3.3. Владеть навыками модернизации программного и аппаратного обеспечения технологий и систем искусственного интеллекта для решения профессиональных задач с учетом требований информационной безопасности в различных предметных областях.	- методологические подходы к выбору и применению методов обработки и распространения знаний с помощью дедукции, индукции и абдукции, согласования экспертных оценок и нечеткого вывода	- выбирать и применять методы обработки и распространения знаний для разработки программных компонентов систем, основанных на знаниях, и приложений	- навыками модернизации программного и аппаратного обеспечения технологий и систем искусственного интеллекта для решения профессиональных задач с учетом требований информационной безопасности в различных

¹ Указываются в соответствии с утвержденными в ОПОП ВО

Код компетенции	Код и наименование индикатора достижения компетенции ¹	Планируемые результаты обучения по дисциплине (модулю)		
		Знать (1)	Уметь (2)	Владеть (3)
				предметных областях

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Общая трудоемкость дисциплины в соответствии с учебным планом составляет 2 зачетные единицы (72 часа).

Трудоемкость отдельных видов учебной работы студентов очной формы обучения приведена в таблице 2.1.

Таблица 2.1. Трудоемкость отдельных видов учебной работы по формам обучения

Вид учебной и внеучебной работы	для очной формы обучения
Объем дисциплины в зачетных единицах	2
Объем дисциплины в академических часах	72
Контактная работа обучающихся с преподавателем (всего), в том числе (час.):	29,25
- занятия лекционного типа, в том числе:	14
- практическая подготовка (если предусмотрена)	–
- занятия семинарского типа (семинары, практические, лабораторные), в том числе:	14
- практическая подготовка (если предусмотрена)	–
- консультация (предэкзаменационная) ²	1
- промежуточная аттестация по дисциплине ³	0,25
Самостоятельная работа обучающихся (час.)	42,75
Форма промежуточной аттестации обучающегося (зачет/экзамен), семестр (ы)	экзамен – 1 семестр

Содержание дисциплины, структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий и самостоятельной работы, для каждой формы обучения представлено в таблице 2.2.

Таблица 2.2. Структура и содержание дисциплины (модуля) для очной формы обучения

Раздел, тема дисциплины (модуля)	Контактная работа, час.							КР / КП	СР, час.	Итого часов	Форма текущего контроля успеваемости, форма промежуточной аттестации [по семестрам]
	Л		ПЗ		ЛР						
	Л	в т.ч. ПП	ПЗ	в т.ч. ПП	ЛР	в т.ч. ПП					
Семестр 1.											

² Числовые данные в данной строке соответствуют трудоемкости, указанной в учебном плане в столбце «Конс. (для гр.)»

³ Числовые данные в данной строке соответствуют трудоемкости, указанной в учебном плане в столбце «КПА»

Раздел, тема дисциплины (модуля)	Контактная работа, час.						КР / КП	СР, час.	Итого часов	Форма текущего контроля успеваемости, форма промежуточной аттестации [по семестрам]
	Л		ПЗ		ЛР					
	Л	в т.ч. ПП	ПЗ	в т.ч. ПП	ЛР	в т.ч. ПП				
Тема 1. Первичная обработка и хранение данных	7				7			21	35	Лабораторная работа 1, тест
Тема 2. Элементы статистического анализа данных	7				7			21,75	35,75	Лабораторная работа 2, тест
Консультации									1	
Контроль промежуточной аттестации									0,25	Экзамен
ИТОГО за семестр:	14				14			42,75	72	
Итого за весь период	14				14			42,75	72	

Примечание: Л – лекция; ПЗ – практическое занятие, семинар; ЛР – лабораторная работа; ПП – практическая подготовка; КР / КП – курсовая работа / курсовой проект; СР – самостоятельная работа

Таблица 3. Матрица соотнесения разделов, тем учебной дисциплины (модуля) и формируемых компетенций

Раздел, тема дисциплины (модуля)	Кол-во часов	Код компетенции	Общее количество компетенций
		ПК-3	
Тема 1. Первичная обработка и хранение данных	35	+	1
Тема 2. Элементы статистического анализа данных	35,75	+	1
Консультация	1	+	1
Промежуточная аттестация по дисциплине	0,25	+	1
Итого	72		

Краткое содержание каждой темы дисциплины (модуля)

Тема 1. Первичная обработка и хранение данных

Виды и источники данных. Загрузка и разделение данных. Объединение данных из разных источников. Очистка данных и заполнение пропусков. Контроль диапазонов. Инструменты для обработки данных. Визуализация данных. Формы представления количественных и качественных данных. Когнитивная визуализация данных. Первичная обработка данных. Сглаживание и нормировка данных. Преобразование данных. Обработка временных рядов. NoSQL базы данных. Большие данные. Хранение и доступ к данным. Виды баз данных. Реляционные СУБД.

Тема 2. Элементы статистического анализа данных

Основы теории вероятностей. Случайные события, понятия и действия над ними. Одномерные и многомерные случайные величины, их типы распределений. Интервальное

оценивание и проверка гипотез, Точечное оценивание в параметрических и непараметрических случаях, сравнение оценок, эффективность. Эмпирическое распределение и описательная статистика. Действия со случайными величинами. Функции случайных величин. Характеристики случайных величин, характеристики зависимости. ЦПТ и ЗБЧ.

5. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ПРЕПОДАВАНИЮ И ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ)

5.1. Указания для преподавателей по организации и проведению учебных занятий по дисциплине (модулю)

Лекционные занятия

Основной формой реализации теоретического обучения является лекция, которая представляет собой систематическое, последовательное изложение преподавателем-лектором учебного материала теоретического характера. Цель лекции – организация целенаправленной познавательной деятельности студентов по овладению программным материалом учебной дисциплины.

Порядок подготовки лекционного занятия включает в себя выполнение следующих этапов:

- изучение требований программы дисциплины;
- определение целей и задач лекции;
- разработка плана проведения лекции;
- подбор литературы (ознакомление с методической литературой, публикациями периодической печати по теме лекционного занятия);
- отбор необходимого и достаточного по содержанию учебного материала;
- определение методов, приемов и средств поддержания интереса, внимания, стимулирования творческого мышления студентов;
- написание конспекта лекции.

Лекция должна включать следующие разделы:

- формулировку темы лекции;
- указание основных изучаемых разделов или вопросов и предполагаемых затрат времени на их изложение;
- изложение вводной части;
- изложение основной части лекции;
- краткие выводы по каждому из вопросов;
- заключение;
- рекомендации литературных источников по излагаемым вопросам.

Самостоятельная работа

Самостоятельная работа – это вид учебной деятельности, которую студент совершает в установленное время и в установленном объеме индивидуально или в группе, без непосредственной помощи преподавателя (но при его контроле), руководствуясь сформированными ранее представлениями о порядке и правильности выполнения действий.

В учебном процессе образовательного учреждения выделяются два вида самостоятельной работы:

- аудиторная – выполняется на учебных занятиях, под непосредственным руководством преподавателя и по его заданию (выполнение самостоятельных работ; выполнение контрольных и практических работ; решение задач);
- внеаудиторная – выполняется по заданию преподавателя, но без его непосредственного участия (подготовка к аудиторным занятиям; изучение учебного материала, вынесенного на самостоятельную проработку; выполнение домашних заданий разнообразного характера; выполнение индивидуальных заданий, направленных на развитие у студентов самостоятельности и инициативы; подготовка к контрольной работе). Внеаудиторные

самостоятельные работы представляют собой логическое продолжение аудиторных занятий, проводятся по заданию преподавателя, который инструктирует студентов и устанавливает сроки выполнения задания.

5.2. Указания для обучающихся по освоению дисциплины (модулю)

Лекция

- Лекция – основной вид обучения в вузе.
- В лекции излагаются основные положения теории, ее понятия и законы, приводятся факты, показывающие связь теории с практикой.
- Накануне лекции необходимо повторить содержание предыдущей лекции (а также теорию по изучаемой теме в школьных учебниках геометрии, если эта тема была представлена в них), а затем посмотреть тему очередной лекции по программе (по плану лекций).
- Полезно вести записи (конспекты) лекций: для непонятных вопросов оставлять место при работе над темой лекции с учебными пособиями.
- Записи лекций следует вести в отдельной тетради, оставляя место для дополнений во время самостоятельной работы.
- При конспектировании лекций выделяйте главы и разделы, параграфы, подчеркивайте основное.

Организация самостоятельной работы

Самостоятельность в учебной работе способствует развитию заинтересованности студента в изучаемом материале, вырабатывает у него умение и потребность самостоятельно получать знания, что весьма важно для специалиста с высшим образованием. Самостоятельная работа студентов представлена в следующих формах:

- работа с учебной литературой и конспектом лекций с целью подготовки к лабораторным занятиям, составление конспектов тем, выносимых на самостоятельную проработку;
- систематическое выполнение домашних работ.

**Таблица 4. Содержание самостоятельной работы обучающихся
для очной формы обучения**

Вопросы, выносимые на самостоятельное изучение	Кол-во часов	Форма работы
Тема 1. Первичная обработка и хранение данных	21	Изучение теоретического материала. Подготовка к лабораторной работе. Подготовка к тесту.
Тема 2. Элементы статистического анализа данных	21,75	Изучение теоретического материала. Подготовка к лабораторной работе. Подготовка к тесту.

5.3. Виды и формы письменных работ, предусмотренных при освоении дисциплины (модуля), выполняемые обучающимися самостоятельно

Письменные работы, выполняемые обучающимися самостоятельно, не предусмотрены.

6. ОБРАЗОВАТЕЛЬНЫЕ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

При реализации различных видов учебной работы по дисциплине могут использоваться электронное обучение и дистанционные образовательные технологии.

Основой для выстраивания аудиторных занятий послужила технология развития критического мышления, которая, интегрируя элементы проблемного, проектного, дискуссионного обучения, позволяет достигать максимальной эффективности в достижении проектируемых компетенций.

6.1. Образовательные технологии

Цели дисциплины достигаются путем сочетания контактной и самостоятельной работы студентов: проведения лекционных занятий, лабораторно-практических занятий на ПК и организации самостоятельной работы студентов.

Лекционные занятия организуются с применением традиционных и инновационных технологий организации учебной деятельности студентов.

На лекциях рассматриваются теоретические основы информатики и вычислительной техники, примеры решения практических задач. Обеспечивается демонстрационная поддержка изложения курса в форме компьютерной презентации. Это способствует передаче большего количества учебного материала обучающимся во время аудиторных занятий и более доходчивому его освоению. В то же время, для студентов первого курса рекомендуется практические примеры разбирать, пользуясь традиционной технологией «доски и мела», поскольку это позволяет включить обучаемого в процесс решения задачи.

На занятиях предусматривается сочетание индивидуальной и групповой форм работы с обязательным общим обсуждением и презентацией выполняемых работ. Лекционные занятия организуются с применением традиционных и инновационных технологий организации учебной деятельности студентов: проблемная лекция. Лекция с заранее запланированными ошибками, проектное обучения.

Таблица 5 – Образовательные технологии, используемые при реализации учебных занятий

Раздел, тема дисциплины (модуля)	Форма учебного занятия		
	Лекция	Практическое занятие, семинар	Лабораторная работа
Тема 1. Первичная обработка и хранение данных	Обзорная лекция	Не предусмотрено	Фронтальный опрос, выполнение практических заданий, тестирование
Тема 2. Элементы статистического анализа данных	Лекция-диалог	Не предусмотрено	Фронтальный опрос, выполнение практических заданий, тестирование

Лабораторно-практические работы выполняются студентами с применением ПК и ориентированы на формирование деятельностных компетентностей. Они заключаются в выполнении сквозного цикла лабораторных работ. В процессе выполнения лабораторных работ достигаются следующие цели:

- изучаются программные средства и технологии обработки информации;
- формируются практические навыки обработки информации различного вида и формы при решении конкретных практических задач;
- формируется навык выявления ошибочных и нестандартных ситуаций и реагирования на них.

На лабораторных занятиях студент вначале знакомится с содержанием работы, пользуясь электронными методическими материалами, размещенными на <http://moodle.asu.edu.ru>, затем выполняет задание и показывает результаты преподавателю. Лабораторные работы, выполняются студентом самостоятельно, возникающие при их выполнении проблемы разрешаются в рамках учебного времени и индивидуальных и групповых консультаций. Для

выставления баллов по итогам выполнения ЛР, студенты прикрепляют файлы с выполненными работами и отчеты на образовательный портал.

Текущая аттестация студентов проводится в форме контрольных работ, представленных в виде компьютерного теста, в ходе которого студент должен продемонстрировать освоение соответствующей технологии.

Для **самостоятельного изучения** теоретического материала дисциплины рекомендуется использовать Internet-ресурсы, информационные базы, методические разработки, специальную учебную и научную литературу.

В рамках организации самостоятельной работы студентам рекомендуется:

- работа с лекционным материалом;
- дополнительная подготовка к лабораторно-практическим работам или выполнение части работы, которую они не успели сделать в аудитории, оформление отчетов;
- выполнение творческих проектов, типовых расчетов, подготовка отчетов к их защите;
- подготовка к компьютерному тестированию;
- подготовка к текущей и промежуточной аттестации (экзамену). Для обеспечения самостоятельной работы разработаны:
- учебно-методическое обеспечение лекционных занятий;
- методические рекомендации по выполнению лабораторно-практических работ, творческих проектов, типовых расчетов, требования к оформлению и представлению отчетов по выполнению;
- тренировочные тестовые задания;
- методические рекомендации к самостоятельной работе студентов.

Оценка результатов самостоятельной работы организуется как единство двух форм: самоконтроль и контроль со стороны преподавателей.

Задача преподавателя состоит в том, чтобы создать условия для выполнения самостоятельной работы (учебно-методическое обеспечение), правильно использовать различные стимулы для реализации этой работы (рейтинговая система), повышать её значимость, и грамотно осуществлять контроль самостоятельной деятельности студента.

Учебные занятия по дисциплине могут проводиться с применением информационно-телекоммуникационных сетей при опосредованном (на расстоянии) интерактивном взаимодействии обучающихся и преподавателя в режимах on-line в формах: видеолекций, лекций-презентаций, видеоконференции, собеседования в режиме чат, форума, чата, выполнения виртуальных практических и/или лабораторных работ и др.

Максимальный объем занятий обучающегося с применением электронных образовательных технологий не должен превышать 25%

6.2. Информационные технологии

При реализации различных видов учебной и вне учебной работы используются следующие информационные технологии:

- использование образовательного сайта <http://moodle.asu.edu.ru> (размещение учебно-методического материала, публикация заданий для предоставления студентами выполненных отчетов по всем видам работ, ознакомление учащихся с оценками и т.д., размещение объявлений, on-line консультации, организация и проведение компьютерного тестирования, обсуждение вопросов в форуме и т.д.), как элемента интерактивного взаимодействия участников образовательного процесса (технологии дистанционного обучения);
- использование ресурсов ЭБС и сети Internet, как источников информации.
- иные информационные системы, сервисы и мессенджеры.

6.3. Программное обеспечение, современные профессиональные базы данных и информационные справочные системы

6.3.1. Программное обеспечение

Наименование программного обеспечения	Назначение
Adobe Reader	Программа для просмотра электронных документов
Moodle	Образовательный портал ФГБОУ ВО «АГУ»
Mozilla FireFox	Браузер
Microsoft Office 2013, Microsoft Office Project 2013, Microsoft Office Visio 2013	Пакет офисных программ
7-zip	Архиватор
Microsoft Windows 10 Professional	Операционная система
Paint .NET	Растровый графический редактор
Blender	Средство создания трёхмерной компьютерной графики
GIMP	Многоплатформенное программное обеспечение для работы над изображениями.
Inkscape	Свободно распространяемый векторный графический редактор, удобен для создания как художественных, так и технических иллюстраций
LibreOffice	Пакет офисных программ.

6.3.2. Современные профессиональные базы данных и информационные справочные системы

1. Электронный каталог Научной библиотеки АГУ на базе MARK SQL НПО «Информ- систем»: <https://library.asu.edu.ru>.
2. Электронная библиотека «Астраханский государственный университет» собственной генерации на электронной платформе ООО «БИБЛИОТЕХ»: <https://biblio.asu.edu.ru>.
3. Электронный каталог «Научные журналы АГУ»: <http://journal.asu.edu.ru/>.
4. Универсальная справочно-информационная полнотекстовая база данных периодических изданий ООО «ИВИС»: <http://dlib.eastview.com/>
5. Электронно-библиотечная система elibrary. <http://elibrary.ru>
6. Справочная правовая система КонсультантПлюс: <http://www.consultant.ru>
7. Информационно-правовое обеспечение «Система ГАРАНТ»: <http://garant-astrakhan.ru>

7. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕГО КОНТРОЛЯ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

7.1. Паспорт фонда оценочных средств

При проведении текущего контроля и промежуточной аттестации по дисциплине (модулю) «Обработка и анализ данных» проверяется сформированность у обучающихся компетенций, указанных в разделе 3 настоящей программы. Этапность формирования данных компетенций в процессе освоения образовательной программы определяется последовательным освоением дисциплин (модулей) и прохождением практик, а в процессе освоения дисциплины (модуля) – последовательным достижением результатов освоения содержательно связанных между собой разделов, тем.

Таблица 6. Соответствие разделов, тем дисциплины (модуля), результатов обучения по дисциплине (модулю) и оценочных средств

Контролируемый раздел, тема дисциплины (модуля)	Код контролируемой компетенции	Наименование оценочного средства
Тема 1. Первичная обработка и хранение данных	ПК-3	Отчет о выполнении ЛПР, тестирование, экзамен
Тема 2. Элементы статистического анализа данных	ПК-3	Отчет о выполнении ЛПР, тестирование, экзамен

7.2. Описание показателей и критериев оценивания компетенций, описание шкал оценивания

В таблицах 7–8 приводятся примерные показатели и критерии оценивания компетенций, шкалы оценивания

Таблица 7. Показатели оценивания результатов обучения в виде знаний

Шкала оценивания	Критерии оценивания
5 «отлично»	демонстрирует глубокое знание теоретического материала, умение обоснованно излагать свои мысли по обсуждаемым вопросам, способность полно, правильно и аргументированно отвечать на вопросы, приводить примеры
4 «хорошо»	демонстрирует знание теоретического материала, его последовательное изложение, способность приводить примеры, допускает единичные ошибки, исправляемые после замечания преподавателя
3 «удовлетворительно»	демонстрирует неполное, фрагментарное знание теоретического материала, требующее наводящих вопросов преподавателя, допускает существенные ошибки в его изложении, затрудняется в приведении примеров и формулировке выводов
2 «неудовлетворительно»	демонстрирует существенные пробелы в знании теоретического материала, не способен его изложить и ответить на наводящие вопросы преподавателя, не может привести примеры

Таблица 8. Показатели оценивания результатов обучения в виде умений и владений

Шкала оценивания	Критерии оценивания
5 «отлично»	демонстрирует способность применять знание теоретического материала при выполнении заданий, последовательно и правильно выполняет задания, умеет обоснованно излагать свои мысли и делать необходимые выводы
4 «хорошо»	демонстрирует способность применять знание теоретического материала при выполнении заданий, последовательно и правильно выполняет задания, умеет обоснованно излагать свои мысли и делать необходимые выводы, допускает единичные ошибки, исправляемые после замечания преподавателя
3 «удовлетворительно»	демонстрирует отдельные, несистематизированные навыки, испытывает затруднения и допускает ошибки при выполнении заданий, выполняет задание по подсказке преподавателя, затрудняется в формулировке выводов
2 «неудовлетворительно»	не способен правильно выполнить задания

7.3. Контрольные задания и иные материалы, необходимые для оценки результатов обучения по дисциплине (модулю)

Полный комплект оценочных средств размещен на <http://moodle.asu.edu.ru>. Допуск студентов осуществляется по расписанию проведения аудиторных занятий и сдачи отчетов по выполнению самостоятельной работы.

Тема 1. Первичная обработка и хранение данных

– Лабораторная работа № 1

Задание 1: Очистка данных

Описание: Вам предоставлен CSV файл с данными о продажах товаров. В данных присутствуют пропущенные значения, дубликаты и некорректные форматы данных.

Задача:

1. Загрузите данные из CSV файла с помощью библиотеки Pandas.
2. Удалите дубликаты строк.
3. Заполните пропущенные значения в столбце "Цена" средним значением цены для данного товара.
4. Преобразуйте столбец "Дата" в формат datetime.
5. Сохраните очищенные данные в новый CSV файл.

Задание 2: Хранение данных в базе данных SQLite

Описание: Вам предоставлен тот же CSV файл с данными о продажах товаров.

Задача:

1. Создайте базу данных SQLite с именем "sales.db".
2. Создайте таблицу "sales" с полями: "id" (INTEGER PRIMARY KEY), "Товар" (TEXT), "Цена" (REAL), "Дата" (TEXT), "Количество" (INTEGER).
3. Загрузите данные из CSV файла в таблицу "sales".
4. Выполните SQL запрос для получения информации о продажах товара "Телефон" за последний месяц.

– Тестовые задания

Задание 1:

Вопрос: Какие из перечисленных методов используются для обработки пропущенных значений в данных?

Варианты ответа:

- a) Замена пропущенных значений средним значением.
- b) Удаление строк с пропущенными значениями.
- c) Замена пропущенных значений медианой.
- d) Игнорирование пропущенных значений.
- e) Замена пропущенных значений наиболее частым значением (модой).

Правильный ответ: a, b, c, e

Задание 2:

Вопрос: Какие библиотеки Python используются для работы с данными в формате CSV?

Варианты ответа:

- a) Pandas
- b) NumPy
- c) Scikit-learn
- d) Matplotlib
- e) CSV

Правильный ответ: a, e

Задание 3:

Вопрос: Какие из перечисленных типов данных могут храниться в базе данных SQLite?

Варианты ответа:

- a) INTEGER
- b) TEXT
- c) REAL
- d) BLOB
- e) LIST

Правильный ответ: a, b, c, d

Задание 4:

Вопрос: Какие методы используются для обнаружения дубликатов в данных?

Варианты ответа:

- a) Сортировка данных и сравнение соседних строк.
- b) Использование хеш-функций.
- c) Поиск по уникальному идентификатору.
- d) Визуальный анализ данных.
- e) Использование алгоритмов машинного обучения.

Правильный ответ: a, b, c

Задание 5:

Вопрос: Какие из перечисленных форматов данных используются для хранения структурированной информации?

Варианты ответа:

- a) CSV
- b) JSON
- c) XML
- d) TXT
- e) MP3

Правильный ответ: a, b, c

Задание 6:

Вопрос: Какие из перечисленных инструментов используются для визуализации данных?

Варианты ответа:

- a) Matplotlib
- b) Seaborn
- c) Pandas
- d) Tableau
- e) Power BI

Правильный ответ: a, b, d, e

Задание 7:

Вопрос: Какие из перечисленных методов используются для нормализации данных?

Варианты ответа:

- a) Min-Max scaling
- b) Z-score normalization
- c) Decimal scaling
- d) Log transformation
- e) One-hot encoding

Правильный ответ: a, b, c

Задание 8:

Вопрос: Какие из перечисленных операций можно выполнять с помощью SQL запросов?

Варианты ответа:

- a) Создание таблиц

- b) Вставка данных
- c) Обновление данных
- d) Удаление данных
- e) Выборка данных

Правильный ответ: a, b, c, d, e

Тема 2. Элементы статистического анализа данных

– Лабораторная работа № 2

Задание 1: Описательная статистика

Описание: Вам предоставлен CSV файл sales_data.csv с данными о продажах товаров в разных магазинах.

Задача:

1. Загрузите данные из файла sales_data.csv с помощью библиотеки Pandas.
2. Вычислите основные описательные статистики для столбца Продажи (например, среднее значение, медиану, стандартное отклонение, минимум, максимум, квартили).
3. Визуализируйте распределение продаж с помощью гистограммы.
4. Вычислите корреляцию между столбцами Продажи и Цена.
5. Интерпретируйте полученные результаты.

Задание 2: Проверка гипотез

Описание: Вам предоставлен CSV файл experiment_data.csv с данными о результатах A/B тестирования двух версий сайта.

Задача:

1. Загрузите данные из файла experiment_data.csv.
2. Сформулируйте нулевую и альтернативную гипотезы о равенстве средних конверсий двух версий сайта.
3. Проверьте гипотезу с помощью t-теста (например, scipy.stats.ttest_ind).
4. Интерпретируйте полученные результаты и сделайте вывод о том, какая версия сайта лучше.

– Тестовые задания

Вопрос 1:

Какие из перечисленных мер являются мерами центральной тенденции?

Варианты ответа:

- a) Среднее арифметическое
- b) Медиана
- c) Мода
- d) Стандартное отклонение
- e) Дисперсия

Правильный ответ: a, b, c

Вопрос 2:

Какие из перечисленных мер являются мерами вариации?

Варианты ответа:

- a) Среднее арифметическое
- b) Медиана
- c) Мода
- d) Стандартное отклонение
- e) Дисперсия
- f) Размах

Правильный ответ: d, e, f

Вопрос 3:

Что показывает коэффициент корреляции?

Варианты ответа:

- a) Силу линейной связи между двумя переменными
- b) Направление связи между двумя переменными
- c) Причинно-следственную связь между двумя переменными
- d) Разницу между средними значениями двух групп

Правильный ответ: a, b

Вопрос 4:

Какие из перечисленных методов используются для проверки гипотез?

Варианты ответа:

- a) t-тест
- b) ANOVA
- c) Хи-квадрат тест
- d) Регрессионный анализ
- e) Кластерный анализ

Правильный ответ: a, b, c

Вопрос 5:

Что такое р-значение в статистике?

Варианты ответа:

- a) Вероятность получить наблюдаемые данные или еще более крайние данные, если нулевая гипотеза верна
- b) Вероятность того, что нулевая гипотеза верна
- c) Вероятность того, что альтернативная гипотеза верна
- d) Уровень значимости

Правильный ответ: a

Вопрос 6:

Какие из перечисленных методов относятся к методам машинного обучения?

Варианты ответа:

- a) Линейная регрессия
- b) Логистическая регрессия
- c) Деревья решений
- d) Кластеризация
- e) t-тест

Правильный ответ: a, b, c, d

Вопрос 7:

Что такое выброс в статистике?

Варианты ответа:

- a) Значение, которое значительно отличается от остальных значений в наборе данных
- b) Среднее значение набора данных
- c) Медиана набора данных
- d) Мода набора данных

Правильный ответ: a

Вопрос 8:

Какие из перечисленных методов используются для визуализации данных?

Варианты ответа:

- a) Гистограмма

- b) Диаграмма рассеяния
- c) Ящик с усами
- d) Круговая диаграмма
- e) t-тест

Правильный ответ: a, b, c, d

Вопрос 9:

Что такое нормальное распределение?

Варианты ответа:

- a) Распределение, которое имеет форму колокола
- b) Распределение, которое симметрично относительно среднего значения
- c) Распределение, которое описывается двумя параметрами: средним значением и стандартным отклонением
- d) Распределение, которое используется для описания многих природных явлений

Правильный ответ: a, b, c, d

Вопрос 10:

Что такое регрессионный анализ?

Варианты ответа:

- a) Метод, который используется для моделирования зависимости между двумя или более переменными
- b) Метод, который используется для проверки гипотез
- c) Метод, который используется для кластеризации данных
- d) Метод, который используется для визуализации данных

Правильный ответ: a

Перечень вопросов, выносимых на экзамен

Общие вопросы:

1. Что такое обработка данных? Опишите основные этапы обработки данных.
2. Какие типы данных существуют? Приведите примеры каждого типа.
3. Что такое анализ данных? Какие цели преследует анализ данных?
4. Какие виды анализа данных существуют? Опишите каждый вид.
5. Какие инструменты и технологии используются для обработки и анализа данных?
6. Что такое Big Data? Какие особенности Big Data?
7. Какие этические вопросы возникают при обработке и анализе данных?

Вопросы по очистке и подготовке данных:

8. Какие проблемы могут возникать с качеством данных?
9. Какие методы используются для очистки данных? Приведите примеры.
10. Что такое пропущенные значения? Как обрабатывать пропущенные значения?
11. Что такое выбросы? Как обнаруживать и обрабатывать выбросы?
12. Что такое нормализация данных? Зачем нужна нормализация данных?
13. Какие методы используются для преобразования данных?

Вопросы по анализу данных:

14. Что такое описательная статистика? Какие меры используются в описательной статистике?
15. Что такое корреляция? Как интерпретировать коэффициент корреляции?
16. Что такое регрессионный анализ? Какие типы регрессии существуют?
17. Что такое классификация? Какие алгоритмы классификации существуют?
18. Что такое кластеризация? Какие алгоритмы кластеризации существуют?
19. Что такое временные ряды? Какие методы используются для анализа временных рядов?

Вопросы по визуализации данных:

20. Зачем нужна визуализация данных?

21. Какие типы графиков используются для визуализации данных? Приведите примеры.
22. Какие инструменты используются для визуализации данных?
23. Какие принципы эффективной визуализации данных?

Вопросы по хранению данных:

24. Какие типы баз данных существуют?
25. Что такое реляционные базы данных?
26. Что такое NoSQL базы данных?
27. Какие преимущества и недостатки разных типов баз данных?
28. Что такое хранилища данных?

Вопросы по применению обработки и анализа данных:

29. Приведите примеры применения обработки и анализа данных в бизнесе.
30. Приведите примеры применения обработки и анализа данных в науке.
31. Какие перспективы развития обработки и анализа данных?

Дополнительные вопросы:

32. Что такое машинное обучение? Как машинное обучение связано с обработкой и анализом данных?
33. Что такое глубокое обучение? Какие преимущества глубокого обучения?
34. Что такое искусственный интеллект? Как искусственный интеллект связан с обработкой и анализом данных?

Полный комплект оценочных материалов по дисциплине (модулю) (фонд оценочных средств) хранится в электронном виде на кафедре, утверждающей рабочую программу дисциплины (модуля).

7.4. Методические материалы, определяющие процедуры оценивания результатов обучения по дисциплине (модулю)

Таблица 10 – Технологическая карта рейтинговых баллов по дисциплине (модулю)

№ п/п	Контролируемые мероприятия	Количество мероприятий / баллы	Максимальное количество баллов	Срок представления
Основной блок				
1.	Ответ на занятии	5/2	10	По расписанию
2.	Выполнение лабораторной работы	10/3	30	По расписанию
Всего			40	–
Блок бонусов				
3.	Посещение занятий	9/1	9	По расписанию
4.	Своевременное выполнение всех заданий	10/1	10	По расписанию
Всего			19	–
Дополнительный блок				
5.	Экзамен		31	
Всего			30	-
ИТОГО			100	-

Таблица 11 – Система штрафов (для одного занятия)

Показатель	Балл
------------	------

Опоздание (два и более)	-2
Не готов к практической части занятия	-3
Нарушение учебной дисциплины	-2
Пропуски лекций без уважительных причин (за одну лекцию)	-2
Пропуск занятий без уважительной причины (за одно занятие)	-2
Нарушение правил техники безопасности	-1
Отсутствие конспектов лекций, семинарских занятий, первоисточников при начислении баллов не учитываются	0

Таблица 12 – Шкала перевода рейтинговых баллов в итоговую оценку за семестр по дисциплине (модулю)

Сумма баллов	Оценка по 4-балльной шкале
90–100	5 (отлично)
85–89	4 (хорошо)
75–84	
70–74	
65–69	3 (удовлетворительно)
60–64	
Ниже 60	2 (неудовлетворительно)

При реализации дисциплины (модуля) в зависимости от уровня подготовленности обучающихся могут быть использованы иные формы, методы контроля и оценочные средства, исходя из конкретной ситуации.

8. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

8.1. Основная литература

1. Анализ данных: учебник для академического бакалавриата / В. С. Мхитарян [и др.] ; под ред. В. С. Мхитаряна. – М.: Издательство Юрайт, 2018. – 490 с. – (Серия : Бакалавр. Академический курс). – ISBN 978-5-534-00616-2. – Режим доступа: www.biblio-online.ru/book/CC38E97A-CCE5-4470-90F1-3B6D35ACC0B4.
2. Миркин Б. Г. Введение в анализ данных: учебник и практикум / Б. Г. Миркин. М.: Издательство Юрайт, 2018. – 174 с. – (Серия: Авторский учебник). – ISBN 978-5-9916-5009-0. – Режим доступа: www.biblio-online.ru/book/46A41F93-BC46-401C-A30E-27C0FB60B9DE.
3. Обработка и анализ данных / Университет ИТМО – Открытое образование. – Режим доступа: <https://openedu.ru/>
4. Элементы статистической обработки данных / Университет ИТМО – Открытое образование. – Режим доступа: <https://openedu.ru/>

8.2. Дополнительная литература

1. Неделько В. М. Основы статистических методов машинного обучения: учебное пособие / В. М. Неделько. – Новосибирск: Новосибирский государственный технический университет, 2010. – 72 с. – ISBN 978-5-7782-1385-2. – Текст: электронный // Электронно-библиотечная система IPR BOOKS: [сайт]. – URL: <https://www.iprbookshop.ru/45418.html>

8.3. Интернет-ресурсы, необходимые для освоения дисциплины (модуля)

1. Электронная библиотечная система издательства ЮРАЙТ, раздел «Легендарные книги». www.biblio-online.ru
2. Электронная библиотечная система IPRbooks. www.iprbookshop.ru

3. Обработка и анализ данных / Университет ИТМО – Открытое образование. – Режим доступа: <https://openedu.ru/>
4. Элементы статистической обработки данных / Университет ИТМО – Открытое образование. – Режим доступа: <https://openedu.ru/>
5. Введение в машинное обучение / Университет ИТМО – Открытое образование. – Режим доступа: <https://openedu.ru/>

9. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Для проведения лекционных занятий необходима мультимедийная аудитория, оснащенная компьютерными рабочими местами студентов.

Для проведения лабораторных занятий необходима аудитория, оснащенная компьютерными рабочими местами студентов и доступом в Интернет.

Для проведения публичной защиты творческих проектов, необходима мультимедийная аудитория с проектором.

Рабочая программа дисциплины (модуля) при необходимости может быть адаптирована для обучения (в том числе с применением дистанционных образовательных технологий) лиц с ограниченными возможностями здоровья, инвалидов. Для этого требуется заявление обучающихся, являющихся лицами с ограниченными возможностями здоровья, инвалидами, или их законных представителей и рекомендации психолого-медико-педагогической комиссии. Для инвалидов содержание рабочей программы дисциплины (модуля) может определяться также в соответствии с индивидуальной программой реабилитации инвалида (при наличии).

10. ОСОБЕННОСТИ РЕАЛИЗАЦИИ ДИСЦИПЛИНЫ (МОДУЛЯ) ПРИ ОБУЧЕНИИ ИНВАЛИДОВ И ЛИЦ С ОГРАНИЧЕННЫМИ ВОЗМОЖНОСТЯМИ ЗДОРОВЬЯ

Рабочая программа дисциплины (модуля) при необходимости может быть адаптирована для обучения (в том числе с применением дистанционных образовательных технологий) лиц с ограниченными возможностями здоровья, инвалидов. Для этого требуется заявление обучающихся, являющихся лицами с ограниченными возможностями здоровья, инвалидами, или их законных представителей и рекомендации психолого-медико-педагогической комиссии. При обучении лиц с ограниченными возможностями здоровья учитываются их индивидуальные психофизические особенности. Обучение инвалидов осуществляется также в соответствии с индивидуальной программой реабилитации инвалида (при наличии).

Для лиц с нарушением слуха возможно предоставление учебной информации в визуальной форме (краткий конспект лекций; тексты заданий, напечатанные увеличенным шрифтом), на аудиторных занятиях допускается присутствие ассистента, а также сурдопереводчиков и тифлосурдопереводчиков. Текущий контроль успеваемости осуществляется в письменной форме: обучающийся письменно отвечает на вопросы, письменно выполняет практические задания. Доклад (реферат) также может быть представлен в письменной форме, при этом требования к содержанию остаются теми же, а требования к качеству изложения материала (понятность, качество речи, взаимодействие с аудиторией и т. д.) заменяются на соответствующие требования, предъявляемые к письменным работам (качество оформления текста и списка литературы, грамотность, наличие иллюстрационных материалов и т. д.). Промежуточная аттестация для лиц с нарушениями слуха проводится в письменной форме, при этом используются общие критерии оценивания. При необходимости время подготовки к ответу может быть увеличено.

Для лиц с нарушением зрения допускается аудиальное предоставление информации, а также использование на аудиторных занятиях звукозаписывающих устройств (диктофонов и т. д.). Допускается присутствие на занятиях ассистента (помощника), оказывающего обучающимся необходимую техническую помощь. Текущий контроль успеваемости

осуществляется в устной форме. При проведении промежуточной аттестации для лиц с нарушением зрения тестирование может быть заменено на устное собеседование по вопросам.

Для лиц с ограниченными возможностями здоровья, имеющих нарушения опорно-двигательного аппарата, на аудиторных занятиях, а также при проведении процедур текущего контроля успеваемости и промежуточной аттестации могут быть предоставлены необходимые технические средства (персональный компьютер, ноутбук или другой гаджет); допускается присутствие ассистента (ассистентов), оказывающего обучающимся необходимую техническую помощь (занять рабочее место, передвигаться по аудитории, прочитать задание, оформить ответ, общаться с преподавателем).