

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«Тольяттинский государственный университет»

Б1.В.02
(индекс дисциплины)

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Теоретические и прикладные основы анализа данных

(наименование дисциплины)

по направлению подготовки
09.04.03 Прикладная информатика

направленность (профиль)
Прикладной анализ данных

Форма обучения: очная

Год набора: 2026

Общая трудоемкость: 5 ЗЕ

Распределение часов дисциплины по семестрам

Семестр	3	Итого
Форма контроля	Экзамен	
Вид занятий		
Лекции	16	16
Лабораторные		
Практические	16	16
Руководство: курсовые работы (проекты) / РГР		
Промежуточная аттестация	0,35	0,35
Контактная работа	32,35	32,35
Самостоятельная работа	112	112
Контроль	35,65	35,65
Итого	180	180

Рабочую программу составили:

Доцент института цифровых технологий, к.э.н., Раченко Т.А.

(должность, ученое звание, степень, Фамилия И.О.)

Рецензирование рабочей программы дисциплины:



Отсутствует



Рецензент

(должность, ученое звание, степень, Фамилия И.О.)

Рабочая программа дисциплины составлена на основании ФГОС ВО и учебного плана направления подготовки (специальности)

09.04.03 Прикладная информатика

Срок действия рабочей программы дисциплины до «31» августа 2028г.

УТВЕРЖДЕНО на заседании института цифровых технологий
(протокол заседания № 1 от «05» сентября 2025 г.)

1. Цель освоения дисциплины

Цель освоения дисциплины – формирование у обучающихся системных знаний и компетенций в области теоретических и прикладных основ анализа данных, включая методологии бизнес-анализа, стандарты, нотации и артефакты работы с требованиями, а также развитие практических навыков управления ИТ-проектами в условиях неопределённости с применением формальных инструментов управления рисками и проблемами проекта.

2. Место дисциплины в структуре ОПОП ВО

Дисциплина Б1.В.02 «Теоретические и прикладные основы анализа данных» относится к части, формируемой участниками образовательных отношений, блока 1 «Дисциплины (модули)» учебного плана направления подготовки 09.04.03 Прикладная информатика, направленность (профиль) «Прикладной анализ данных».

Освоение дисциплины базируется на знаниях, умениях и навыках, сформированных при изучении следующих предшествующих дисциплин и практик: Методы и инструментальные средства прикладного анализа данных (2 семестр); Производственная практика (научно-исследовательская работа) 1 (1 семестр); Производственная практика (научно-исследовательская работа) 2 (2 семестр).

Знания, умения и навыки, приобретённые в ходе изучения дисциплины, необходимы для успешного освоения следующих дисциплин и прохождения практик: Машинное обучение (4 семестр); Методы интерпретации и визуализации анализа больших данных (4 семестр); Производственная практика (научно-исследовательская работа) 3 (3 семестр); Производственная практика (научно-исследовательская работа) 4 (4 семестр); Производственная практика (преддипломная практика) (3 семестр) и Выполнение и защита выпускной квалификационной работы (4 семестр).

3. Планируемые результаты обучения

Формируемые и контролируемые компетенции (код и наименование)	Индикаторы достижения компетенций (код и наименование)	Планируемые результаты обучения
ПК-2 Способен управлять проектами в области ИТ в условиях неопределенностей с применением формальных инструментов управления рисками и проблемами проекта	ПК-2.1 Знает понятия проектного менеджмента в области ИТ и инструментальные средства	Знать: – основные понятия и методологии проектного менеджмента в ИТ; – классификацию требований к ИТ-проектам и их свойства; – инструментальные средства управления проектами (Jira, Trello, Microsoft Project и др.).
	ПК-2.2 Умеет управлять проектами в области ИТ с применением инструментальных средств	Уметь: – выбирать и обосновывать инструментальные средства для управления ИТ-проектами; – применять методы бизнес-анализа и работы с

Формируемые и контролируемые компетенции (код и наименование)	Индикаторы достижения компетенций (код и наименование)	Планируемые результаты обучения
		<p>требованиями в проектной деятельности;</p> <ul style="list-style-type: none"> – управлять проектами в условиях неопределённости, используя формальные инструменты управления рисками.
	<p>ПК-2.3 Владеет инструментарием проектного менеджмента в области ИТ</p>	<p>Владеть:</p> <ul style="list-style-type: none"> – навыками работы с инструментальными средствами проектного менеджмента (системы отслеживания задач, диаграммы Ганта, инструменты управления рисками); – приёмами формализации и документирования требований к ИТ-проектам; – методами идентификации, анализа и реагирования на проектные риски.

4. Структура и содержание дисциплины

Модуль	Вид учебной работы	Наименование тем занятий	Семестр	Объём, ч	Баллы (max)	Интерактив, ч	Формы текущего контроля
1. Основные положения и концепции работы с данными	лекция	Тема 1. Основы работы с данными	3	2	–	–	–
	лекция	Тема 2. Технологии сбора данных	3	2	–	–	–
	лекция	Тема 3. Методы работы с данными	3	2	–	–	–
	практ.	Практическая работа №1. Сбор, очистка и анонимизация образовательных данных	3	4	12	–	Отчёт по практической работе (защита)
	самост.	Подготовка к тестовым заданиям	3	32	–	–	–
2. Основные методы обработки и представления данных	лекция	Тема 4. Этапы решения задачи анализа данных и их взаимосвязи	3	2	–	–	–
	лекция	Тема 5. Системы аналитической обработки данных	3	2	–	–	–
	лекция	Тема 6. Этапы исследования данных	3	2	–	–	–
	лекция	Тема 7. Тестирование, интерпретация данных	3	2	–	–	–
	лекция	Тема 8. Использование результатов анализа данных	3	2	–	–	–

Модуль	Вид учебной работы	Наименование тем занятий	Семестр	Объём, ч	Баллы (max)	Интерактив, ч	Формы текущего контроля
	практ.	Практическая работа №2. Визуальный анализ и построение дашбордов для педагогической аналитики	3	4	12	—	Отчёт по практической работе (защита)
	практ.	Практическая работа №3. Построение модели прогнозирования успеваемости студентов с использованием машинного обучения	3	4	13	—	Отчёт по практической работе (защита)
	практ.	Практическая работа №4. Анализ поведенческих паттернов студентов в онлайн-курсах с применением кластеризации	3	4	13	—	Отчёт по практической работе (защита)
	самост.	Подготовка к тестовым заданиям	3	32	—	—	—
	пром. аттест.	Промежуточная аттестация (экзамен)	3	0,35	50	—	—
Итого				180	100	—	

Схема расчета итогового балла

- Текущий рейтинг (сумма баллов за отчёты по практическим работам) — максимум 50.
- Итоговый контроль (экзамен) — максимум 50.
- Итоговый балл = Текущий рейтинг + Результат экзамена (максимум 100).

Экзамен допускается при набранном рейтинге не менее 55 баллов (или по иному порогу, установленному ИЦТ).

5. Образовательные технологии

В рамках учебного курса предусмотрены следующие образовательные технологии:

- технология традиционного обучения: лекции, практические работы, самостоятельная работа;
- технология проектного обучения: реализация и защита отчетов по практическим работам.

Для студентов всех форм обучения предусмотрено получение консультационной помощи. Особое внимание необходимо уделить самостоятельному изучению нормативных источников и рекомендованной литературы.

В качестве текущего контроля при изучении курса предусмотрены защиты отчетов по практическим работам.

6. Методические указания по освоению дисциплины

Самостоятельная работа студентов (СРС) – работа с лекционным материалом, подготовка к практическим занятиям, изучение тем, вынесенных на самостоятельную проработку; подготовка к экзамену.

Самостоятельная работа студентов проводится с целью углубления и расширения теоретических знаний; развития познавательных способностей и активности студентов; самостоятельности, ответственности и организованности, творческой инициативы; формирования самостоятельности мышления, способности к саморазвитию, самосовершенствованию и самореализации.

Контроль выполненной самостоятельной работы осуществляется индивидуально, на занятии, при защите рефератов и творческих проектов.

6.1. Рекомендации по подготовке к лекционным занятиям

Изучение дисциплины требует систематического и последовательного накопления знаний, следовательно, пропуски отдельных тем не позволяют глубоко освоить предмет.

В ходе лекционных занятий задавать преподавателю уточняющие вопросы с целью уяснения теоретических положений, разрешения спорных ситуаций.

Студент может дополнить список использованной литературы современными источниками, не представленными в списке рекомендованной литературы, и в дальнейшем использовать собственные подготовленные учебные материалы при написании курсовых и выпускных квалификационных работ.

6.2. Рекомендации по подготовке к практическим занятиям

Студентам следует:

- до очередного практического занятия по рекомендованным литературным источникам проработать теоретический материал, соответствующей темы занятия;
- при подготовке к практическим занятиям следует обязательно использовать не только лекции, учебную литературу, но и другие источники;
- в начале занятий задать преподавателю вопросы по материалу, вызвавшему затруднения в его понимании и освоении при решении задач, заданных для самостоятельного решения;
- на занятии доводить каждую задачу до окончательного решения, демонстрировать понимание проведенных расчетов (анализов, ситуаций), в случае затруднений обращаться к преподавателю.

Для того чтобы практические занятия приносили максимальную пользу, необходимо помнить, что упражнение и решение задач проводятся по рассмотренному на лекциях материалу и связаны, как правило, с детальным разбором отдельных вопросов лекционного курса. Следует подчеркнуть, что только после усвоения лекционного материала с

определенной точки зрения (а именно с той, с которой он излагается на лекциях) он будет закрепляться студентом на практических занятиях как в результате обсуждения и анализа лекционного материала, так и с помощью решения проблемных ситуаций, задач. При этих условиях студент не только хорошо усвоит материал, но и научится применять его на практике, а также получит дополнительный стимул (и это очень важно) для активной проработки лекции.

При самостоятельном решении задач нужно обосновывать каждый этап решения, исходя из теоретических положений курса. Если студент видит несколько путей решения проблемы (задачи), то нужно сравнить их и выбрать самый рациональный. Полезно до начала вычислений составить краткий план решения проблемы (задачи). Решение проблемных задач или примеров следует излагать подробно, вычисления располагать в строгом порядке, отделяя вспомогательные вычисления от основных. Решения при необходимости нужно сопровождать комментариями, схемами, чертежами и рисунками.

Следует помнить, что решение каждой учебной задачи должно доводиться до окончательного логического ответа, которого требует условие, и по возможности с выводом. Полученный ответ следует проверить способами, вытекающими из существа данной задачи. Полезно также (если возможно) решать несколькими способами и сравнить полученные результаты. Решение задач данного типа нужно продолжать до приобретения твердых навыков в их решении.

6.3. Рекомендации по подготовке к экзамену

Подготовка к экзамену способствует закреплению, углублению и обобщению знаний, получаемых, в процессе обучения, а также применению их к решению практических задач. Готовясь к зачету, студент ликвидирует имеющиеся пробелы в знаниях, углубляет, систематизирует и упорядочивает свои знания. На зачете студент демонстрирует то, что он приобрел в процессе обучения по учебной дисциплине.

Необходимо ориентировать студентов на систематическую подготовку к занятиям в течение семестра, что позволит использовать время экзаменационной сессии для систематизации знаний.

7. Оценочные средства

7.1. Паспорт оценочных средств

Семестр	Код контролируемой компетенции (или ее части)	Наименование оценочного средства
3	ПК-2 (способен управлять проектами в области ИТ в условиях неопределенностей с применением формальных инструментов управления рисками и проблемами проекта)	– Отчёты по практическим работам №1–4 (текущий контроль) – Типовые задания закрытого и открытого типа для проверки сформированности компетенции (№1–300) – Вопросы к экзамену (промежуточная аттестация)

7.2. Типовые задания или иные материалы, необходимые для текущего контроля

7.2.1. Комплект отчетов по заданиям, выполненным на практических занятиях

(наименование оценочного средства)

Типовой пример практических работ

Практическая работа №1

Тема: сбор, очистка и анонимизация образовательных данных: построение «цифрового следа» студента

Цель: освоить этапы подготовки данных для анализа в образовательной среде с соблюдением принципов информационной безопасности и требований GDPR/ФЗ-152.

Задачи:

- Сгенерировать синтетический датасет, имитирующий данные студентов (демография, оценки, посещаемость, активность в LMS).
- Выполнить предварительный анализ: выявить пропуски, выбросы, дубликаты.
- Применить методы анонимизации: удаление прямых идентификаторов, псевдонимизация, обобщение возрастных групп, добавление шума к числовым показателям.
- Оформить отчёт с обоснованием выбора методов с точки зрения этики и безопасности.

Формируемые компетенции: способность работать с персональными данными, соблюдать законодательство о защите данных, проводить первичную обработку данных.

Планируемые результаты:

- Знать основные этапы ETL (Extract, Transform, Load) применительно к образовательным данным.

- Уметь генерировать синтетические данные, выявлять и устранять аномалии, применять методы анонимизации.

- Владеть библиотеками pandas, numpy, faker.

Инструменты: Python (pandas, numpy, faker, random), Jupyter Notebook.

Краткое содержание (этапы):

1. Генерация датасета студентов (1000 записей) и их академических показателей (оценки, посещаемость, активность).
2. Объединение данных.
3. Проверка на пропуски и дубликаты, анализ выбросов (IQR, boxplot).
4. Анонимизация: удаление ФИО, псевдонимизация student_id, группировка возраста, добавление шума к числовым показателям.
5. Сохранение очищенного и анонимизированного датасета.

Форма отчёта: отчёт с описанием этапов, скриншотами кода и результатами анализа, выводами о применённых методах анонимизации.

Практическая работа №2

Тема: визуальный анализ и построение дашбордов для педагогической аналитики

Цель: научиться визуализировать образовательные данные для поддержки принятия решений преподавателями и администрацией вуза.

Задачи:

- Построить статические визуализации (распределение оценок, динамика посещаемости, корреляция активности и успеваемости) средствами Python.
- Разработать интерактивный дашборд в Power BI / Tableau с фильтрами по группам, курсам, семестрам.
- Добавить элементы раннего предупреждения (early warning): условное форматирование, KPI, индикаторы риска отчисления.
- Подготовить отчёт с интерпретацией полученных графиков и дашборда.

Формируемые компетенции: способность визуализировать данные, интерпретировать результаты анализа, создавать инструменты мониторинга учебного процесса.

Планируемые результаты:

- Знать основные виды графиков для образовательной аналитики.

- Уметь строить гистограммы, диаграммы рассеяния, boxplot, использовать условное форматирование.

- Владеть инструментами Power BI (или Tableau) для создания дашбордов.

Инструменты: Python (matplotlib, seaborn), Power BI Desktop (или Tableau).

Краткое содержание (этапы):

1. Загрузка анонимизированного датасета из ПР №1.
2. Построение в Python:
 - гистограмма итоговых баллов с порогами риска;
 - динамика посещаемости по семестрам;
 - scatter-plot «активность – успеваемость» с цветовой маркировкой по уровню риска.
3. Сохранение данных в формате CSV для Power BI.
4. Создание дашборда:
 - загрузка данных, создание мер;
 - визуализации (гистограмма, boxplot, точечная диаграмма);
 - фильтры по группе, направлению, семестру, возрасту;
 - элементы early warning (условное форматирование таблицы, KPI с цветовой индикацией).
5. Подготовка отчёта с анализом визуализаций и сценариями использования дашборда.

Форма отчёта: файл дашборда (PBIX) и текстовый отчёт с описанием функциональности и выводами по анализу.

Практическая работа №3

Тема: построение модели прогнозирования успеваемости студентов с использованием машинного обучения

Цель: освоить применение алгоритмов машинного обучения для прогнозирования академической успеваемости.

Задачи:

- Выделить целевую переменную (бинарную: риск отчисления / успеваемость, или непрерывную: итоговый балл).

- Обучить и сравнить 2–3 модели (логистическая регрессия, дерево решений, случайный лес).
- Оценить качество моделей по метрикам: accuracy, precision, recall, F1-score, AUC-ROC.
- Проанализировать важность признаков.
- Подготовить отчёт с обоснованием выбора модели и интерпретацией результатов.

Формируемые компетенции: способность применять методы машинного обучения в образовательных задачах, оценивать и интерпретировать качество моделей.

Планируемые результаты:

- Знать метрики классификации, принципы работы базовых алгоритмов.
- Уметь проводить предобработку данных для ML, разделять выборку, обучать модели, оценивать их качество.
- Владеть библиотеками scikit-learn, pandas.

Инструменты: Python (scikit-learn, pandas, numpy), Jupyter Notebook.

Краткое содержание (этапы):

1. Загрузка и подготовка датасета (из ПР №1).
2. Создание целевой переменной: «успевающий» (итоговый балл ≥ 60) vs «риск».
3. Разделение выборки на train/test.
4. Обучение моделей:
 - логистическая регрессия;
 - дерево решений;
 - случайный лес.
5. Оценка качества (confusion matrix, accuracy, precision, recall, F1, ROC-AUC).
6. Анализ важности признаков (feature importance).
7. Формулирование выводов о наиболее значимых факторах успеваемости.

Форма отчёта: Jupyter Notebook с кодом, визуализациями, таблицами метрик и выводами.

Практическая работа №4

Тема: анализ поведенческих паттернов студентов в онлайн-курсах с применением кластеризации

Цель: выявить типы обучающихся на основе их цифрового поведения для адаптации образовательных траекторий.

Задачи:

- Сформировать набор поведенческих признаков (среднее время сессии, количество посещений в неделю, доля завершённых модулей, задержка сдачи заданий).
- Применить алгоритмы кластеризации: K-means, иерархическая кластеризация, DBSCAN.
- Оценить качество кластеризации (силуэт, интерпретируемость).
- Дать педагогическую интерпретацию полученным кластерам.
- Предложить стратегии адаптации контента или обратной связи для каждого кластера.

Формируемые компетенции: способность сегментировать обучающихся на основе цифрового следа, интерпретировать результаты кластеризации в педагогическом контексте.

Планируемые результаты:

- Знать основные алгоритмы кластеризации и методы выбора числа кластеров.
- Уметь проводить feature engineering, нормализацию, визуализировать кластеры (PCA, t-SNE, дендрограммы).
- Владеть библиотеками scikit-learn, scipy, matplotlib.

Инструменты: Python (scikit-learn, scipy, pandas, matplotlib, seaborn).

Краткое содержание (этапы):

1. Генерация или загрузка логов активности студентов (500–1000 записей).
2. Feature engineering: расчёт поведенческих метрик.
3. Стандартизация признаков.
4. Определение оптимального числа кластеров для K-means (метод локтя, силуэт).
5. Кластеризация:
 - K-means с оптимальным K;
 - иерархическая кластеризация (построение дендрограммы, выбор порога);

- DBSCAN (подбор параметров `eps`, `min_samples`).

6. Сравнение результатов, выбор наилучшего метода по силуэту и интерпретируемости.

7. Визуализация кластеров в пространстве главных компонент (РСА).

8. Анализ средних значений признаков по кластерам, присвоение педагогических ярлыков (активные, прокрастинаторы, выборочные, отстающие).

9. Разработка рекомендаций по адаптации контента и педагогической поддержки для каждого типа.

Форма отчёта: Jupyter Notebook с кодом, графиками (локоть, силуэт, дендрограмма, РСА-проекция, профили кластеров), таблицей характеристик кластеров и подробными рекомендациями.

Общие рекомендации по выполнению практических работ

- Все работы выполняются в среде Python (Jupyter Notebook, Google Colab) с использованием указанных библиотек.

- Для работы №2 допускается использование Power BI Desktop (или Tableau) для создания дашборда.

- Данные на всех этапах являются синтетическими (генерируются) или обезличенными, что соответствует принципам информационной безопасности.

- По каждой работе оформляется отчёт в соответствии с установленными требованиями (пояснительная записка, скриншоты, код, выводы).

Критерии оценивания практических работ

Каждая практическая работа оценивается по совокупности представленного отчёта и его защиты. Максимальная сумма баллов за выполнение и защиту **всех** практических работ составляет **50** (распределение баллов между работами может быть равномерным или дифференцированным в зависимости от сложности).

Балльное распределение

Практическая работа	Тема	Максимальный балл
№1	Сбор, очистка и анонимизация образовательных данных	12
№2	Визуальный анализ и построение дашбордов для педагогической аналитики	12
№3	Построение модели прогнозирования успеваемости студентов с использованием машинного обучения	13

Практическая работа	Тема	Максимальный балл
№4	Анализ поведенческих паттернов студентов в онлайн-курсах с применением кластеризации	13

Оценка каждой работы производится на основе следующих критериев (сумма баллов по критериям не должна превышать максимальный балл, установленный для работы). Баллы выставляются целыми числами.

№	Критерий	Пояснение	Максимальный балл по критерию (для работы)
1	Наличие всей существенной информации	Отчёт содержит все этапы: цель, описание данных, код (или ссылка на Notebook), результаты, визуализации, выводы. Для ПР №2 обязательно прилагается файл дашборда.	2
2	Точность и полнота предоставляемых сведений	Все числовые показатели, метрики, графики верны. Отсутствуют пропуски обязательных элементов (обоснование метода анонимизации в ПР №1, метрики качества моделей в ПР №3, интерпретация кластеров в ПР №4).	2
3	Непротиворечивость информации	Данные, визуализации и выводы логически согласованы.	1
4	Правильность интерпретаций и выводов	Выводы обоснованы результатами, отражают прикладной смысл работы (этика, сценарии использования, факторы успеваемости, педагогические рекомендации).	2
5	Степень достижения цели	Реализованы все задачи, предусмотренные описанием работы, получены ожидаемые результаты (очищенный датасет, дашборд, обученные модели, интерпретируемые кластеры).	2
6	Обоснованность применяемого решения	Выбор методов (способ очистки, алгоритм, визуализация) пояснён со ссылкой на цели и свойства данных.	2
7	Грамотность используемых формулировок	Текст отчёта логически связан, терминология корректна.	1

Требования к оформлению отчётов

1. Отчёт по каждой практической работе выполняется в цифровом виде и сдаётся через электронную среду (например, Jupyter Notebook с пояснениями, PDF-документ, ссылка на Google Colab).
2. Для работ, связанных с дашбордами (**ПР №2**), вместе с отчётом предоставляется файл дашборда (PBIX, Tableau Workbook) или ссылка на опубликованную версию.
3. Для работ с кодом (**ПР №1, №3, №4**) отчёт может быть оформлен в виде Jupyter Notebook с ячейками markdown, содержащими пояснения, и ячейками кода.
4. При оформлении текстового документа (PDF, Word) необходимо придерживаться следующих правил:
 - формат страницы – А4.
 - каждый структурный элемент (теоретическая часть, практическая часть, приложения) начинается с новой страницы.
 - в заголовках не допускаются переносы слов.
 - все таблицы, рисунки, листинги кода должны быть выполнены в соответствии с требованиями действующих стандартов (ГОСТ, либо внутренние требования кафедры).
5. При использовании внешних источников данных обязательно указывается их происхождение (ссылка, описание генерации).

Типовые задания для оценки сформированности компетенций

ПК-2. Способен управлять проектами в области ИТ в условиях неопределенностей с применением формальных инструментов управления рисками и проблемами проекта

код и наименование компетенции

Задания на проверку сформированности компетенции ПК-2 (1–50)

Задания разделены на закрытый (выбор ответа) и открытый (развёрнутый ответ) типы, охватывают все модули дисциплины и могут использоваться для текущего контроля и подготовки к экзамену. Все формулировки приведены на русском языке.

Задания закрытого типа (1–25)

Задание 1.

Выберите несколько правильных вариантов ответа.

Какие из перечисленных методов относятся к управлению рисками проекта?

- а) идентификация рисков
- б) качественный и количественный анализ рисков
- в) разработка пользовательского интерфейса
- г) планирование реагирования на риски

Правильный ответ: а, б, г

Задание 2.

Выберите несколько правильных вариантов ответа.

Какие инструментальные средства управления проектами в области ИТ?

- а) Jira
- б) Trello
- в) Microsoft Excel
- г) Microsoft Project

Правильный ответ: а, б, г

Задание 3.

Выберите один правильный вариант ответа.

Что такое ETL-процесс в контексте подготовки данных?

- а) этап визуализации данных

- б) процесс извлечения, преобразования и загрузки данных
- в) метод машинного обучения
- г) способ анонимизации данных

Правильный ответ: б

Задание 4.

Выберите несколько правильных вариантов ответа.

Какие виды шкал измерения данных существуют?

- а) номинальная
- б) порядковая
- в) интервальная
- г) абсолютная

Правильный ответ: а, б, в

Задание 5.

Выберите один правильный вариант ответа.

Какой метод используется для обнаружения выбросов в данных?

- а) метод главных компонент (PCA)
- б) межквартильный размах (IQR)
- в) логистическая регрессия
- г) one-hot encoding

Правильный ответ: б

Задание 6.

Выберите несколько правильных вариантов ответа.

Какие библиотеки Python используются для визуализации данных?

- а) matplotlib
- б) seaborn
- в) pandas
- г) plotly

Правильный ответ: а, б, г

Задание 7.

Выберите один правильный вариант ответа.

Что такое дашборд (dashboard)?

- а) программа для редактирования текста
- б) инструмент для мониторинга ключевых показателей в реальном времени
- в) база данных
- г) алгоритм машинного обучения

Правильный ответ: б

Задание 8.

Выберите несколько правильных вариантов ответа.

Какие методы относятся к ансамблевым алгоритмам машинного обучения?

- а) бэггинг (Bagging)
- б) бустинг (Boosting)
- в) метод k-ближайших соседей
- г) стекинг (Stacking)

Правильный ответ: а, б, г

Задание 9.

Выберите один правильный вариант ответа.

Для чего используется метод главных компонент (PCA)?

- а) для классификации объектов
- б) для уменьшения размерности данных
- в) для кластеризации
- г) для поиска ассоциативных правил

Правильный ответ: б

Задание 10.

Выберите несколько правильных вариантов ответа.

Какие метрики используются для оценки качества моделей классификации?

- а) точность (accuracy)
- б) полнота (recall)
- в) среднеквадратичная ошибка (MSE)
- г) F1-мера

Правильный ответ: а, б, г

Задание 11.

Выберите один правильный вариант ответа.

Что такое переобучение (overfitting) модели?

- а) модель слишком хорошо обобщает новые данные
- б) модель слишком точно подстроилась под обучающие данные и плохо обобщает
- в) модель не может обучиться на обучающих данных
- г) модель показывает одинаковые результаты на обучении и тесте

Правильный ответ: б

Задание 12.

Выберите несколько правильных вариантов ответа.

Какие методы относятся к методам обработки пропущенных значений?

- а) удаление строк с пропусками
- б) заполнение средним или медианой
- в) применение one-hot encoding
- г) интерполяция

Правильный ответ: а, б, г

Задание 13.

Выберите один правильный вариант ответа.

Что такое нормализация данных?

- а) приведение данных к единому масштабу, например, к диапазону [0,1]
- б) удаление дубликатов
- в) кодирование категориальных переменных
- г) добавление шума к числовым показателям

Правильный ответ: а

Задание 14.

Выберите несколько правильных вариантов ответа.

Какие алгоритмы относятся к методам кластеризации?

- а) K-means
- б) иерархическая кластеризация
- в) DBSCAN
- г) логистическая регрессия

Правильный ответ: а, б, в

Задание 15.

Выберите один правильный вариант ответа.

Что такое SQL и где он используется?

- а) язык программирования для создания мобильных приложений
- б) язык структурированных запросов для работы с реляционными базами данных
- в) библиотека для визуализации данных
- г) система управления проектами

Правильный ответ: б

Задание 16.

Выберите несколько правильных вариантов ответа.

Какие виды диаграмм наиболее подходят для визуализации распределения одной переменной?

- а) гистограмма
- б) диаграмма рассеяния (scatter plot)
- в) ящик с усами (box plot)
- г) тепловая карта (heatmap)

Правильный ответ: а, в

Задание 17.

Выберите один правильный вариант ответа.

Для чего используется метод кодирования one-hot encoding?

- а) для масштабирования числовых признаков
- б) для преобразования категориальных переменных в бинарные векторы
- в) для удаления выбросов
- г) для заполнения пропусков

Правильный ответ: б

Задание 18.

Выберите несколько правильных вариантов ответа.

Какие факторы влияют на выбор программного обеспечения для управления проектами?

- а) потребности проекта

- б) требования безопасности
- в) наличие современной оргтехники
- г) высокоскоростной интернет

Правильный ответ: а, б

Задание 19.

Выберите один правильный вариант ответа.

Какая метрика используется для оценки качества регрессионных моделей?

- а) F1-мера
- б) средняя абсолютная ошибка (MAE)
- в) площадь под ROC-кривой (AUC-ROC)
- г) точность (ассигасу)

Правильный ответ: б

Задание 20.

Выберите несколько правильных вариантов ответа.

Какие методы анонимизации персональных данных вы знаете?

- а) удаление прямых идентификаторов
- б) псевдонимизация
- в) добавление шума к числовым показателям
- г) шифрование без возможности восстановления

Правильный ответ: а, б, в, г

Задание 21.

Выберите один правильный вариант ответа.

Что такое Agile в контексте управления ИТ-проектами?

- а) методология управления проектами, основанная на итеративном подходе
- б) инструмент для построения диаграмм Ганта
- в) база данных для хранения требований
- г) язык программирования

Правильный ответ: а

Задание 22.

Выберите несколько правильных вариантов ответа.

Какие этапы включает процесс CRISP-DM?

- а) бизнес-понимание
- б) подготовка данных
- в) разработка интерфейса
- г) оценка модели

Правильный ответ: а, б, г

Задание 23.

Выберите один правильный вариант ответа.

Что такое диаграмма Ганта?

- а) график, отображающий распределение частот
- б) инструмент для визуализации сроков выполнения задач проекта
- в) метод кластеризации
- г) способ оценки качества модели

Правильный ответ: б

Задание 24.

Выберите несколько правильных вариантов ответа.

Какие преимущества управления программными проектами можно выделить?

- а) сокращение общих расходов за счёт оптимизации операций
- б) улучшение коммуникации между заинтересованными сторонами
- в) снижение рисков, негативно влияющих на организацию
- г) автоматическое написание кода

Правильный ответ: а, б, в

Задание 25.

Выберите один правильный вариант ответа.

Что такое кросс-валидация (cross-validation)?

- а) метод визуализации данных
- б) способ оценки качества модели путём разбиения данных на несколько частей
- в) алгоритм кластеризации
- г) метод анонимизации данных

Правильный ответ: б

Задания открытого типа (26–50)

Задание 26.

Дайте развернутый ответ.

Что представляет собой понятие «проектный менеджмент в области информационных технологий (ИТ)»?

Правильный ответ:

Проектный менеджмент в области информационных технологий — это процесс управления проектами, связанными с разработкой, внедрением и поддержкой информационных систем и технологий. Он включает планирование, организацию, контроль ресурсов, управление рисками и коммуникациями для достижения целей проекта в установленные сроки и бюджет.

Задание 27.

Дайте развернутый ответ.

Что понимают под инструментами проектного менеджмента? Приведите примеры.

Правильный ответ:

Инструменты проектного менеджмента — это программные средства, помогающие управлять проектом на всех этапах жизненного цикла. Примеры: Microsoft Project (планирование, диаграммы Ганта), Jira (отслеживание задач, гибкие методологии), Trello (канбан-доски), Slack (коммуникации), Redmine, Asana.

Задание 28.

Дайте развернутый ответ.

Дайте характеристику инструмента Jira.

Правильный ответ:

Jira — система управления проектами с открытым исходным кодом (прим.: проприетарная, но имеется открытая версия для небольших команд), которая позволяет создавать задачи, управлять ими, отслеживать прогресс, настраивать рабочие процессы, вести спринты и общаться с командой проекта. Широко используется в гибких методологиях (Scrum, Kanban).

Задание 29.

Дайте развернутый ответ.

Какие методы управления рисками проекта вы знаете? Приведите примеры их применения в ИТ-проектах.

Правильный ответ:

Методы включают идентификацию рисков (мозговой штурм, анализ документации), качественный анализ (матрица вероятности и последствий), количественный анализ (моделирование Монте-Карло), планирование реагирования (уклонение, передача, снижение, принятие) и мониторинг. Пример: риск срыва сроков в ИТ-проекте можно снизить, используя гибкие методологии и резервирование времени.

Задание 30.

Дайте развернутый ответ.

Что такое «консолидация данных»? Какие задачи она решает?

Правильный ответ:

Консолидация данных — это комплекс методов и процедур, направленных на извлечение данных из различных источников, обеспечение необходимого уровня их информативности и качества, преобразования в единый формат, в котором они могут быть загружены в хранилище данных или аналитическую систему. Решает проблемы фрагментарности, противоречивости и неоднородности данных.

Задание 31.

Дайте развернутый ответ.

Опишите основные этапы решения задачи анализа данных.

Правильный ответ:

1. Постановка задачи и бизнес-понимание.
2. Сбор данных из внутренних и внешних источников.
3. Предобработка данных (очистка, нормализация, кодирование, работа с пропусками).
4. Разведочный анализ данных (EDA) и визуализация.
5. Выбор и обучение модели (если применимо).
6. Оценка качества модели.

7. Интерпретация результатов и внедрение.

Задание 32.

Дайте развернутый ответ.

В чём различие между нормализацией и стандартизацией данных? Приведите примеры.

Правильный ответ:

Нормализация приводит данные к определённому диапазону, обычно $[0,1]$ или $[-1,1]$, например, Min-Max scaling. Стандартизация приводит данные к нулевому среднему и единичному дисперсии (Z-масштабирование). Нормализация используется, когда нужно сохранить отношение к границам, стандартизация — когда важно сравнивать признаки с разными единицами измерения.

Задание 33.

Дайте развернутый ответ.

Что такое «хранилище данных»? Каковы его основные характеристики?

Правильный ответ:

Хранилище данных — это предметно-ориентированная, интегрированная, неизменяемая во времени и поддерживающая исторические данные совокупность, предназначенная для поддержки принятия управленческих решений. Основные характеристики: интеграция данных из разных источников, поддержка временных срезов, ориентация на аналитические запросы.

Задание 34.

Дайте развернутый ответ.

Какие методы кодирования категориальных переменных вы знаете? В чём их особенности?

Правильный ответ:

One-hot encoding — создание бинарных столбцов для каждой категории (подходит для номинальных признаков). Label encoding — присвоение числовых меток категориям (подходит для порядковых признаков). Целевое кодирование (mean encoding) — замена категории на среднее значение целевой переменной.

Задание 35.

Дайте развернутый ответ.

Объясните назначение диаграммы «ящик с усами» (box plot). Как интерпретировать её элементы?

Правильный ответ:

Box plot используется для визуализации распределения числовых данных и выявления выбросов. Элементы: медиана (линия внутри ящика), межквартильный размах (IQR) — высота ящика, усы — границы нормальных значений (обычно $Q1 - 1,5 * IQR$ и $Q3 + 1,5 * IQR$), точки за усами — выбросы.

Задание 36.

Дайте развернутый ответ.

Что такое «задача классификации»? Приведите примеры из реальной жизни.

Правильный ответ:

Классификация — задача отнесения объекта к одному из заранее известных классов на основе его признаков. Примеры: определение спама в email, прогнозирование оттока клиентов, диагностика заболеваний по симптомам.

Задание 37.

Дайте развернутый ответ.

Опишите алгоритм k-means. Как выбрать количество кластеров?

Правильный ответ:

K-means — алгоритм кластеризации, который разбивает данные на K кластеров, минимизируя сумму квадратов расстояний от точек до центроидов кластеров. Выбор K осуществляется с помощью метода локтя (зависимость суммы квадратов от K) или силуэтного коэффициента.

Задание 38.

Дайте развернутый ответ.

Что такое «переобучение» модели? Как его обнаружить и предотвратить?

Правильный ответ:

Переобучение — ситуация, когда модель слишком точно подстраивается под обучающие данные и теряет способность обобщать. Обнаруживается по большой разнице между

метриками на обучении и на валидации. Предотвращается: регуляризацией, кросс-валидацией, уменьшением сложности модели, увеличением данных, ранней остановкой.

Задание 39.

Дайте развернутый ответ.

Какие возможности предоставляет библиотека pandas в Python?

Правильный ответ:

Pandas предоставляет структуры данных (Series, DataFrame) для работы с табличными данными, функции чтения/записи данных в различных форматах (CSV, Excel, SQL), инструменты для очистки, фильтрации, группировки, объединения и преобразования данных, а также базовую визуализацию.

Задание 40.

Дайте развернутый ответ.

Что такое «воспроизводимый анализ» (reproducible research)? Какие практики его обеспечивают?

Правильный ответ:

Воспроизводимый анализ — это возможность независимого повторения исследования и получения тех же результатов. Практики: использование Jupyter Notebook, версионный контроль (Git), фиксация случайных seed, документирование кода и данных, автоматизация пайплайнов (Airflow), контейнеризация (Docker).

Задание 41.

Дайте развернутый ответ.

Какие методы интерпретации моделей «чёрного ящика» вы знаете?

Правильный ответ:

SHAP (Shapley Additive Explanations) — разложение предсказания по вкладу каждого признака. LIME (Local Interpretable Model-agnostic Explanations) — локальные аппроксимации. PDP (Partial Dependence Plots) — зависимость предсказания от одного признака. ICE (Individual Conditional Expectation) — индивидуальные кривые.

Задание 42.

Дайте развернутый ответ.

Как оценить практическую значимость полученной модели, а не только статистическую точность?

Правильный ответ:

Необходимо оценить бизнес-метрики (например, ROI, снижение рисков), провести A/B-тестирование на реальных данных, оценить стоимость ошибок (confusion matrix с весами), интерпретируемость модели, удобство внедрения и поддержки.

Задание 43.

Дайте развернутый ответ.

Что такое «АВ-тестирование»? Как результаты анализа данных используются в нём?

Правильный ответ:

А/В-тестирование — эксперимент, в котором сравниваются две версии (А и В) для выявления статистически значимых различий в метриках. Результаты анализа данных (статистические тесты, доверительные интервалы, р-значения) используются для принятия решения о внедрении изменений.

Задание 44.

Дайте развернутый ответ.

Какие этические проблемы могут возникать при анализе данных?

Правильный ответ:

Нарушение приватности, дискриминация и предвзятость моделей (bias), отсутствие согласия на сбор данных, непрозрачность алгоритмов, возможность манипуляции поведением пользователей. Требуется соблюдение законодательства (GDPR, ФЗ-152), анонимизация, fairness-метрики.

Задание 45.

Дайте развернутый ответ.

Как проверить, что модель продолжает работать корректно после внедрения (мониторинг дрейфа данных)?

Правильный ответ:

Необходимо отслеживать дрейф данных (data drift) и дрейф концепта (concept drift) с помощью статистических тестов (например, PSI — Population Stability Index), сравнивать

распределения входных признаков и качество прогнозов на новых данных, использовать автоматические алерты при отклонениях.

Задание 46.

Дайте развернутый ответ.

Что такое «проектный треугольник» в управлении ИТ-проектами?

Правильный ответ:

Проектный треугольник — модель, описывающая ограничения проекта: объём (scope), время (time), стоимость (cost). Изменение одного ограничения влияет на другие. Качество часто считается четвёртым элементом. В ИТ-проектах управление этими тремя параметрами критически важно для успеха.

Задание 47.

Дайте развернутый ответ.

Опишите разницу между гибкими (Agile) и традиционными (Waterfall) методологиями управления проектами.

Правильный ответ:

Waterfall — последовательный подход, где каждый этап завершается до начала следующего; подходит для проектов с чёткими требованиями. Agile — итеративный подход с короткими циклами (спринтами), частой обратной связью, адаптацией к изменениям; подходит для проектов с высокой неопределённостью.

Задание 48.

Дайте развернутый ответ.

Какие способы оценки качества моделей машинного обучения вы знаете?

Правильный ответ:

Для классификации: accuracy, precision, recall, F1-score, AUC-ROC. Для регрессии: MAE, MSE, RMSE, R^2 . Кросс-валидация (k-fold) позволяет более надёжно оценить обобщающую способность. Для несбалансированных классов используют специальные метрики (F1, AUC).

Задание 49.

Дайте развернутый ответ.

Что такое «риск-менеджмент» в ИТ-проектах? Опишите основные этапы.

Правильный ответ:

Риск-менеджмент — процесс идентификации, анализа, планирования реагирования и мониторинга рисков. Этапы: 1) идентификация рисков, 2) качественный анализ (оценка вероятности и влияния), 3) количественный анализ (при необходимости), 4) планирование реагирования, 5) мониторинг и контроль.

Задание 50.

Дайте развернутый ответ.

Как составить отчёт по результатам аналитического исследования? Какие разделы он должен включать?

Правильный ответ:

Отчёт должен включать: введение (цель, задачи), описание данных и источников, методы анализа, результаты (визуализации, таблицы), интерпретацию результатов, выводы и рекомендации, ограничения исследования, приложения (код, дополнительные таблицы).

7.3. Оценочные средства для промежуточной аттестации по итогам освоения дисциплины

7.3.1. Вопросы к промежуточной аттестации (экзамену)

Семестр 3

Модуль 1. Основы работы с данными: сбор, предобработка, визуализация

1. Что такое анализ данных? Какие задачи решает анализ данных в современном бизнесе и науке?
2. Перечислите основные этапы решения задачи анализа данных. Опишите их содержание.
3. Какие типы данных (структурированные, неструктурированные, полуструктурированные) существуют? Приведите примеры.
4. Что такое «источники данных»? Какие бывают внутренние и внешние источники?
5. Объясните понятия «данные», «информация», «знания». В чём их различие?
6. Что такое «хранилище данных»? Каковы его основные характеристики?
7. Что представляет собой процесс ETL (Extract, Transform, Load)? Какие задачи решает?
8. Какие методы сбора данных вы знаете? В чём преимущества и недостатки опросов, логирования, веб-скрапинга?
9. Что такое «первичные данные» и «вторичные данные»? Приведите примеры использования.
10. Какие виды шкал измерения данных (номинальная, порядковая, интервальная, относительная) существуют? Приведите примеры.
11. Что такое «очистка данных» (data cleaning)? Перечислите основные задачи очистки.
12. Какие методы работы с пропущенными значениями вы знаете? В каких случаях какой метод предпочтительнее?
13. Как обнаружить и обработать выбросы (аномалии) в данных?

14. Что такое «нормализация» и «стандартизация» данных? В чём их отличие и когда применяются?
15. Как бороться с дубликатами в наборе данных?
16. Что такое «дискретизация» данных? Приведите примеры.
17. Какие методы кодирования категориальных переменных вы знаете (one-hot encoding, label encoding и др.)? В чём их особенности?
18. Что такое «консолидация данных»? Какие проблемы решает?
19. Какие инструменты (программные) используются для подготовки данных? Назовите и кратко охарактеризуйте.
20. Как провести первичный разведочный анализ данных (EDA)? Какие методы и визуализации при этом используются?
21. Какие виды диаграмм используются для визуализации распределения одной переменной? Приведите примеры.
22. Какие графики применяются для визуализации взаимосвязи двух числовых переменных?
23. Что такое «гистограмма», «полигон частот», «кумулятивный график»? Для каких целей они используются?
24. Объясните назначение диаграммы «ящик с усами» (box plot). Как интерпретировать её элементы?
25. Какие методы визуализации подходят для сравнения категориальных данных?
26. Что такое «тепловая карта» (heatmap) и когда она применяется?
27. Как визуализировать многомерные данные? Приведите примеры (scatter plot matrix, параллельные координаты).
28. Какие библиотеки визуализации в Python вы знаете? Назовите основные возможности matplotlib, seaborn, plotly.
29. Какие правила эффективной визуализации данных вы знаете? Как избежать искажения информации?
- Модуль 2. Методы моделирования, инструментальные средства и интерпретация результатов**
30. Что такое «задача классификации»? Приведите примеры из реальной жизни.
31. Какие алгоритмы классификации вы знаете? Кратко опишите принцип работы логистической регрессии, деревьев решений, случайного леса.
32. Что такое «задача кластеризации»? В чём отличие от классификации?
33. Опишите алгоритм k-means. Как выбрать количество кластеров?
34. Что такое «задача регрессии»? Приведите примеры. Какие метрики используются для оценки качества регрессионных моделей?
35. Какие методы регрессии вы знаете? В чём отличие линейной регрессии от полиномиальной?
36. Что такое «поиск ассоциативных правил»? Опишите алгоритм Apriori. Приведите пример практического применения.
37. Что такое «редукция размерности»? Для чего она нужна? Опишите метод главных компонент (PCA).
38. Какие методы относятся к описательным моделям? Какие – к прогнозным?
39. Что такое «переобучение» модели? Как его обнаружить и предотвратить?
40. Какие способы оценки качества моделей машинного обучения вы знаете (cross-validation, hold-out)?
41. Как выбрать функцию потерь (loss function) для задачи регрессии? Для задачи классификации?
42. Что такое «байесовский подход» в анализе данных? В чём его отличие от классического?
43. Какие методы ансамблирования (бэггинг, бустинг, стекинг) вы знаете? Приведите примеры алгоритмов.

44. Как интерпретировать коэффициенты линейной регрессии? Что они означают?
45. Какие языки программирования наиболее часто используются для анализа данных? Назовите их основные библиотеки.
46. Что такое «аналитическая платформа»? Приведите примеры (Tableau, Power BI, Qlik).
47. Какие возможности предоставляет библиотека pandas в Python? Для чего используется?
48. Для каких целей применяется библиотека scikit-learn?
49. Что такое «SQL» и как он используется в анализе данных?
50. Какие инструменты для работы с большими данными (Spark, Hadoop) вы знаете?
51. Как автоматизировать пайплайны анализа данных? Какие инструменты используются (Airflow, Luigi, etc.)?
52. Какие онлайн-платформы (Kaggle, Google Colab, Jupyter Notebook) используются для анализа данных? В чём их особенности?
53. Что такое «воспроизводимый анализ» (reproducible research)? Какие практики его обеспечивают?
54. Как оценить практическую значимость полученной модели, а не только статистическую точность?
55. Какие методы интерпретации моделей «чёрного ящика» вы знаете (SHAP, LIME)?
56. Как представить результаты анализа данных для заинтересованных сторон (stakeholders)?
57. Что такое «дашборд» (dashboard)? Какие элементы он должен содержать для эффективного мониторинга?
58. Как организовать внедрение результатов анализа в бизнес-процессы?
59. Какие этические проблемы могут возникать при анализе данных (приватность, предвзятость)?
60. Какие способы защиты данных при анализе вы знаете (анонимизация, дифференциальная приватность)?
61. Что такое «АВ-тестирование»? Как результаты анализа данных используются в нём?
62. Как проверить, что модель продолжает работать корректно после внедрения (мониторинг дрейфа данных)?
63. Как составить отчёт по результатам аналитического исследования? Какие разделы он должен включать?
64. Что такое «проектный треугольник» в управлении ИТ-проектами?
65. Опишите разницу между гибкими (Agile) и традиционными (Waterfall) методологиями управления проектами.

7.3.2. Критерии и нормы оценки

Экзамен проводится в письменной форме, максимальное количество баллов за экзамен – **50**.
Оценка знаний на экзамене производится по следующим критериям:

Баллы	Традиционная оценка	Критерии
45–50	отлично	Студент исчерпывающе и грамотно отвечает на вопросы экзаменационного билета. Ответы полные, развёрнутые, демонстрируют глубокое понимание теоретического материала и умение применять его на практике. Допускается одна небольшая неточность, исправленная при ответе на дополнительные вопросы.
35–44	хорошо	Студент даёт полный, развёрнутый ответ на один из вопросов билета, на второй вопрос – тезисные высказывания, либо допускает небольшие неточности, но в целом владеет материалом. На дополнительные вопросы отвечает кратко, но без грубых ошибок.
25–34	удовлетворительно	Студент не может дать полный ответ на один из вопросов билета, либо отвечает на оба вопроса только в виде кратких тезисов, без достаточной детализации. На дополнительные вопросы отвечает неуверенно, допускает ошибки.
менее 25	неудовлетворительно	Студент не отвечает на вопросы билета либо допускает фундаментальные ошибки, свидетельствующие об отсутствии необходимых знаний. На дополнительные вопросы ответить не может.

Оценивание

1. **Допуск к экзамену** осуществляется при набранном текущем рейтинге (сумма баллов за практические работы) не менее **55** баллов (по 100-балльной шкале) или по иному порогу, установленному кафедрой.

2. **Итоговая оценка** по дисциплине формируется как сумма баллов текущего рейтинга (максимум 50) и баллов за экзамен (максимум 50).
Максимальный итоговый балл – **100**.

3. **Шкала перевода итогового балла в традиционную оценку** (при необходимости):

- 85–100 баллов – «отлично»
- 70–84 балла – «хорошо»
- 55–69 баллов – «удовлетворительно»
- менее 55 баллов – «неудовлетворительно»

8. Учебно-методическое и информационное обеспечение дисциплины

8.1. Обязательная литература

№ п/п	Авторы, составители	Заглавие (заголовок)	Тип (учебник, учебное пособие, учебно- методическое пособие, практикум, др.)	Год издания	Количество в научной библиотеке / Наименование ЭБС
1	Маккинли, У.	Python и анализ данных : учебное пособие / У. Маккинли ; перевод с англ. А. Слинкина. — 2-е изд. — Саратов : Профобразование, 2021. — 482 с. — ISBN 978-5-4488-0146-7.	Учебное пособие	2021	ЭБС «IPRbooks»
2	Брюс, П., Брюс, Э.	Практическая статистика для специалистов Data Science : пер. с англ. — СПб. : БХВ-Петербург, 2021. — 304 с. — ISBN 978-5-9775-6785-5.	Учебное пособие	2021	ЭБС «Лань»
3	Вандер Плас, Дж.	Python для сложных задач: наука о данных и машинное обучение : пер. с англ. — СПб. : Питер, 2021. — 576 с. — ISBN 978-5-4461-1464-7.	Учебное пособие	2021	ЭБС «Лань»
4	Уортен, Н.	Визуализация данных в Python с помощью matplotlib и seaborn : учебное пособие / Н. Уортен ; пер. с англ. — М. : ДМК Пресс, 2022.	Учебное пособие	2022	ЭБС «Znanium»

№ п/п	Авторы, составители	Заглавие (заголовок)	Тип (учебник, учебное пособие, учебно- методическое пособие, практикум, др.)	Год издания	Количество в научной библиотеке / Наименование ЭБС
		— 350 с. — ISBN 978-5-97060-969-9.			
5	Джеймс, Г., Уиттон, Д., Хасты, Т., Тибширани, Р.	Введение в статистическое обучение с примерами на языке R : пер. с англ. — М. : ДМК Пресс, 2022. — 552 с. — ISBN 978-5- 97060-978-1.	Учебник	2022	ЭБС «IPRbooks»

8.2. Дополнительная литература

№ п/п	Авторы, составители	Заглавие (заголовок)	Тип (учебник, учебное пособие, учебно- методическое пособие, практикум, др.)	Год издания	Количество в научной библиотеке / Наименование ЭБС
1	Агравал, Д., Филипп, А.,	Анализ данных: от отчетности к решению : пер. с англ. — М. : Альпина Паблишер, 2023. — 320 с. —	Учебное пособие	2023	ЭБС «Znanium»

№ п/п	Авторы, составители	Заглавие (заголовок)	Тип (учебник, учебное пособие, учебно- методическое пособие, практикум, др.)	Год издания	Количество в научной библиотеке / Наименование ЭБС
	Ганьон, А.	ISBN 978-5-9614-7878-9.			
2	Прохоренок, Н. А.	Python. Самое необходимое : учебное пособие. — СПб. : БХВ-Петербург, 2022. — 512 с. — ISBN 978-5-9775-6833-3.	Учебное пособие	2022	ЭБС «Лань»
3	Кетков, Ю. Л., Кетков, А. Ю.	Анализ данных и машинное обучение в Python : учебное пособие. — М. : Горячая линия – Телеком, 2021. — 424 с. — ISBN 978-5-9912-0856-3.	Учебное пособие	2021	ЭБС «IPRbooks»
4	Леоненков, А. В.	Статистический анализ данных в MS Excel : учебное пособие. — М. : ИНФРА-М, 2022. — 320 с. — ISBN 978-5-16-017152-4.	Учебное пособие	2022	ЭБС «Znanium»
5	Брусенцев, А. Г.	Анализ данных и процессов. Ч.1. Методы статистического анализа данных : учебное пособие / А. Г. Брусенцев. — Белгород : Белгородский государственный технологический университет им. В.Г. Шухова, 2020. — 63 с. — ISBN 978-5-361-00540-6.	Учебное пособие	2020	ЭБС «IPRbooks»
6	Алексеев, С.	Анализ данных в социологии : учебно-методическое	Учебно-	2019	ЭБС «IPRbooks»

№ п/п	Авторы, составители	Заглавие (заголовок)	Тип (учебник, учебное пособие, учебно- методическое пособие, практикум, др.)	Год издания	Количество в научной библиотеке / Наименование ЭБС
	А.	пособие / С. А. Алексеев ; под ред. Л. Г. Шевчука. — Казань : Казанский национальный исследовательский технологический университет, 2019. — 92 с. — ISBN 978-5-7882-2617-0.	методическое пособие		
7	Истомина, А. П.	Анализ данных качественных исследований : лабораторный практикум / сост. А. П. Истомина. — Ставрополь : Северо-Кавказский федеральный университет, 2018. — 108 с.	Лабораторный практикум	2018	ЭБС «IPRbooks»
8	Шнарева, Г. В., Пономарева, Ж. Г.	Анализ данных : учебно-методическое пособие / Г. В. Шнарева, Ж. Г. Пономарева. — Симферополь : Университет экономики и управления, 2019. — 129 с.	Учебно-методическое пособие	2019	ЭБС «IPRbooks»
9	Синева, И. С.	Анализ данных в среде R. Ч. 1 : учебное пособие / И. С. Синева. — Москва : Московский технический университет связи и информатики, 2018. — 32 с.	Учебное пособие	2018	ЭБС «IPRbooks»

8.3. Перечень профессиональных баз данных и информационных справочных систем

- Научная электронная библиотека elibrary <http://elibrary.ru>
- Электронно-библиотечная система издательства «Лань» <https://e.lanbook.com/>

8.4. Перечень программного обеспечения

№ п/п	Наименование ПО	Реквизиты договора (дата, номер, срок действия)
1	Microsoft Office Standard	Договор № 690 от 19.05.2015г., срок действия – бессрочно; Договор № 727 от 20.07.2016г., срок действия – бессрочно
2	DreamSpark в составе: Microsoft Visio; Microsoft Visual Studio; Microsoft Access; Microsoft Project	До 01.07.2020. Продлевается каждые 3 года

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

№ п/п	Наименование оборудованных учебных кабинетов, лабораторий, мастерских и др. объектов для проведения практических и лабораторных занятий, помещений для самостоятельной работы обучающихся (номер аудитории)	Перечень основного оборудования
1	Компьютерный класс. Учебная аудитория для проведения занятий лекционного типа. Учебная аудитория для проведения занятий семинарского типа. Учебная аудитория для проведения лабораторных работ. Учебная аудитория для курсового проектирования (выполнения курсовых работ). Учебная аудитория для проведения групповых и индивидуальных консультаций Учебная аудитория для проведения занятий текущего контроля и промежуточной аттестации (УЛК-408).	Компьютер (монитор 17", системный блок Intel (R) Celeron (R) 2,66 GHz / 1 Gb / 80 Gb), маршрутизатор 2801 Router, коммутатор Catalyst, экран/интерактивная доска Smart Board TV, проектор Acer P1303W., стол преподавательский, стол ученический, стол компьютерный, стул, доска аудиторная (маркерная).
2	Учебная аудитория для проведения занятий лекционного типа. Учебная аудитория для проведения занятий семинарского типа. Учебная аудитория для курсового проектирования (выполнения курсовых работ). Учебная аудитория для проведения групповых и индивидуальных консультаций. Учебная аудитория для проведения занятий текущего контроля и	Стол ученический двухместный (моноблок), стол преподавательский, стул, доска аудиторная (меловая).

№ п/ п	Наименование оборудованных учебных кабинетов, лабораторий, мастерских и др. объектов для проведения практических и лабораторных занятий, помещений для самостоятельной работы обучающихся (номер аудитории)	Перечень основного оборудования
	промежуточной аттестации (УЛК-413).	
3	<p>Учебная аудитория для проведения занятий лекционного типа.</p> <p>Учебная аудитория для проведения занятий семинарского типа. Учебная аудитория для курсового проектирования (выполнения курсовых работ). Учебная аудитория для проведения групповых и индивидуальных консультаций. Учебная аудитория для проведения занятий текущего контроля и промежуточной аттестации (УЛК-418).</p>	<p>Учебная аудитория для проведения занятий лекционного типа.</p> <p>Учебная аудитория для проведения занятий семинарского типа. Учебная аудитория для курсового проектирования (выполнения курсовых работ). Учебная аудитория для проведения групповых и индивидуальных консультаций. Учебная аудитория для проведения занятий текущего контроля и промежуточной аттестации.</p>
4	Помещение для самостоятельной работы студентов (Г-401)	Стол ученический, стул, компьютер с выходом в сеть интернет.